The Economics of Non-Market Goods and Resources

Applications of Simulation Methods in Environmental and Resource Economics

Riccardo Scarpa and Anna Alberini (Eds.)

Series Editor
Ian J. Bateman

The Economics of Non-Market Goods and Resources

Springer

APPLICATIONS OF SIMULATION METHODS IN ENVIRONMENTAL AND RESOURCE ECONOMICS

THE ECONOMICS OF NON-MARKET GOODS AND RESOURCES

VOLUME 6

Series Editor: Dr. Ian J. Bateman

Dr. Ian J. Bateman is Professor of Environmental Economics at the School of Environmental Sciences, University of East Anglia (UEA) and directs the research theme Innovation in Decision Support (Tools and Methods) within the Programme on Environmental Decision Making (PEDM) at the Centre for Social and Economic Research on the Global Environment (CSERGE), UEA. The PEDM is funded by the UK Economic and Social Research Council. Professor Bateman is also a member of the Centre for the Economic and Behavioural Analysis of Risk and Decision (CEBARD) at UEA and Executive Editor of Environmental and Resource Economics, an international journal published in cooperation with the European Association of Environmental and Resource Economists. (EAERE).

Aims and Scope

The volumes which comprise *The Economics of Non-Market Goods and Resources* series have been specially commissioned to bring a new perspective to the greatest economic challenge facing society in the 21st Century; the successful incorporation of non-market goods within economic decision making. Only by addressing the complexity of the underlying issues raised by such a task can society hope to redirect global economies onto paths of sustainable development. To this end the series combines and contrasts perspectives from environmental, ecological and resource economics and contains a variety of volumes which will appeal to students, researchers, and decision makers at a range of expertise levels. The series will initially address two themes, the first examining the ways in which economists assess the value of non-market goods, the second looking at approaches to the sustainable use and management of such goods. These will be supplemented with further texts examining the fundamental theoretical and applied problems raised by public good decision making.

For further information about the series and how to order, please visit our Website
www.springeronline.com

Applications of Simulation Methods in Environmental and Resource Economics

Edited by

Riccardo Scarpa

University of York, U.K.,
and University of Waikato,
New Zealand

and

Anna Alberini

University of Maryland, U.S.A.

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 1-4020-3683-3 (HB)
ISBN-13 978-1-4020-3683-5 (HB)
ISBN-10 1-4020-3684-1 (e-book)
ISBN-13 978-1-4020-3684-2 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springeronline.com

Printed on acid-free paper

Printed in the Netherlands.

This volume is dedicated to our respective partners and immediate family.

Contents

List of Figures

List of Tables

Contributing Authors

Anna Alberini is Associate Professor at the Department of Agricultural and Resource Economics, University of Maryland, U.S.A.

Jorge E. Araña is Associate Professor Department of Applied Economic Analysis University of Las Palmas de Gran Canaria, Spain.

Ken Baerenklau is Assistant Professor in economics, in the Economics Department of the University of California, Riverside, U.S.A.

Giovanni Baiocchi is Lecturer in Economics, in the Economics Department of the university of Durham, U.K.

Michel Bierlaire is Maître d'Enseignement et de Recherche at the Institute of Mathematics of the Ecole Polytechnique Fédérale de Lausanne (EPFL), Extraordinary Professor at the University of Stellenbosch, South-Africa, and Research Affiliate at the Department of Civil and Environmental Engineering of the Massachusetts Institute of Technology, U.S.A.

William S. Breffle is Managing Economist at Stratus Consulting, Inc. in Boulder, Colorado, U.S.A.

Joseph Cooper is Deputy Director, Resource Economics Division Economic Research Service, United States Department of Agriculture, Washington, DC, U.S.A.

Silvia Ferrini is Research Fellow at the Environment Department, University of York, U.K. and Research Doctor Department of Statistics, University of Florence, Italy.

Margarita Genius is Lecturer of Economics at the Department of Economics, University of Crete, Greece.

William Greene is Professor of Economics in the Economics Department of the Stern Business School, University of New York, U.S.A.

Wade L. Griffin is Professor, Department of Agricultural Economics Department at Texas A&M University, U.S.A.

Roger von Haefen is Assistant Professor in the Department of Agricultural and Resource Economics at the University of Arizona. During 2004, He was a visiting scholar in the Department of Economics and Center for Environmental Science and Policy at Stanford University, CA, U.S.A.

David A. Hensher is Professor of Management, Faculty of Economics and Business, and Director of the Institute of Transport Studies. University of Sydney, New South Wales.

Stephane Hess is Researcher in the Centre for Transport Studies at Imperial College London, and a part-time employee of Rand Europe, U.K.

Garth Holloway is Reader in Agricultural Economics and Statistics at the University of Reading, Berkshire, U.K.

Xavier Irz is Lecturer at the Department of Agricultural and Food Economics, University of Reading, U.K.

David F. Layton is Associate Professor at the Daniel J. Evans School of Public Affairs, University of Washington, Seattle, U.S.A.

Carmelo León is Professor, Department of Applied Economic Analysis, University of Las Palmas de Gran Canaria, Spain.

Richard A. Levine is Associate Professor of Statistics, Department of Mathematics and Statistics, San Diego State University, U.S.A.

John G. McPeak is Assistant Professor in the Department of Public Administration of the Maxwell School at Syracuse University. He also holds a courtesy appointment in the Department of Economics at Syracuse University, U.S.A.

Klaus Moeltner is Assistant Professor at the Department of Resource Economics, University of Nevada, Reno, U.S.A.

Edward R. Morey is Professor of Economics, Department of Economics at the University of Colorado-Boulder, U.S.A.

Daniel Phaneuf is Associate Professor in the Department of agricultural and resource economics at North Carolina State University where is also Associate director of NCSU's Center for Environmental and Resource Economics Policy, North Carolina, U.S.A.

John W. Polak is Professor of Transport Demand in the Department of Civil and Environmental Engineering and Head of the Centre for Transport Studies at Imperial College London, U.K.

Bill Provencher is Associate Professor at the Department of Agricultural and Applied Economics, University of Wisconsin, Madison, U.S.A.

John M. Rose is Lecturer in Transportation and Logistics Management in the Institute of Transport Studies, Faculty of Economics and Business at the University of Sydney, New South Wales.

Riccardo Scarpa is Professor, Department of Economics, Management School, University of Waikato, N.Z. and senior Lecturer Environment Department, University of York, U.K.

Garrett Sonnier is Doctoral Student University of California at Los Angeles, Anderson School of Management, CA, U.S.A.

Elisabetta Strazzera is Associate Professor of Economics at the University of Cagliari, Italy.

David Tomberlin is Research Economist National Marine Fisheries Service, Santa Cruz, CA, U.S.A.

Kenneth Train is Adjunct Professor of Economics and Public Policy at the University of California, Berkeley and Vice-President of National Economic Research Associates, U.S.A.

Donald M. Waldman is Professor, Department of Economics, University of Colorado at Boulder, U.S.A.

Melvyn Weeks is Lecturer in Economics at the University of Cambridge, U.K.

Kenneth G. Willis is Professor, School of Architecture landscape and planning, University of Newcastle, Newcastle upon Tyne, U.K.

Richard T. Woodward is Associate Professor in the Department of Agricultural Economics Department at Texas A&M University, Texas, U.S.A.

Yong-Suhk Wui is Assistant Professor in the Department of Business Administration and Economics, University of Arkansas at Pine Bluff, U.S.A.

Andreas Ziegler is Researcher at the Centre for European Economic Research (ZEW), Department Environmental and Resource Economics, Environmental Management, Germany.

Foreword

The growth of simulation based methods over the past fifteen years, together with the accompany improvements in computational power, has ushered in a new era of applied research in economics generally and environmental and resource economics in particular. The analyst is no longer constrained to choose from among a limited set of models simply because they yield a convenient functional form for, say, the log-likelihood function in a discrete choice model or the posterior distribution in a Bayesian analysis. Instead, one can specify more flexible models that allow agents to reveal patterns in their behavior through the data. Maximum simulated likelihood procedures can be used to allow for complex patterns of correlations among choice alternatives in a mixed logit or multinomial probit model of discrete choice, rather than having to impose a priori the rigid structure of the more traditional multinomial or nested logit models. Bayesians can specify prior distributions that reflect their actual prior beliefs, rather than being limited to a convenient set of conjugate priors. Indeed, it is increasingly the case that the questions that can be addressed in applied research are constrained by limitations in the available data, rather than by the models that can be feasibly estimated.

Despite the promise offered by simulation methods, many practitioners continue to avoid their use, daunted by the perceived difficulty of the techniques themselves or the prospect of programming the routines in software packages such as GAUSS or Matlab. Moreover, graduate students often receive little or no training in simulation methods and are forced to learn the necessary tools on their own. This volume provides a valuable resource in this learning process. Alberini and Scarpa have gathered in one place eighteen papers covering a wide range of simulation issues and techniques with applications to environmental and natural resource problems. The topics range from estimation procedures, such as Train and Weeks (in the opening chapter) use of a Bayesian framework and MCMC methods to obtain estimates for the mixed logit model, to discussions of important computational issues (e.g., the choice of random number generators and the tradeoffs between using Gaussian quadrature versus simulation methods for integration). The authors, all experienced practitioners in the use of simulation methods, provide valuable and practical insights into the proce-

dures. In most cases, step-by-step algorithms are provided for the simulation method under discussion, with the underlying computer code available from the respective authors. For those of us who have made extensive use of the mixed logit code graciously provided by Kenneth Train in the past, it is clear that such computer code can substantially lower the learning costs associated with simulation based methods. Just as important, the authors provide insights into the limitations and potential pitfalls associated with use of simulation based methods. Simulation methods greatly expand the set of models that can be feasibly estimated in a given setting. Unfortunately, this makes it all the more important to understand the underlying limitations of a model and how the ways in which they are structured and estimated, rather than the data itself, can determine the outcomes of an analysis. The complex econometric methods that simulation techniques allow are still no substitute for sound economic modeling and careful data collection. The insights provided in this volume should help practitioners in the proper use of simulation based methods.

Joseph Herriges
Iowa State University
Ames, Iowa, February 2005

Acknowledgments

As any editor of a similar volume will know, and it is quite easy to imagine, this product can rarely be completed without the contribution of many. Credit must be given where credit is due.

The volume collects the work of 36 authors in 18 papers, plus the editors' assistants. We are indebted to all of them, for their work, patience and support. A special 'thank you' must go to Kenneth Train for both the initial inspiration and the encouragement throughout. Had it not been for his work, this volume would never have been conceived.

In the process of scientific review of the manuscripts we took advantage of the generosity of time of many colleagues who are not co-authors of any chapter in the volume. We are thankful to—amongst others—Elisabetta Cherchi (University of Cagliari), Andreas Kontoleon (University of Cambridge), Gary Koop (University of Leister), and Luis Rizzi (Catholic University of Santiago, Chile) for their constructive comments during the reviewing effort.

We are also very grateful to those who pined over the translation from MSword documents into the LaTeX typeset. These include many that helped along the way, but we feel we should name the members of the editorial team at the Environment Department of the University of York, who really worked at the "coal-face": Nesha Beharry, Chung-Ting Chang and Jan Minx for their hard and proficient work. I am sure that for some of them the 'discovery' of LaTeX involved substantial effort.

The help and guidance provided by Springer publisher, and in particular by Esther Verdries, are also gratefully acknowledged.

Introduction

Anna Alberini
University of Maryland, U.S.A.
aalberini@arec.umd.edu

Riccardo Scarpa
University of Waikato, N.Z. and University of York, U.K.
rs24@york.ac.uk

1. Background and motivation

This volume collects a series of empirical research papers in environmental and resource economics with one common feature: They all make use of simulation methods (SMs). The rapid development of the computational power in computers that we have experienced in the last twenty years has brought about momentous changes in the techniques used in applied economics.

This has unleashed an unprecedented growth in the versatility of behavioural models that can be empirically investigated, provided that adequate data become available. The new fruits of this increased computational power are now being harvested throughout economics, with many applications in environmental and resource economics.

Anthony Fisher, in his 2004 keynote speech to the European Association of Environmental and Resource Economists in Budapest was asked to list what he thought were the three most promising areas of research in the discipline. In his answer he mentioned the promise of benefit transfer, experimental economics and simulation based methods.

The growth in computational power and the widespread adoption of SMs has also posed new questions. With new research frontiers opening up to the

profession, new challenges have emerged, and now need to be faced and successfully tackled. At the same time old distinctions between paradigms—such as the classical *vs* the Bayesian dichotomy—have sometimes become blurred and hence lost some of their rationale and methodological bite (as shown in chapter 2), and become even complementary (as shown in chapters 7 and 8).

The considerable dynamism that characterizes this area of research makes it difficult to regularly incorporate such advances into conventional university courses. It is our perception that—perhaps with the exception of large research establishments—most post-graduate training in environmental and resource economics today does not include adequate exposure to SMs. We hope that this volume will help fill this gap, and that PhD supervisors will refer their supervisees to it for a primer on the subject.

This collection was hence conceived to bring together in a single volume a significant selection of research papers by leading practitioners in this field. Each chapter has been peer-reviewed either by contributors or by other qualified reviewers (see acknowledgements). At the same time the structure of writing was deliberately pedagogical in nature.

Many areas and research questions within natural resource and environmental economics naturally lend themselves to the application of simulation methods. The prevalence of articles on non-market valuation in this volume mirrors the large role played by this line of research in the profession at the moment.

As any researcher using SMs will recognize, the writing of specific codes to implement simulation algorithms is a key component of a researcher's toolkit. Most of the contributors to this collection have very kindly agreed to making their respective codes available (of course, without any guarantee). Considering the amount of time that this activity normally requires, their generosity should—in our opinion—generate substantial external benefits. We very much hope that those readers who will use these routines in their own research will take the time to mention the source of these routines in their published works.

The chapters of the volume were arranged on the basis of some themes. In what follows we briefly discuss each theme and its chapters.

2. Heterogeneity in discrete choice models

Simulation methods are increasingly being used to estimate discrete-choice models, and the related welfare measures for non-market goods. Recent research in this area has attempted to incorporate constraints and checks so as to ensure that measures of marginal prices, willingness to pay, and other relevant welfare statistics are behaviourally plausible in the presence of unobserved heterogeneity. For this reason the first four chapters in this volume are concerned with heterogeneity in discrete choice models.

In the first chapter, Kenneth Train and Melvyn Weeks explore the consequences of framing heterogeneity in the *WTP* space, in contrast with the more common option of heterogeneity in preference space. The estimation is conducted in a Bayesian framework using Monte Carlo Markov Chains and the authors highlight correlation issues previously overlooked in the literature. The novelty of introducing parameter heterogeneity *directly* in the expenditure function is no doubt very appealing and it has already proven to be a preferred approach by many non-market valuation analysts since its introduction by Cameron, 1988 and Cameron and James, 1987 in the contingent valuation literature.

In Chapter 2, William Greene, David Hensher and John Rose focus on the use of mixed logit with bounded parameter distributions to achieve behaviourally plausible *WTP* distributions and on the derivation of individual-specific *WTP* estimates conditional on observed choices, which they compare with the conventional population-based statistics. They pay special attention to large and implausible *WTP* values. The technique they propose is implemented in Nlogit, a popular econometric package. We expect that the approach they suggest will be used in numerous applied papers in the future, especially because some preliminary evidence seems to suggest that the distribution of these conditional estimates in the sample is less prone to the presence of behaviorally implausible value estimates.

Heterogeneity of *WTP* distributions is also the focus of chapter 3, authored by David Layton and Klaus Moeltner who present a contingent valuation study of *WTP* to avoid power outages. The authors face numerous econometric challenges and propose an approach that deals with all of these based on a Gamma distribution with heterogeneity addressed by a scale parameter with log-normal distribution. Such creative use of mixed and flexible models is one of the gifts delivered by advances in SMs.[1]

In the final chapter of this section (chapter 4) Stephane Hess, Michel Bierlaire and John Polak put forward a potential solution to the problem of confounding of correlation between alternatives and taste variation across individuals. To separately address these issues without confounding they use a mixed generalized extreme value model. As the use of mixed logit error-component models becomes more established in the environmental economics literature (see for example Herriges and Phaneuf, 2002 and chapter 13 in this volume) this line of research will become increasingly important. The estimation is conducted using BIOGEME, the software developed by Bierlaire and available from this author.

[1] These authors employed MatLab code and have decided to make it available upon request.

In chapter 5, Joe Cooper specifies a multinomial probit model (MNP) for stated-preference adoption or non-adoption of one of five best management practices under the federal Environmental Quality Incentive Program given a hypothetical payment to the farmer. Estimation of the MNP requires maximum simulated likelihood methods. Cooper further incorporates the semi-parametric approach to modeling binary data willingness-to-accept (*WTA*) responses first introduced by Creel and Loomis, 1997, and illustrates this model using the data from a survey of U.S. farmers. A likelihood ratio test confirms Cooper's suspicions that farmer *WTA* for one practice is not independent of that for another practice. An approach for predicting *WTA* for one practice conditional on the bids for the other practice is then proposed.

In chapter 6, Andreas Ziegler is motivated by a broad question about the role of environmental regulation and firm self-regulation: What factors drive firms to adopt environmental product innovations? Specifically, is the adoption of such innovations affected by market conditions, including competitive pressure, client relations, etc.? Ziegler compares the multinomial logit model with the multinomial probit model, where the latter is estimated using the Geweke-Hajivassiliou-Keane simulator. Ziegler finds that the simulated ML estimates of the multinomial probit model are unreliable, a result that he attributes, among other things, to the nature of the multinomial model itself, where the only explanatory variables are firm characteristics that do not vary across alternatives. He concludes that the multinomial probit model offers few new insights above and beyond those of the multinomial logit.

3. Bayesian applications

The following section collects five chapters that employ Bayesian estimation techniques. They either compare these to classical ones, or show how these can supplement classical ones. Of course this is another area in which SMs have had a tremendous impact. In the opening chapter of this section (chapter 7), Ken Train and Garrett Sonnier illustrate the use of efficient Bayesian estimation in deriving taste-parameter distributions with bounds that reflect behavioural expectations and full correlation across parameters of the utility function. The associate GAUSS code has been available for some time from the web-page of Prof. Train, and it is very flexible. Their empirical application reports on a study of preferences for environmentally friendly car engine solutions, comparing electric, hybrid and conventional engines.

Chapter 8, which we judge very useful from a teaching point of view. is contribution by Roger van Haefen and Dan Phaneuf, two leading researchers in the field of demand systems for quality-differentiated goods in the Kuhn-Tucker framework. This line of research has benefited greatly from the use of SMs. In this chapter the various challenges that the profession has had to

face to implement this elegant approach to the travel-cost method are carefully described, and the role of SMs highlighted in both parameter estimation and derivation of welfare measures form moose hunting. The authors propose three estimation approaches, one of which uses Bayesian techniques, and make their GAUSS code available to the reader. Hopefully this chapter will lead to a more widespread use of this approach, which is so far the most theoretically consistent with consumer theory.

Holloway, Tomberlin and Irz in chapter 9 move our attention to the use of SMs to estimate production efficiency in a trawl fishery with a purely Bayesian hierarchical approach based on specifically developed error-component model and MCMC-assisted estimation. Estimation of efficiency frontiers is a 'classic' problem in empirical production economics and of high relevance in the economics of fishery resources.[2] Powerful and robust Bayesian techniques of this kind are obvious assets in the tool-kit of applied resource economists.

In chapter 10 Layton and Levine address the issue of how to incorporate previous information of past studies of pilot and pre-test surveys to improve the quality of estimation from the final survey data. This is a natural context of application for a Bayesian analytical framework, especially in multi-stage data collection in which sequenced updating of posteriors is possible. As computer aided survey administration becomes more widespread such techniques will become of greater value. The authors illustrate the value of their approach using stated preference data on surveys for the protection of the northern spotted owl in Seattle, Washington.

Araña and León in chapter 11 illustrate how to analyze discrete-choice contingent valuation data single- and double-bounded in a Bayesian framework. They run Monte Carlo experiments to compare conventional maximum likelihood analysis with Bayesian ones and find the latter improves the performance of the model, particularly with relatively small samples. This result is of great interest given the cost of contingent valuation surveying.[3]

4. Simulation methods in dynamic models

Two chapters in this book are dedicated to the use of SMs for solving and exploring the characteristics of dynamic models. This area of research is of germane importance in natural resource economics.

Richard Woodward, Wade Griffin and Yong-Suhk Wui focus on solutions and on approximation to solutions of dynamic programming (DP) models that are tied to large simulation models. They compare and discuss the pros and cons of a direct approach, in which the simulation model is embedded in the

[2]The authors employ MatLab and the code is available upon request.
[3]Their GAUSS code is available upon request from the authors.

DP algorithm, to an indirect one in which functional relationships are approximated using an econometrically estimated meta-model. In their application they tackle a complex management problem in the red snapper fishery in the Gulf of Mexico. The policy variables examined here are (i) the total allowable catch (TAC) for the red snapper fishery, and (ii) the distribution of the TAC between commercial fishermen and recreational anglers. Woodward *et al.* find that the metamodeling approach to be less computationally burdensome, but the direct approach is superior in terms of plausibility of results, consistency with economic theory, and forecasting performance.

In chapter 15, Bill Provencher and Kenneth Baerenklau make a valuable point in emphasizing that most empirical dynamic models—which address issues such as land development decisions, livestock grazing rates, and timing of timber harvesting—have been based on stylized reduced-form specifications. They therefore present an approach to structural form estimation and apply it to the timber harvesting problem of Brazee and Mendelsohn, 1988. They also discuss the difficulty of distinguishing whether microeconomic data are generated by static or dynamic behaviour even with maintained structural assumptions about the form of the intra-period utility or profit function. We find their contribution to present an important pedagogical perspective.

5. Monte Carlo experiments

Chapters 13 and 12 are dedicated to one of the major work-horses of SMs: context-specific Monte Carlo experiments devised to explore the finite sample properties of estimators for which econometric theory provides the researcher only with asymptotic results.

Riccardo Scarpa, Silvia Ferrini and Ken Willis—in chapter 13—focus on the econometrics of choice experiments in the format commonly used for non-market valuation in environmental economics. In this context it is customary to include in each choice set the "status quo" response option in addition to the other alternatives. Economists and psychologists have, however, worried about status-quo biases and other undesirable response effects induced by the inclusion of this response option, and in this chapter Scarpa *et al.* investigate these effects. They compare three models, two of which are commonly used by practitioners (conditional logit with one status-quo alternative-specific constant and nested logit), while the third is a more flexible mixed logit error component model which nests the other two under specific conditions. Their Monte Carlo results suggest that the mixed logit error component model they proposed is robust to mis-specification errors over a range of commonly employed sample sizes, and should be preferred over the other more commonly employed two.[4]

[4]Their GAUSS code is made available upon request.

In chapter 12, Margarita Genius and Elisabetta Strazzera follow up on an influential paper by Carson, Groves and Machina, 2000, in that they study models appropriate for checking the incentive-compatibility properties of dichotomous choice contingent valuation questions. Briefly, in single-bounded referendum questions the respondent is asked to say whether he would vote in favor or against a public program if the cost to him or her is X. The responses to these referendum questions are incentive-compatible, but statistically inefficient. To refine information about *WTP*, researchers usually include dichotomous-choice follow-up questions. However, the follow-up questions are not incentive compatible.

This paper focuses on bivariate models of *WTP*, e.g. models in which the response to the initial and follow-up payment question are assumed to be driven by two different-and unobserved-*WTP* amounts. The two latent *WTP* amounts are allowed to be correlated. Specifically, the authors consider (i) a bivariate model that restricts the coefficients in the first and second equation to be identical, (ii) a bivariate model with a shift (following Carson et al.), and (iii) a model that truncates the bivariate distribution to allow for the incentives to be different for individuals who answer no to the initial payment question, and individuals who answer yes to the initial payment question.

The authors also worry about using bivariate normal distributions when the true distribution is not normal, and introduce a Joe copula, i.e., a joint distribution for two variables with specified marginals. For each of the three possible data generating processes ((i), (ii) and (iii)), the authors fit all proposed models, examine the effect of fitting a bivariate normal model when the distribution of the latent *WTP* amounts is not a bivariate normal, and experiment with Joe copula models. They find that the latter are flexible, perform well, and have a good track record of convergence, especially when the models based on the traditional bivariate approach do not converge easily.

6. Computational aspects

Although computational aspects are dealt with throughout the volume, the pedagogical intention of this volume on SMs in environmental and resource economics required that we examine such issues in some detail. Accordingly, the last three chapters deal with these specific aspects of simulation-based methods.

Perhaps one of the main concerns in simulation studies using Monte Carlo techniques is the quality of pseudo-random generators. Giovanni Baiocchi deals with these and related aspects such as reproducibility of results and reporting in chapter 16. He reports the outcomes of a battery of tests specifically meant to assess the quality of different pseudo-random generators employed in software commonly used by resource and environment economists. Many of

these outcomes are surprising. We believe readers will find this chapter very useful for learning about potential pitfalls of simulation-based tools.

Bill Breffle, Ed Morey and Donald Waldman in chapter 17—instead— focus on simulation noise in a comparison between quadrature and simulation techniques with pseudo-random draws in approximating integrals without a closed-form in binary probit models. Their context of application is stated preference data from a sample of Green Bay anglers. They find that in some circumstances quadrature affords computational gains with respect to simulation-based estimation.[5]

Finally, simulation noise is also the main theme of chapter 18 by John McPeak. Using his data-set on land use decisions by Kenyan pastoralists, he focuses on simulation noise due to various sources in the bivariate tobit model. He considers how variable characteristics influence parameter variability across estimation runs, and identifies specific characteristics that influence variability in his results. While McPeak concludes that in his dataset simulation noise is not large enough to lead the analyst to incorrect conclusions, his concern is of high relevance to SMs practitioners, who—in our opinion—should systematically carry out tests to check whether this is indeed the case. As a consequence his approach is of general interest.[6]

7. Terminology

In an ideal world there would be a one-to-one mapping between terms and concepts. Initially we intended to standardize the terminology throughout the book in as much as possible. Although we tried, we now feel we did not go very far with it. As a result, for example, many acronyms of models across chapters may refer to different econometrics specifications. While an apology in this sense is due to the reader, we feel that the diversity of terms in this book reflects that in the current literature. Hopefully, in future there will be a natural evolution of technical jargon towards some kind of standard.

[5]Their GAUSS code is made available upon request.

[6]John McPeak also uses code for GAUSS and he makes it available upon request.

Chapter 1

DISCRETE CHOICE MODELS IN PREFERENCE SPACE AND WILLINGNESS-TO-PAY SPACE

Kenneth Train
Department of Economics, University of California, Berkeley
train@econ.berkeley.com

Melvyn Weeks
University of Cambridge, Cambridge, U.K.
mw217@econ.cam.ac.uk

Abstract In models with unobserved taste heterogeneity, distributional assumptions can be placed in two ways: (1) by specifying the distribution of coefficients in the utility function and deriving the distribution of willingness to pay (*WTP*), or (2) by specifying the distribution of *WTP* and deriving the distribution of coefficients. In general the two approaches are equivalent, in that any mutually compatible distributions for coefficients and *WTP* can be represented in either way. However, in practice, convenient distributions, such as normal or log-normal, are usually specified, and these convenient distributions have different implications when placed on *WTP*'s than on coefficients. We compare models that use normal and log-normal distributions for coefficients (called models in preference space) with models using these distributions for *WTP* (called models in *WTP* space). We find that the models in preference space fit the data better but provide less reasonable distributions of *WTP* than the models in *WTP* space. Our findings suggests that further work is needed to identify distributions that either fit better when applied in *WTP* space or imply more reasonable distributions of *WTP* when applied in preference space.

Keywords: Mixed logit, random parameters, random willingness to pay.

1. Introduction

In many applications of discrete choice models with random coefficients, the price coefficient is held constant, especially when the goal is to estimate the distribution of consumers' willingness to pay for alternative attributes (e.g., Revelt and Train, 1998; Goett *et al.*, 2000; Layton and Brown, 2000; Scarpa *et al.*, 2002; Hensher *et al.*, 2004) and/or to infer the willingess to pay of individual consumers from their observed choices and the population distribution (Train, 2003, Ch. 11; Scarpa *et al.*, 2005; Greene *et al.*, 2005.) This restriction allows the distributions of willingness to pay (*WTP*) to be calculated easily from the distributions of the non-price coefficients, since the two distributions take the same form. For example, if the coefficient of an attribute is distributed normally, then *WTP* for that attribute, which is the attribute's coefficient divided by the price coefficient, is also normally distributed. The mean and standard deviation of *WTP* are simply the mean and standard deviation of the attribute coefficient scaled by the inverse of the (fixed) price coefficient. The restriction also facilitates estimation. As Ruud (1996) points out, a model with all random coefficients, including the price coefficient, can be practically unidentified empirically, especially in datasets with only one observed choice for each decision-maker.

A fixed price coefficient,[1] however, implies that the standard deviation of unobserved utility, which is called the scale parameter, is the same for all observations. Louviere (2003) discusses the importance of recognizing that the scale parameter can, and in many situations clearly does, vary randomly over observations and that ignoring this variation in estimation can lead to erroneous interpretation and conclusions. For example, if the price coefficient is constrained to be fixed when in fact scale varies over observations, then the variation in scale will be erroneously attributed to variation in *WTP*.

In this paper we investigate alternative ways to specify random coefficients and *WTP* when the price coefficient varies. Cameron and James (1987) and Cameron (1988) introduced the concept of parameterizing a fixed-coefficient model in terms of *WTP* rather than coefficients. We extend their analysis to models with random coefficients, where distributional assumptions and restrictions can be placed on the coefficients or on the *WTP*'s. The two approaches are formally equivalent, in the sense that any distribution of coefficients translates into some derivable distribution of *WTP*'s, and vice-versa. However, the two approaches differ in terms of numerical convenience under any given distributional assumptions. For example, a model with an attribute coefficient that is normally distributed and a price coefficient that is log-normal implies that *WTP* for the attribute is distributed as the ratio of a normal to a log-normal.

[1] Or, more generally, any fixed coefficient, or uncorrelated random coefficients.

A researcher working directly in *WTP* space is unlikely to choose this inconvenient distribution for *WTP*'s. Conversely, a model with normal *WTP* and log-normal price coefficient implies that the attribute coefficient is the product of a normal and log-normal, which is a distribution that has never, to our knowledge, been applied in preference space. Restrictions are also asymmetric. For example, uncorrelated preference coefficients translate into *WTP*'s that are correlated in a particular way that would be hard to implement and test in the context of *WTP* distributions, and vice-versa.

We estimate and compare models that are parameterized in terms of coefficients, called "models in preference space," and models parameterized in terms of *WTP*, called "models in *WTP* space." For the models in preference space, a convenient distribution is specified for the coefficients, and the parameters of this distribution (such as its mean and variance) are estimated. The distribution of *WTP*'s is then derived from the estimated distribution of coefficients. This is currently the standard practice for application of choice models. For the models in *WTP* space, convenient distributions are specified for the *WTP*'s and the price coefficient. The parameters of this distribution are estimated, from which the estimated distribution of utility coefficients is derived.

We find that models using convenient distributions in preference space fit the data better, both within sample and out-of-sample, than models using convenient distributions in *WTP* space. However, the distributions of *WTP* that are derived from these models have unreasonably large variance, which translates into an untenable implication that many people are willing to pay an enormous amount of money to have or avoid an attribute. Stating the conclusions in combination: the models that fit better give less reasonable distributions for *WTP*. These results suggests that alternative distributional specifications are needed that either fit the data better when applied in *WTP* space or imply more reasonable *WTP* distributions when applied in preference space.

Our analysis and findings mirror those of Sonnier, Ainslee, and Otter (2003), with one exception. In similar comparisons as ours they find that their models in preference space fit the within-sample data better than their models in *WTP* space but provide unreasonably large variances in *WTP*. In these regards, their results match ours. However, they find that their models in *WTP* space attain better out-of-sample fit than their models in preference space, which is opposite of what we find. Sonnier *et al.* (2003) use a different method for evaluating out-of-sample fit than we do, which might account for the difference. However, differences like this one are to be expected over different datasets, since the issue under investigation is the performance of various distributional specifications and the appropriate distribution is necessarily situation-dependent.

2. Specification

In this section we describe the two types of models. Decision-makers are indexed by n, alternatives by j, and choice situations by t. To facilitate discussion, we specify utility as separable in price, p, and non-price attributes, x:

$$U_{njt} = -\alpha_n p_{njt} + \beta_n' x_{njt} + \epsilon_{njt} \tag{2.1}$$

where α_n and β_n vary randomly over decision-makers and ϵ_{njt} is i.i.d. We assume ϵ_{njt} is distributed extreme value, though the analysis is the analogous for other distributions. The variance of ϵ_{njt} can be different for different decision-makers: $Var(\epsilon_{njt}) = k_n^2(\pi^2/6)$, where k_n is the scale parameter for decision-maker n.

Though the utility specification is not yet normalized, the current formulation allows us to clarify the circumstances under which the scale parameter can be expected to vary over decision-makers. A random scale parameter is conceptually different from random values for α and β. α_n and β_n represent the tastes of person n, and these parameters vary over decision-makers because different people have different tastes. In contrast, the scale parameter does not represent a term within the utility function in any given choice situation but rather the standard deviation of utility over different choice situations. By allowing the scale parameter to be random, the researcher gives a variance to a variance. The question arises: what would cause the variance of ϵ to vary? Two prominent situations arise:

1 The unobserved term ϵ might reflect factors that are actually random or quixotic from the decision-maker's perspective, rather than, as in the usual derivation, factors that are known to the decision-maker but unknown by the researcher. In this situation, the variance of ϵ reflects the degree of randomness in the decision-maker's process, which can be expected to differ over decision-makers. This concept of randomness is particularly relevant with stated preference data, where respondents differ in their attention to the task and in their constructs of unlisted attributes. However, randomness in behavior can arise in revealed preference data as well.

2 In panel data settings, each decision-maker faces a sequence of choice situations with unobserved factors differing in each choice situation. It is reasonable to believe in this situation that the variance of these unobserved factors over choice situations for each decision-maker is different for different decision-makers, even when the unobserved factors are known to the decision-maker and unobserved only by the researcher.

These two situations also clarify the converse: When ϵ represents factors that are known to the decision-maker but unknown by the researcher, and only one

choice situation is observed for each decision-maker such that each observation represents a different decision-maker, there is perhaps little need or meaning to allowing the scale parameter to vary over decision-makers. In this circumstance, the scale parameter captures variance over observations in factors that the researcher does not observe; this variance is defined on the researcher, not the decision-maker, and takes a given (i.e., fixed) value for the researcher.

Dividing utility (2.1) by the scale parameter does not affect behavior and yet results in a new error term that has the same variance for all decision-makers:

$$U_{njt} = -(\alpha_n/k_n)p_{njt} + (\beta_n/k_n)'x_{njt} + \varepsilon_{njt} \tag{2.2}$$

where ε_{njt} is i.i.d. type-one extreme value, with constant variance $\pi^2/6$. The utility coefficients are defined as $\lambda_n = (\alpha_n/k_n)$ and $c_n = (\beta_n/k_n)$, such that utility is written:

$$U_{njt} = -\lambda_n p_{njt} + c_n' x_{njt} + \varepsilon_{njt} \tag{2.3}$$

Note that if k_n varies randomly, then the utility coefficients are correlated, since k_n enters the denominator of each coefficient. Specifying the utility coefficients to be independent implicitly constrains the scale parameter to be constant. If the scale parameter varies and α_n and β_n are fixed, then the utility coefficients vary with perfect correlation. If the utility coefficients have correlation less than unity, then α_n and β_n are necessarily varying in addition to, or instead of, the scale parameter.

Equation (2.3) is called the model in preference space. Willingness to pay for an attribute is the ratio of the attribute's coefficient to the price coefficient: $w_n = c_n/\lambda_n$. Using this definition, utility can be rewritten as

$$U_{njt} = -\lambda_n p_{njt} + (\lambda_n w_n)' x_{njt} + \varepsilon_{njt}, \tag{2.4}$$

which is called utility in *WTP* space. Under this parameterization, the variation in *WTP*, which is independent of scale, is distinguished from the variation in the price coefficient, which incorporates scale.[2]

The utility expressions are equivalent of course. Any distribution of λ_n and c_n in (2.3) implies a distribution of λ_n and w_n in (2.4), and vice-versa. The general practice has been to specify distributions in preference space, estimate the parameters of those distributions, and derive the distributions of *WTP* from these estimated distributions in preference space (e.g., Train, 1998.) While fully general in theory, this practice is usually limited in implementation by the use of convenient distributions for utility coefficients. Convenient distributions for utility coefficients do not imply convenient distributions for *WTP*, and

[2]Any coefficient can be used as the base that incorporates scale, with each other coefficient expressed as the product of this coefficient and a term that is independent of scale. The only reason to use the price coefficient as the base is that the scale-free terms become *WTP*'s, which are easy to interpret.

vice-versa. As stated above, if the price coefficient is distributed log-normal and the coefficients of non-price attributes are normal, then *WTP* is the ratio of a normal term to a log-normal term. Similarly, normal distributions for *WTP* and a log-normal for the price coefficient implies that the utility coefficients are the product of a normal term and a log-normal term. The placement of restrictions is similarly asymmetric. It is fairly common for researchers to specify uncorrelated utility coefficients; however, this restriction implies that scale is constant, as stated above, and moreover that *WTP* is correlated in a particular way. It is doubtful that a researcher in specifying uncorrelated coefficients is actually thinking that *WTP* is correlated in this way. Similarly, uncorrelated *WTP*, which the researcher might want to assume or test, implies a pattern of correlation in utility coefficients that is difficult to implement in preference space.

The issue becomes: does the use of convenient distributions and restrictions in preference space or *WTP* space result in more accurate and reasonable models? The answer is necessarily situationally dependent, since the true distributions differ in different applications. However, some insight into the issue can be obtained by comparisons on a given dataset. This is the topic of the next section.

3. Data

We use the stated-preference data collected by Train and Hudson (2000) on households' choice among alternative-fueled vehicles, including gas, electric, and hybrid gas-electric vehicles. 500 respondents were presented with 15 choice situations apiece. For each choice situation, the respondent was given a card that described three vehicles and was asked to state which of the vehicles he/she would choose to buy. Each vehicle was described in terms of the following variables:

- Engine type (gas, electric, or hybrid),
- Purchase price, in dollars,
- Operating cost, in dollars per month,
- Performance (grouped into three levels, which we call "low," "medium," and "high," [3]
- Range between recharging/refueling, in hundreds of miles,

[3]Performance was described on the card in terms of top speed and seconds required to reach 60 mph. However, these two components were not varied independently, and only three combinations of the two components were utilized.

- Body type (10 types ranging from mini car to large van).

Each of the attributes varied over choice situations and over respondents. Range varied for electric vehicles but was constant for gas and hybrid vehicles, since the purpose of this variable was to determine consumers' response to the relatively restricted range of electric vehicles. All but a few respondents completed the fifteen choice tasks, giving a total of 7,437 observations for estimation. These data have been previously used by Hess *et al.* (2003) and Train and Sonnier (2005) for other purposes. We use the data to compare specifications in preference and *WTP* space.

4. Estimation

4.1 Uncorrelated coefficients in preference space

Our first model is specified in preference space with a random coefficient for each variable and no correlation over coefficients. As discussed above, uncorrelated coefficients implies that the scale parameter is fixed. This model can therefore be seen as a version that does not allow for random scale. It is compared with models, described below, that allow random scale.

For this and other models in preference space, the attributes that are desirable, or undesirable, for everyone are given log-normally distributed coefficients. These attributes are: price, operating cost, range, a dummy for medium performance or higher, and a dummy for high performance. The coefficient for the first of the performance variables captures the extra utility associated with increasing performance from low to medium, while the coefficient for the second performance variable reflects the extra utility associated with increasing performance from medium to high. Price and operating cost are entered as negative, since the log-normal distribution implies positive coefficients. The other attributes can be either desirable or undesirable, depending on the views and tastes of the consumer. These attributes are: dummies for electric and hybrid engines, whose coefficients reflect the value of these engine types relative to gas; and dummies for each body type except mid-sized car, whose coefficients reflect the value of these body types relative to a mid-sized car (holding other attributes constant, of course.) The coefficients of these variables are given normal distributions.

The model, and all the ones which follow, was estimated by Bayesian MCMC procedures, using diffuse priors. These procedures for mixed logit models are described by Train (2003) in general and by Train and Sonnier (2005) in relation to these particular data. 10,000 iterations were used as "burn-in" after which every tenth draw was retained from 10,000 additional iterations, providing a total 1,000 draws from the posterior distribution of the parameters. Previous analysis of these data by Train and Sonnier, as well as our own analysis, indicates that the MCMC sequences converged within the burn-in period.

Table 1.1. Model in Preference Space with Uncorrelated Coefficients

Attribute	*Parameter*	*Estimate*	*St. error*
Price in $10,000's	Mean of ln(–coeff.)	–0.2233	0.0508
	Variance of ln(–coeff.)	0.5442	0.0635
Operating cost in $/month	Mean of ln(–coeff.)	–3.5540	0.0993
	Variance of ln(–coeff.)	0.7727	0.1449
Range in 100's of miles	Mean of ln(coeff.)	–0.7272	0.1298
	Variance of ln(coeff.)	0.3317	0.1209
Electric engine	Mean of coeff.	–1.9453	0.1354
	Variance of coeff.	1.6492	0.2820
Hybrid engine	Mean of coeff.	0.8331	0.1102
	Variance of coeff.	1.4089	0.1797
High performance	Mean of ln(coeff.)	–3.0639	0.3546
	Variance of ln(coeff.)	3.3681	0.8493
Medium or high performance	Mean of ln(coeff.)	–1.3030	0.2630
	Variance of ln(coeff.)	1.4041	0.5204
Mini car	Mean of coeff.	–3.0325	0.1767
	Variance of coeff.	3.5540	1.0535
Small car	Mean of coeff.	–1.3966	0.1240
	Variance of coeff.	1.3086	0.4290
Large car	Mean of coeff.	–0.4008	0.1272
	Variance of coeff.	1.3084	0.7080
Small SUV	Mean of coeff.	–0.8499	0.1072
	Variance of coeff.	0.7032	0.3655
Midsize SUV	Mean of coeff.	0.2490	0.1449
	Variance of coeff.	0.9772	0.3548
Large SUV	Mean of coeff.	–0.1295	0.1765
	Variance of coeff.	2.4334	0.9578
Compact pickup	Mean of coeff.	–1.3201	0.1507
	Variance of coeff.	1.3209	0.4484
Full-sized pickup	Mean of coeff.	–0.7908	0.1544
	Variance of coeff.	3.1370	0.8326
Minivan	Mean of coeff.	–0.5219	0.1441
	Variance of coeff.	2.6569	0.6334
Log likelihood at convergence		–6,297.81	

The Bernstein-von Mises theorem states that, under fairly benign conditions, the mean of the Bayesian posterior is a classical estimator that is asymptotically equivalent to the maximum likelihood estimator. Also, the variance of the posterior is the asymptotic variance of this estimator. See Train (2003) for an explanation with citations. Therefore, even though the model is estimated

Table 1.2. Mean and standard deviations of coefficients and *WTP*, implied by estimated parameters of model in preference space (Table 1.1)

Attribute	*Coefficient Mean*	*Coefficient Std. dev.*	*WTP Mean*	*WTP Std. dev.*
Price in $10,000's	–1.0499	0.8948		
Operating cost in $/month	–0.0421	0.0453	–0.0690	0.1130
Range in 100's of miles	0.5701	0.3576	0.9365	1.1077
Electric engine	–1.9453	1.2842	–3.1957	3.8605
Hybrid engine	0.8331	1.1870	1.3703	2.8062
High performance	0.2518	1.1829	0.4164	2.7611
Medium or high performance	0.5483	0.9581	0.9004	2.1917
Mini car	–3.0325	1.8852	–4.9773	5.8563
Small car	–1.3966	1.1439	–2.2938	3.1446
Large car	–0.4008	1.1439	–0.6598	2.5314
Small SUV	–0.8499	0.8386	–1.3952	2.1607
Midsize SUV	0.2490	0.9885	0.4060	2.1527
Large SUV	–0.1295	1.5599	–0.2120	3.3620
Compact pickup	–1.3201	1.1493	–2.1702	3.0874
Full-sized pickup	–0.7908	1.7712	–1.3032	3.9653
Minivan	–0.5219	1.6300	–0.8621	3.5859

by Bayesian procedures, the results can be interpreted from a purely classical perspective.

Table 1.1 gives estimation results for our model in preference space with uncorrelated coefficients. The estimate for each parameter is the mean of the 1,000 draws from the posterior, and the standard error of the estimate is the standard deviation of these draws. Presenting the results in this way facilitates interpretation by researchers who maintain a classical perspective: the estimates and standard errors can be interpreted the same as if they had been obtained by maximum likelihood procedures. The results can also, of course, be interpreted from a Bayesian perspective, with the mean and standard deviation of the draws providing summary information about the posterior. The log-likelihood value given at the bottom of table 1.1 is calculated in the classical way at the parameter estimates.[4]

For the log-normally distributed coefficients, the estimates in Table 1.1 are the mean and variance of the log of coefficient, which are difficult to interpret directly. Table 1.2 gives the estimated mean and standard deviation of the co-

[4]A Bayesian log-likelihood would be calculated by integrating the log-likelihood over the posterior or, as described by Sonnier *et al.* (2003), by integrating the inverse of the log-likelihood over the posterior and then taking the inverse.

efficients themselves, derived from the estimated parameters in Table 1.1. The estimates seem generally reasonable. Electric vehicles are considered worse than gas vehicles by the vast majority of the population, even if the two types of vehicles could cost the same and have the same range. The mean and standard deviation of the electric vehicle coefficient imply that 94 percent of the population place a negative value of electric vehicles relative to gas. Hybrid vehicles, on the other hand, are preferred to gas vehicles by most consumers, if they were to cost the same. The estimated mean and standard deviation imply that 75 percent have a positive coefficient for the hybrid dummy. Performance is valued at a decreasing rate, as expected. The average utility associated with moving from low to medium performance is greater than that for moving from medium to high performance (0.5483 and 0.2518 respectively.) The standard deviation of the range coefficient is much lower than of the two performance variables. This difference indicates that consumers are more similar in their desire for extra range than in their value for higher top speed and acceleration. The body type coefficients seem reasonable, with mid-sized cars and SUVs being preferred, on average, to either smaller or larger versions (holding price and operating cost constant). And pickups are valued less, on average, than comparably sized SUVs.

The estimated parameters in preference space imply distributions of *WTP*. A draw from the estimated distribution of *WTP* for an attribute is simulated by taking a draw from the estimated distribution of the attribute's coefficient and dividing by a draw from the estimated distribution of the price coefficient. Statistics for the distribution of *WTP* are obtained by taking numerous such draws and calculating the requisite statistic for these draws. The estimated mean and standard deviation of the *WTP* for each attribute is given in the final two columns of Table 1.2.

The most distinguishing aspect of the estimated distributions of *WTP* is the prevalence of large standard deviations. The standard deviation exceeds the mean for all *WTP*'s, and are more than twice the means for eight of the fifteen. These large standard deviations imply that a nontrivial share of people are willing to pay enormous amounts of money to obtain/avoid some attributes. For example, ten percent the population is estimated to have a *WTP* for range that exceeds 2. Given the units for price and range, a *WTP* over 2 means that the consumer is willing to pay more than $20,000 to have an extra 100 miles of range. Similarly, ten percent of the population is estimated to be willing to pay over $20,000 to move from low to medium performance. We return to this issue after presenting results of a model estimated in *WTP* space, where the distribution of *WTP* is estimated directly rather than derived from estimated coefficient distributions.

As stated above, a model with uncorrelated coefficients in preference space implies correlated *WTP*, with the correlation being the fairly arbitrary outcome

Table 1.3. Correlations between *WTP* for attributes, implied by estimated parameters of model in preference space (Table 1.1)

Attribute	*Op. cost*	*Range*	*Electric*	*Hybrid*	*Hi Perf*	*Med Perf*
Operating cost	1.0000	0.3687	–0.3627	0.2129	0.0679	0.1784
Range	0.3687	1.0000	–0.5029	0.2965	0.0958	0.2496
Electric	–0.3627	–0.5029	1.0000	–0.2855	–0.0929	–0.2411
Hybrid	0.2129	0.2965	–0.2855	1.0000	0.0584	0.1433
High perf	0.0679	0.0958	–0.0929	0.0584	1.0000	0.0439
Med–hi Perf	0.1784	0.2496	–0.2411	0.1433	0.0439	1.0000

(in the sense that the researcher does not specify it directly) of the estimated means and variances of the coefficients themselves. The correlation of *WTP* over attributes is given in Table 1.3. To conserve space, the correlation matrix does not contain the body types. As the table indicates, correlations among *WTP*'s are fairly large; researchers assuming uncorrelated coefficients might not be aware that they are implicitly assuming fairly large correlations among *WTP*'s.

4.2 Uncorrelated *WTP*'s in *WTP* space

We estimated a model with utility specified as in equation (2.4), where the coefficient of each non-price attribute is the product of the *WTP* for that attribute times the price coefficient. This model allows for random scale. If only scale varies, then the correlation between each pair of coefficients is one; correlations below one in coefficients imply that *WTP* varies as well as scale.

The price coefficient $-\lambda_n$ is given a log-normal distribution. The elements of ω_n (*WTP*'s) associated with operating cost, range, and the two performance variables are also specified to be log-normal, while the elements of ω_n associated with engine and body types are, instead, normal. The *WTP*'s are assumed to be uncorrelated over attributes. Note, of course, that when *WTP* for an attribute is normally distributed and the price coefficient is log-normal, the coefficient of the attribute is not normal (as in the previous model). Also, as stated above, uncorrelated *WTP* implies correlated coefficients (unlike the previous model), due to the common influence of the price coefficient on each other coefficient. The current model differs from the previous one in both of these ways.

Table 1.4 gives the estimation results. The log-likelihood is considerably lower than that for the model in Table 1.1. However, the distributions of *WTP* seem more reasonable. Comparing Table 1.5 with Table 1.2, the main dis-

Table 1.4. Model in *WTP* Space with Uncorrelated *WTP*'s

Attribute	*Parameter*	*Estimate*	*St. error*
Price in $10,000's	Mean of ln(–coeff.)	–0.0498	0.0602
	Variance of ln(–coeff.)	0.9014	0.1234
Operating cost in $/month	Mean of ln(*WTP*)	–3.4106	0.1100
	Variance of ln(*WTP*)	0.7847	0.1530
Range in 100's of miles	Mean of ln(*WTP*)	–0.4045	0.1286
	Variance of ln(*WTP*)	0.2706	0.0939
Electric engine	Mean of *WTP*	–2.5353	0.2369
	Variance of *WTP*	1.9828	0.4443
Hybrid engine	Mean of *WTP*	0.8738	0.1090
	Variance of *WTP*	2.1181	0.2745
High performance	Mean of ln(*WTP*)	–1.8854	0.2840
	Variance of ln(*WTP*)	1.7172	0.5898
Medium or high performance	Mean of ln(*WTP*)	–1.7380	0.2917
	Variance of ln(*WTP*)	2.4701	0.7310
Mini car	Mean of *WTP*	–3.4645	0.1894
	Variance of *WTP*	6.5767	1.3889
Small car	Mean of *WTP*	–1.5992	0.1451
	Variance of *WTP*	1.7010	0.5337
Large car	Mean of *WTP*	–0.6148	0.1716
	Variance of *WTP*	1.9353	0.6750
Small SUV	Mean of *WTP*	–1.0671	0.1287
	Variance of *WTP*	0.8203	0.5776
Midsize SUV	Mean of *WTP*	0.2173	0.1611
	Variance of *WTP*	1.8544	0.4389
Large SUV	Mean of *WTP*	–0.7559	0.2923
	Variance of *WTP*	8.2263	2.3072
Compact pickup	Mean of *WTP*	–1.4752	0.1398
	Variance of *WTP*	1.2675	0.5266
Full-sized pickup	Mean of *WTP*	–1.1230	0.1843
	Variance of *WTP*	5.7762	1.2558
Minivan	Mean of *WTP*	–0.7406	0.1827
	Variance of *WTP*	3.9847	0.9252
Log likelihood at convergence		–6,362.13	

tinction is that the means and especially the standard deviations of *WTP*'s are smaller for the model in *WTP* space than the model in preference space. This difference means that there is a smaller share with unreasonably large *WTP*'s. For example, the model in *WTP* space implies that 1.7 percent are estimated to be willing to pay more than $20,000 for 100 miles of extra range, while, as stated above, the model in preference space implies over 10 percent. Sim-

Table 1.5. Mean and standard of preference coefficients and *WTP*, implied by estimated parameters of model in *WTP* space (Table 1.4)

Attribute	*Coefficient Mean*	*Coefficient Std. dev.*	*WTP Mean*	*WTP Std. dev.*
Price in $10,000's	–1.4934	1.8123		
Operating cost in $/month	–0.0732	0.1616	–0.0489	0.0531
Range in 100's of miles	1.1406	1.7027	0.7636	0.4257
Electric engine	–3.7870	5.6565	–2.5353	1.4081
Hybrid engine	1.3053	3.7585	0.8738	1.4554
High performance	0.5335	1.7974	0.3584	0.7563
Medium or high performance	0.8951	4.5679	0.6047	1.9542
Mini car	–5.1712	8.6579	–3.4645	2.5645
Small car	–2.3849	4.1887	–1.5992	1.3042
Large car	–0.9180	3.4259	–0.6148	1.3912
Small SUV	–1.5914	2.8561	–1.0671	0.9057
Midsize SUV	0.3151	3.1997	0.2173	1.3618
Large SUV	–1.1336	6.8725	–0.7559	2.8682
Compact pickup	–2.2029	3.7700	–1.4752	1.1258
Full-sized pickup	–1.6858	5.9893	–1.1230	2.4034
Minivan	–1.1161	4.8729	–0.7406	1.9962

ilarly, but not as dramatically, the share who are willing to pay over $20,000 to move from low to medium performance is estimated to be 6 percent in the model in *WTP* space, which is less than the 10 percent implied by the model in preference space.

In conclusion, for both preference coefficients and *WTP* values, the indirect way of estimating the distributions results in larger means and standard deviations than when the distributions are estimated directly. As discussed above, the larger standard deviations in *WTP* imply implausible shares of the population willing to pay large amounts for an attribute. The meaning of larger means and standard deviations of coefficients is not clear.

Table 1.6 gives the correlations between coefficients that are implied by the estimated distributions of *WTP* and the price coefficient. The correlations are fairly high, due to the fact that each *WTP* is multiplied by the common price coefficient. These high correlations suggest that models with uncorrelated coefficients in preference space are incompatible empirically (as well as theoretically, of course) with independent *WTP*'s and price coefficient. Researchers, when considering independence over attributes, must be careful in distinguishing whether they want to assume that *WTP*'s are independent or

Table 1.6. Correlations between preference coefficients of attributes, implied by estimated parameters of model in *WTP* space (Table 1.4)

Attribute	*Price*	*Op. cost*	*Range*	*Electric*	*Hybrid*	*Hi Perf*	*Med Perf*
Price	1.0000	0.5526	0.8117	–0.8080	0.4157	0.3570	0.2242
Op cost	0.5526	1.0000	0.4481	-0.4456	0.2322	0.2087	0.1281
Range	0.8117	0.4481	1.0000	-0.6532	0.3375	0.2895	0.1796
Electric	–0.8080	–0.4456	–0.6532	1.0000	–0.3343	–0.2853	–0.1857
Hybrid	0.4157	0.2322	0.3375	-0.3343	1.0000	0.1439	0.0945
Hi perf	0.3570	0.2087	0.2895	-0.2853	0.1439	1.0000	0.0794
Med/Hi Perf	0.2242	0.1281	0.1796	–0.1857	0.0945	0.0794	1.0000

that utility coefficients are independent, since independence of one implies non-independence of the other.

4.3 Correlated coefficients and *WTP*

In general, neither coefficients nor *WTP*'s are independent. We estimated a model in preference space with correlated coefficients and a model in *WTP* space with correlated *WTP*'s. The model in preference space incorporates random scale, since it allows correlation between all coefficients. The two models (in preference space and *WTP* space) are therefore the same in allowing for random scale and differ only in the distributional assumptions for coefficients and *WTP*. Both models assume a log-normal price coefficient. The model in preference space assumes normal and log-normal non-price coefficients, which implies that *WTP*'s are distributed as the ratio of a normal or log-normal to a log-normal. The model in *WTP* space assumes normal and log-normal *WTP*'s, which implies coefficients that are the product of a log-normal with a normal or log-normal.

To save space, we do not present the estimates of these model; they are available to interested readers upon request. The results are consistent with those obtained above, namely: (1) the model in preference space obtains a higher log-likelihood, but (2) the estimated distribution of *WTP* is more reasonable (with smaller means and variances) for the model in *WTP* space. In addition, several conclusions can be drawn concerning correlations:

- The hypothesis that coefficients in preference space are uncorrelated can be rejected. The model in preference space attains a log-likelihood of -6,178.12 with correlated coefficients, compared to -6,297.81 for the model given in Table 1.1 with uncorrelated coefficients. The likelihood ratio test statistic is therefore 239.4 for the hypothesis that all 120 covariances are zero, which is greater than the 99-percentile value of the chi-square with 120 degrees of freedom.

- The estimated correlations among coefficients are generally small or moderate in size. 47 of the 160 correlations are below 0.1 in magnitude, and only 12 are above .4 in magnitude.

- The model in *WTP* space attains a log-likelihood of -6,228.31 when the *WTP*'s and price coefficient are all allowed to be correlated and -6,362.13 when they are constrained to be uncorrelated. The hypothesis of no correlation can be rejected.

- The estimated correlations between *WTP*'s for the model in *WTP* space are generally small or moderate, similar to the estimated correlations between coefficients for the model in preference space.

- The correlations among coefficients that are derived from the model in *WTP* space are considerably larger in magnitude than those estimated directly in the model in preference space. Similarly, the correlations among *WTP*'s that are derived from the model in preference space are considerably larger than those estimated directly in the model in *WTP* space. These findings are similar to those given above for variances, i.e., that larger variances in coefficients are obtained when they are estimated indirectly instead of directly, and larger variances in *WTP*'s are obtained when estimated indirectly than directly. It seems that the process of combining estimated distributions (dividing a normal by a log-normal for *WTP* or multiplying a normal by a log-normal for a coefficient) tends to inflate the estimated variances and covariances.

Sonnier *et al.* (2003) estimated models in preference space and *WTP* space, using the terms "linear models" and "nonlinear models" instead of our terminology to denote that the random customer-level parameters enter utility linearly in the former and nonlinearly in the later. Their results are consistent with our main conclusions, in that they obtained better within-sample fit for their model in preference space but more reasonable *WTP* distributions for their model in *WTP* space. However, their results differ from ours in one regard. They performed out-of-sample analysis and concluded that their model in *WTP* space fits better out-of-sample, even though it fits worse in-sample.

To examine this issue, we divided our sampled respondents into two equal-sized sub-samples, estimated each model on one sub-sample, and evaluated the log-likelihood of the estimated models on the other sub-sample. In each comparison (estimation on first half with evaluation on the second half, and estimation on the second half with evaluation on the first half), the model in preference space obtained a higher log-likelihood than the model in *WTP* space on the out-of-estimation sub-sample.

Our results therefore differ in this regard from those of Sonnier *et al.* (2003). The difference can perhaps be explained by the fact that we used a somewhat

different method to evaluate out-of-sample fit than they did. We estimated on half the respondents using all of their choice situations and then calculated the log-likelihood for all the choice situations for the other half of the respondents, while they estimated the model on all but one choice situation for each respondent and then calculated the log-likelihood for this one "hold-out" choice situation for each respondent.

However, there is no reason to expect the same results in different settings, since the answer to the question "Which distributions fit better?" is necessarily situation-dependent. The purpose of the explorations is to focus our attention on the relation between distributions of coefficients and distributions of *WTP*, rather than to attempt to identify the appropriate distributions to use in all situations.

5. Conclusions

This paper examines consumer choice among alternative-fueled vehicles, including gas, electric, and gas/electric hybrids. The empirical results indicate that the vast majority of consumers would need to be compensated through a lower price (i.e., have a negative willingness to pay) for electric vehicles relative to gas vehicles, even if operating cost, performance, and range were the same. In contrast, most consumers are willing to pay extra for a hybrid relative to a gas vehicle with the same non-price attributes. This result is consistent with the market experience in the U.S. The few electric cars that have been introduced in the U.S. have fared poorly in the market, and models are being discontinued. In contrast, the initial offerings of hybrids have been relatively popular, and more models, such as hybrid SUVs, are being launched.

Discrete choice models were estimated with convenient distributions (normal and log-normal) in preference space and in willingness-to-pay *WTP* space. The models in preference space were found to fit the data better, both within-sample and out-of-sample, than the models in *WTP* space. However, the models in *WTP* space provided more reasonable distributions of *WTP*, with fewer consumers having untenably large *WTP*'s, than the models in preference space. This comparison implies that research is needed to identify distributions that fit the data better when applied in *WTP* space and/or provide more reasonable distributions of *WTP* when applied in preference space.

Chapter 2

USING CLASSICAL SIMULATION-BASED ESTIMATORS TO ESTIMATE INDIVIDUAL *WTP* VALUES

A Mixed Logit Case Study of Commuters

William H. Greene
Department of Economics, Stern School of Business,
New York University, New York, USA
wgreene@stern.nyu.edu

David A. Hensher
Institute of Transport Studies, Faculty of Economics and Business,
The University of Sydney, NSW 2006 Australia
Davidh@its.usyd.edu.au

John M. Rose
Institute of Transport Studies, Faculty of Economics and Business,
The University of Sydney, NSW 2006 Australia
Johnr@its.usyd.edu.au

Abstract A number of papers have recently contrasted classical inference estimation methods for logit models with Bayesian methods. It has been argued that two particularly appealing features of the Bayesian approach are its relative simplicity in estimation, and its ability to derive, individual-specific willingness to pay (*WTP*) measures that are less problematic than the classical approaches in terms of extreme values and unexpected signs. This paper challenges this claim by deriving both population derived *WTP* measures and individual-specific values based on the classical mixed logit model, establishing the extent of unacceptable valuations. Our aim is not to estimate Bayesian contrasts per se but to show that the classical inference approach is likewise straightforward – indeed the individual-specific estimates are a by-product of the parameter estimation process. We also reveal the benefits of calculating *WTP* measures from ratios of individual param-

eters which are behaviourally more appealing approximations to the true values of each individual, in contrast to draws from population distributions that run the risk of allocating two parameters that are poorly juxtaposed in a relative sense, resulting in extreme value estimates. Our results suggest that while extreme values and unexpected signs cannot be ruled out (nor can they in the Bayesian framework), the overall superiority of the Bayesian method appears overstated. Both approaches have merit.

Keywords: *WTP*, mixed logit simulator.

1. Introduction

Discrete choice models are a primary source of estimates of willingness to pay (henceforth *WTP*) for specific attributes such as travel time savings. As choice modelling matures into a suite of models with increasing degrees of behavioural richness, typified by the progression from multinomial logit, nested logit (NL), cross-correlated NL and mixed logit, analysts are increasingly exploring the deep parameterisation of *WTP*, using classical inference approaches, as a way of accommodating the heterogeneity of trade-offs in a sampled population. Such distributions of *WTP* can be derived from a set of moments portraying the population-level profile of a distribution (i.e., a mean and standard deviation with a specific analytical distribution (e.g., normal, log-normal, triangular, constrained triangular, uniform etc.)) or from parameters that are unique to each sampled individual. For an application of the former, see Sillano and Ortuzar (2004). In this study, we consider individual-specific parameters that are derived from conditional distributions in which known choices are taken into account in the spirit of the Bayesian posterior distributions.

With a growing interest in the Bayesian approach and claims that it is a more attractive paradigm than classical inference methods, the objective of this paper is to show how easy it is to obtain the equivalent information on individual parameters within the classical inference framework, and to derive such rich indicators of *WTP* distributions. We contrast the 'Bayesian-like' estimates with the *WTP* derived from the population moments, the latter more commonly associated with classical inference. From a policy perspective the empirical evidence is very revealing, suggesting that the aggregation inherent in the population approach in which parameters used to derive *WTP* estimates are random in the numerator and the denominator, appears (on our initial evidence) to explode the mean *WTP* due to the presence of a percentage of extreme values. This is in contrast to that associated with the *WTP* derived from the individual-specific parameterisation that has behaviourally sensible distributions.

The paper is organised as follows. We begin with a brief overview of the Bayesian approach specifically to gain an appreciation of its contribution then

follow with a summary of the mixed logit model that will deliver the parameters in a classical inference setting. The data setting is then presented (a stated mode choice experiment for commuting trips in Sydney in 2003), followed by the findings and implications for deriving *WTP* from alternative interpretations of the mixed logit outputs.

2. The Bayesian approach

Bayesian methods are often promoted as behaviourally different from, and preferable to classical estimation methods currently used in estimation of advanced discrete choice models such as mixed logit. Brownstone (2001), Chen et al. (2000), Geweke (1999) and Train (2001) provide useful overviews of the Bayesian perspective. Use of information on priors (as structural parameters) and posterior individual-specific parameter estimates from conditional utility functions are included as information to capture sources of heterogeneity.

The key difference between Bayesian and classical approaches is that Bayesians treat the nature of the randomness differently. In the classical view, the randomness is part of the model; it is the heterogeneity of the taste parameters, across individuals. In the Bayesian approach, the randomness 'represents' the uncertainty in the mind of the analyst (conjugate priors notwithstanding). Therefore, from the classical viewpoint, there is a 'true' distribution of the parameters across individuals. From the Bayesian viewpoint, in principle, there could be two analysts with different, both legitimate, but substantially different priors, who therefore could obtain very different, albeit both legitimate, posteriors.

Prior knowledge about parameters, $\boldsymbol{\theta}$, is gathered in a *prior* distribution, $\pi(\boldsymbol{\theta})$. The sampling distribution, or likelihood function, is given by $f(\mathbf{X}|\boldsymbol{\theta})$ where $\mathbf{X}$ contains all the sample data in the study. After observing the data, the information about $\boldsymbol{\theta}$ is given by the *posterior* distribution which is derived using Bayes Theorem;

$$\Pr(\boldsymbol{\theta}|x) = \frac{f(x|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(x|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \tag{2.1}$$

We note for the purposes explained below, that the posterior density is functionally equivalent to the conditional distribution of the parameters given the data. All inference is based on this posterior distribution. The usual Bayes estimator is the mean of the posterior distribution, and Bayesian confidence bands are typically given by the narrowest region of the posterior distribution with the specified coverage probability. Bayesian confidence regions are interpreted as fixed regions containing the random parameter $\boldsymbol{\theta}$ with the specified coverage probability (i.e., the 'highest posterior density' interval). This is different from the classical confidence region, which is a region with random endpoints that contain the true value $\boldsymbol{\theta}$ with the specified probability over independent re-

peated realisations of the data (Brownstone, 2001). Classical inference therefore depends on the distribution of unobserved realisations of the data, whereas Bayesian inference conditions on the observed data. Bayesian posterior inference is also exact and does not rely on asymptotic approximations to a true sampling distribution.

The Bayesian approach requires the a priori specification of prior distributions for all of the model parameters. In cases where this prior is summarising the results of previous empirical research, specifying the prior distribution is a useful exercise for quantifying previous knowledge (such as the alternative currently chosen). In most circumstances, however, the prior distribution cannot be fully based on previous empirical work. The resulting specification of prior distributions based on the analyst's subjective beliefs is the most controversial part of Bayesian methodology. Poirier (1988) argues that the subjective Bayesian approach is the only approach consistent with the usual rational actor model to explain individuals' choices under uncertainty. More importantly, the requirement to specify a prior distribution enforces intellectual rigour on Bayesian practitioners.[1] All empirical work is guided by prior knowledge and the subjective reasons for excluding some variables and observations are usually only implicit in the classical framework. The simplicity of the formula defining the posterior distribution hides some difficult computational problems, explained in Brownstone (2001).[2]

Huber and Train (2001), in particular, have explored the empirical similarities and differences between hierarchical Bayes and classical estimators in the context of estimating reliable individual-level parameters from sampled population data as a basis of market segmentation. The ability to combine information about the aggregate distributions of preferences with individuals' choices to derive conditional estimates of the individual parameters is very attractive. They conclude, however, that the empirical results are virtually equivalent conditional estimates of marginal utilities of attributes for individuals.[3] What this

[1]Bayesians are encouraged to carry out sensitivity analysis (although this is not often undertaken) across other reasonable prior distributions to demonstrate that their empirical results are not just reflections of their prior beliefs (Brownstone 2001).

[2]Computing the posterior distribution typically requires integrating over β and this can be difficult for the number of parameters frequently encountered in choice modelling. Until recently Bayesians solved this problem by working with conjugate families. These are a family of prior distributions linked to a family of likelihood functions where the posterior distribution is in the same family as the prior distribution. For example, the Beta family is a conjugate prior for the binomial with fixed number of trials. Koop and Poirier (1993) have developed and applied a conjugate prior for the conditional (and multinomial) logit model, but there do not appear to be tractable conjugate priors for other GEV discrete choice models. Recent applications have circumvented these difficulties through the use of Gibbs Sampling and Markov chain Monte Carlo Methods.

[3]Allenby and Rossi (1999) have carried out an extensive Bayesian analysis of discrete brand choice and discussed a number of methodological issues relating to the estimation of individual level preferences. In comparison of the Bayesian and classical methods, they state the simulation based classical methods are

debate has achieved in particular is to show classical estimation choice modellers that there is indeed more information in their estimation procedure that enables one to improve on the behavioural explanation within sample.[4] Recent developments in classical inference methods that are rich in deep parameters enable the analyst to obtain information that is Bayesian-like. The mixed-logit model is one choice specification with this capability which we use as the model for a classical inference.

3. The Mixed Logit Model

To illustrate the "mixed logit"[5] model we start by assuming that a sampled individual ($q = 1, \ldots, Q$) faces a choice among J alternatives in each of T choice situations. Individual q is assumed to consider the full set of offered alternatives in choice situation t and to choose the alternative with the highest utility. The utility associated with each alternative j as evaluated by each individual q in choice situation t, is represented in a discrete choice model by a utility expression of the general form:

$$U_{jtq} = \boldsymbol{\beta}_q' \mathbf{x}_{jtq} + \varepsilon_{jtq} \tag{3.1}$$

where $\mathbf{x}_{jtq}$ is the full vector of explanatory variables, including attributes of the alternatives, socio-economic characteristics of the individual and descriptors of the decision context and choice task itself in choice situation t. The complexity of the choice task in stated choice experiments as defined by number of choice situations, number of alternatives, attribute ranges, data collection method, etc., can be included to condition specific parameters associated with attributes of alternatives. The components $\boldsymbol{\beta}_q$ and ε_{jtq} are not observed by the analyst and are treated as stochastic influences. Note that the first of these, unlike its counterpart in other models, is assumed to vary across individuals.

---

likely to be extremely cumbersome and are approximate whereas the Bayesian methods are much simpler and are exact in addition. As to whether the Bayesian estimates are exact while sampling theory estimates are approximate, one must keep in mind what is being characterised by this statement. The two estimators are not competing for measuring the same population quantity with alternative tools. In the Bayesian approach, the 'exact' computation is of the analysts posterior belief about the distribution of the parameter (conditioned, one might note on a conjugate prior virtually never formulated based on prior experience), not an exact copy of some now revealed population parameter. The sampling theory 'estimate' is of an underlying 'truth' also measured with the uncertainty of sampling variability. The virtue of one over the other is not established on any but methodological grounds – no objective, numerical comparison is provided by any of the preceding or the received literature.

[4]Within-sample priors such as the actual choice can help a great deal. When applying a model out-of-sample then Bayesians need some subjective priors. Hensher and Jones (2004) discuss the problems in taking individual parameters from either a Bayesian or classical-inference setting and mapping them to observations in a hold-out sample. This is an empirical question. We were unable to find (in our data on firm failures) any exogenous criteria that delivered a predictable mapping. Thus the preference for using population moments reported from the data used in model estimation.

[5]It is also referred to in various literatures as random parameter logit (RPL), mixed multinomial logit (MMNL), kernel logit, hybrid logit and error components logit.

Individual heterogeneity is introduced into the utility function through $\boldsymbol{\beta}_q$. Thus, we write

$$\begin{aligned} \boldsymbol{\beta}_q &= \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_q + \boldsymbol{\Gamma}\mathbf{v}_q \\ &= \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_q + \boldsymbol{\eta}_q \end{aligned} \tag{3.2}$$

or $\beta_{qk} = \beta_k + \boldsymbol{\delta}_k'\mathbf{z}_q + \eta_{qk}$ where β_{qk} is a random term whose distribution over individuals depends in general on underlying parameters, $\mathbf{z}_q$ is observed data and $\boldsymbol{\eta}_q$ denotes a vector of K random components in the set of utility functions in addition to the J random elements in $\boldsymbol{\eta}_{tq}$. Since β_q may contain alternative-specific constants, η_{qk} may also vary across choices and, in addition, may thus induce correlation across choices. Note that $\boldsymbol{\beta}_q$ and its components are structural parameters ($\boldsymbol{\beta}$,$\boldsymbol{\Delta}$,$\boldsymbol{\Gamma}$) and choice situation invariant characteristics of the person, $\mathbf{z}_q$. It does not vary across choice situations or across choices (save for the extent that components of $\mathbf{x}_{jtq}$ are choice-specific).

The mixed logit class of models assumes a general distribution for β_{qk} and an IID extreme value type 1 distribution for ε_{jtq}. That is, β_{qk} can take on different distributional forms such as normal, lognormal, uniform or triangular. Denote the marginal joint density of $[\beta_{q1}, \beta_{q2}, \ldots, \beta_{qk}]$ by $f(\boldsymbol{\beta}_q|\boldsymbol{\Omega}, \mathbf{z}_q)$ where the elements of $\boldsymbol{\Omega}$ are the underlying parameters of the distribution of $\boldsymbol{\beta}_q, (\boldsymbol{\beta}, \boldsymbol{\Delta}, \boldsymbol{\Gamma}\,)$ and $\mathbf{z}_q$ is observed data-specific to the individual, such as socio demographic characteristics. For a given value of $\boldsymbol{\beta}_q$, the *conditional* probability for choice j in choice situation t is multinomial logit, since the remaining error term is IID extreme value:

$$L_{jtq}(\boldsymbol{\beta}_q|\mathbf{X}_{tq}, \mathbf{z}_q, \boldsymbol{\eta}_q) = \frac{\exp(\boldsymbol{\beta}_q'\mathbf{z}_{jtq})}{\sum_j \exp(\boldsymbol{\beta}_q'\mathbf{z}_{jtq})} \tag{3.3}$$

Equation 3.3 is the simple multinomial logit model, but with the proviso that, for each sampled individual, we have additional information defined by $\boldsymbol{\beta}_q$. This is where the use of the word '*conditional*' applies – the probability is conditional on $\boldsymbol{\eta}_q$, that is, on $\mathbf{v}_q$, and $\mathbf{z}_q$. This additional information influences the choice outcome.

The *unconditional* choice probability is the expected value of the logit probability over all the possible values of $\boldsymbol{\beta}_q$, that is, integrated over these values, weighted by the density of $\boldsymbol{\beta}_q$. From equation 3.2, we see that this probability density is induced by the random component in the model for $\boldsymbol{\beta}_q$, $\mathbf{v}_q$ (Hensher and Greene, 2003). Thus, the unconditional choice probability is:

$$\begin{aligned} \Pr(\mathbf{X}_{tq}, \mathbf{z}_q, \boldsymbol{\Omega}) &= \int_{\boldsymbol{\beta}_q} L_{jtq}(\boldsymbol{\beta}_q|\mathbf{X}_{tq}, \mathbf{z}_q, \boldsymbol{\eta}_q) f(\boldsymbol{\beta}_q|\mathbf{z}_q, \boldsymbol{\Omega}) d\boldsymbol{\beta}_q \\ &= \int_{\boldsymbol{\eta}_q} L_{jtq}(\boldsymbol{\beta}_q|\mathbf{X}_{tq}, \mathbf{z}_q, \boldsymbol{\eta}_q) f(\boldsymbol{\eta}_q|\mathbf{z}_q, \boldsymbol{\Omega}) d\boldsymbol{\eta}_q \end{aligned} \tag{3.4}$$

where, once again, $\boldsymbol{\beta}_q = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_q + \boldsymbol{\eta}_q$. Thus, the unconditional probability that individual q will choose alternative j given the specific characteristics of

their choice set and the underlying model parameters is equal to the expected value of the conditional probability as it ranges over the possible values of $\boldsymbol{\beta}_q$. The random variation in $\boldsymbol{\beta}_q$ is induced by the random vector $\boldsymbol{\eta}_q$, hence that is the variable of integration in (3.4). The log likelihood function for estimation of the structural parameters is built up from these unconditional probabilities, aggregated for individual q over the T choice situations and the choices actually made:

$$\ln L = \sum_{q=1}^{Q} \ln \int_{\boldsymbol{\beta}_q} \prod_{t=1}^{T} L_{jtq}(\boldsymbol{\beta}_q|\mathbf{X}_{tq}, \mathbf{z}_q, \boldsymbol{\eta}_q) f(\boldsymbol{\beta}_q|\mathbf{z}_q, \boldsymbol{\Omega}) d\boldsymbol{\beta}_q \tag{3.5}$$

Details on estimation of the parameters of the mixed logit model by maximum simulated likelihood may be found in Train (2003) and in the manual for NLOGIT v.3.

One can construct estimates of individual-specific preferences by deriving the individual's conditional distribution based (within-sample) on their own choices (i.e., prior knowledge). For convenience, let $\mathbf{Y}_q$ denote the observed information on choices by individual q, and let $\mathbf{X}_q$ denote all elements of $\mathbf{x}_{jtq}$ for all j and t. Using Bayes' Rule, we find the conditional density for the random parameters,

$$\mathbf{H}(\boldsymbol{\beta}_q|\mathbf{Y}_q, \mathbf{X}_q, \mathbf{z}_q, \boldsymbol{\Omega}) = \frac{f(\mathbf{Y}_q|\boldsymbol{\beta}_q, \mathbf{X}_q, \mathbf{z}_q, \boldsymbol{\Omega}) \Pr(\boldsymbol{\beta}_q|\mathbf{z}_q, \boldsymbol{\Omega})}{f(\mathbf{Y}_q|\mathbf{X}_q, \mathbf{z}_q, \boldsymbol{\Omega})}. \tag{3.6}$$

The left hand side gives the density of the random parameter vector given the underlying parameters and the data on individual q. In the numerator of the right hand side, the first term gives the probability in the conditional likelihood—this is in (3.4). The second term gives the marginal probability density for the random $\boldsymbol{\beta}_q$ given in (3.2) with the assumed distribution of $\boldsymbol{\eta}_q$. The denominator is the unconditional choice probability for the individual—this is given by (3.4). Note that the denominator in (3.6) is the integral of the numerator, as given in (3.4). This result will be used to estimate the person-specific parameters, utilities, or choice probabilities as a function of the underlying parameters of the distribution of the random parameters. Estimation of the individual specific value of $\boldsymbol{\beta}_q$ is then done by computing an estimate of the mean from this distribution. Note that this conditional mean is a direct analog to its counterpart in the Bayesian framework, the posterior mean in (3.1). We return to this type of computation in the next section.

The choice probability in the denominator of (3.6) generally cannot be calculated exactly because the integral in (3.4) will not have a closed form. The integral is approximated by simulation. For a given value of the parameters, $\boldsymbol{\Omega}$, and the observed data, $\mathbf{z}_q$, a value of $\boldsymbol{\beta}_q$ is drawn from its distribution based on (3.2). Using this draw, the logit formula (3.3) for $L_{jtq}(\boldsymbol{\beta})$ is calculated. This

process is repeated for many draws, and the mean of the resulting $L_{jtq}(\boldsymbol{\beta})$'s is taken as the approximate choice probability giving the simulated probability in equation (3.7),

$$\hat{f}(\mathbf{Y}_q|\mathbf{X}_q, \mathbf{z}_q, \boldsymbol{\Omega}) = \frac{1}{R}\sum_{r=1}^{R} L_{jq}(\boldsymbol{\beta}_{qr}, \mathbf{X}_q, \mathbf{z}_q, \boldsymbol{\Omega}, \boldsymbol{\eta}_{qr}), \tag{3.7}$$

where R is the number of replications (i.e., draws of $\boldsymbol{\beta}_{qr}$), $\boldsymbol{\beta}_{qr}$ is the r^{th} draw, and the right hand side is the simulated probability that an individual chooses alternative j.[6] In our application of this model, we will use the structure in (3.2), $\boldsymbol{\beta}_q = \boldsymbol{\beta}_q + \boldsymbol{\Delta}_q\mathbf{z}_q + \boldsymbol{\Gamma}\mathbf{v}_q$ where the fixed underlying parameters are $\boldsymbol{\Omega} = (\boldsymbol{\beta}, \boldsymbol{\Delta}, \boldsymbol{\Gamma})$, $\boldsymbol{\beta}$ is the fixed mean of the distribution, $\mathbf{v}_q$ is a set of person-specific influences (also referred to as 'heterogeneity around the mean'), $\boldsymbol{\Delta}$ is a matrix of parameters, $\mathbf{v}_q$ is a vector of uncorrelated random variables with known variances on the diagonal of a variance-covariance matrix, $\boldsymbol{\Sigma}$, and is a lower triangular matrix which, because $Var[\boldsymbol{\beta}_q] = \boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}'$, allows free variances and correlations of the parameters. Thus, a 'draw' from the distribution of $\boldsymbol{\beta}_q$ consists of a 'draw' from the distribution of $\mathbf{v}_q$ which is then used to compute $\boldsymbol{\beta}_q$ as shown above.

4. An Empirical Example

The empirical study uses a mode choice data set of 223 commuting trips by a sample of residents of the north-west sector of the Sydney metropolitan area interviewed in 2003. The centerpiece of the data collection activity is a stated choice experiment in which each sampled individual reviewed T = 10 mode choice scenarios and chose their preferred alternative.[7] The main mode alternatives are car, existing bus, existing train, existing busway, new light rail (LR), new heavy rail (HR) and new busway (BW). Each public transport mode has an access and an egress component (see below). The data was collected using a Computer Aided Survey Instrument (CAPI), with all data being automatically captured in the CAPI into a data base formatted for immediate choice model estimation. More details are given in Hensher and Rose (2003).

Table 2.1 shows the descriptive statistics for the sample. 50.4% of the sample is male and the mean age of the sample is 43.1 years. The average household size is 3.78 and the annual personal income is $64,100. Table 2.1 shows

[6] By construction, this is a consistent estimator of P_j for any R; its variance decreases as R increases. It is strictly positive for any R, so that $\ln(SP_j)$ is always defined in a log-likelihood function. It is smooth (i.e., twice differentiable) in parameters and variables, which helps in the numerical search for the maximum of the likelihood function. The simulated probabilities sum to one over alternatives. Train (1998) provides further commentary on this.

[7] Readers unfamiliar with stated choice methods may refer to Louviere et al. (2000).

Table 2.1. Descriptive Statistics for Commuters

	N	*Mean/Percent*	*Std. Dev.*	*Min.*	*Max.*
Age	223	43.10	12.50	24	70
Hours worked per week	223	37.60	14.60	0	70
Annual Personal Income ($000's)	223	64.10	41.80	0	140
Household size	223	3.78	2.30	1	8
No. of children in household	223	1.05	1.09	0	4
Gender (male =1)	223	50.40		0	1

Table 2.2. Percentage of Commuters Who Had a Motor Vehicle Available for the Trip

Car available	*Frequency*	*Percent*
yes	199	89.24
no	24	10.76
Total	223	100

the descriptive statistics. The mean age is 43.1 years with an average annual gross personal income of $64,100. Table 2.2 reports that 89.24 percent of the sample had a car available for the surveyed trip. Of the 223 respondents interviewed as part of the commuter sample, 199 had access to a motor vehicle for the surveyed trip. This represents 89.24 percent of the sample.

4.1 Stated Choice Experimental Design

The experimental design has 47 variables (46 in four levels and one in six levels for the blocks) in 60 runs; yielding six blocks of ten scenarios each. The D-optimal design is almost orthogonal with maximum correlations between 0.06.[8] The design allows the estimation of all alternative-specific main effects. Within each block the runs have been randomized to control for order effect.

[8]The design minimises the correlations between attributes and maximises the amount of information captured by each choice task. In designing choice experiments, knowledge of the parameters or at least some priors (like signs) for each attribute provides a useful input. Insights from past studies determined their approximate values. A preferred choice experiment design is one that maximises the determinant of the covariance matrix, which is itself a function of the estimated parameters. The design developed herein takes into account the expected signs of the parameters (e.g., negative for the time and cost attributes). We found that in so doing, the search eliminates dominant alternatives, which is sensible since dominant alternatives do not give any useful information if we know the signs of the parameters. The method used finds the D-optimality plan very quickly. Carlsson and Martinsson (2003) have recently shown, using Monte Carlo simulation, that D-optimal designs, like orthogonal designs, produce unbiased parameter estimates but that the former have lower mean squared errors.

Table 2.3. Trip Attributes in Stated Choice Design

For existing public transport modes	*For new public transport modes*	*For the existing car mode*
Fare (one-way)	Fare (one-way)	Running Cost
In-vehicle travel time	In-vehicle travel time	In-vehicle Travel time
Waiting time	Waiting time	Toll Cost (One way)
Access Mode: Walk time	Transfer waiting time	Daily Parking Cost
Car time	Access Mode: Walk time	Egress time
Bus time	Car time	
Bus fare	Bus time	
Egress time	Access Mode Fare (one-way)	
	Bus fare	
	Egress time	

There are different task configurations: with/without car, inter/intra regional, new LR and New HR versus new HR and new BW. A maximum number of complete designs have to be filled within each configuration. This is achieved in the field as follows: if the first respondent has a car on an intra regional trip with new LR & HR he is randomly assigned to a block (e.g., block three). If the second respondent is in the exact same configuration she sees the next immediate block (e.g., block four) otherwise she sees another randomly assigned block in one of the other configurations. Once all blocks in a configuration have been viewed, we randomly start at with another block. The trip attributes associated with each mode are summarised in Table 2.3.

Each design attribute has four levels. These were chosen as the following variations around the base level: -25%, 0%, +25%, +50%. The base times and costs used for new modes are shown in Table 2.4 where the locations are rail or busway stations. An example of a stated choice screen is shown as Figure 1.

Table 2.4. Base times and costs for new public transport modes

	Dollars $	*Busway* *Min.*	*Heavy Rail* *Min.*	*Light Rail* *Min.*
Mungerie Road	1.8	33	22	3
Burns Road	1.0	27	18	27
Norwest Business Park	1.0	22.5	15	22.5
Hill Centre	1.0	18	12	18
Norwest Business Park	1.0	22.5	9	13.5
Norwest Business Park	1.0	7.5	5	7.5

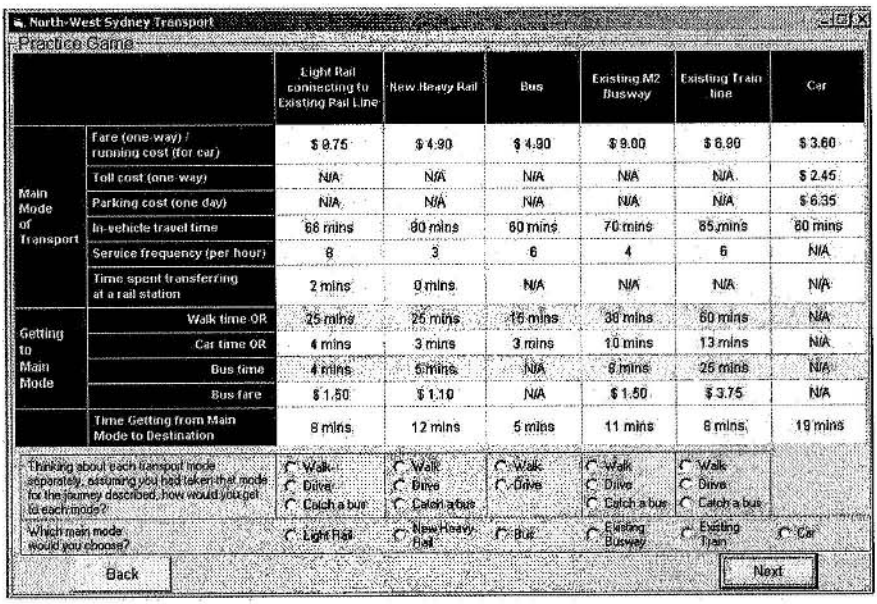


Figure 2.1. Example stated preference choice screen

4.2 Findings

The final multinomial logit and mixed logit models are given in Table 2.5. The overall goodness of fit (pseudo-R^2) is similar for MNL and mixed logit. The log likelihood function for the random parameters model is considerably larger than that for the MNL, which does suggest the improvement due to the broader specification. However, a direct test against the null hypothesis of the simpler model is not possible here. In our random parameters formulation, we

have specified that $\beta_{qk} = \beta_k + |\beta_k|\nu_{qk}$ where $\mathbf{v}_{qk}$ has a tent distribution, this constrains the range of the coefficient to one side of zero. As such, though the random parameters model introduces the new source of variation in the model, the individual heterogeneity, it does not introduce any new parameters. Therefore, the two models are not nested. (A simple likelihood ratio test would have zero degrees of freedom.) The statistically significant standard deviation parameters for all time and cost attributes (with a constrained triangular distribution) suggests that there is a structural advantage in selecting the mixed logit specification.

In-vehicle cost, in-vehicle time for the main mode and egress time from the main mode were specified as generic across all public transport modes and alternative-specific for car. The access and wait times were best represented as generic within all rail alternatives and likewise within all bus alternatives. For car, the parking cost was treated separately. The personal gross income of the respondent and their gender have a statistically significant influence on choice between public transport and car. All others things equal, the probability of choosing public transport decreases as personal income increases, less so for males.

All parameters associated with attributes of the modes (except the mode-specific constants) are specified as random parameters and are statistically significant. A constrained triangular distribution was chosen as the analytical representation of the behavioural profile of preference heterogeneity for each modal attribute.[9] Hensher and Greene (2003) show that when the mean parameter equals its spread (i.e. $\beta_q = \beta + \beta\nu_q$, where ν_q has support from -1 to $+1$), the density starts at zero, rises linearly to the mean, and then declines to zero again at twice the mean. It is peaked, as one would expect, bounded below at zero and above at a reasonable value that is estimated. The distribution is symmetric so that the mean is easy to interpret. Thus, the entire distribution is within the positive range. This is important when deriving estimates of *WTP* for specific attributes which have no logic in a two signs domain (for a specific empirical context). We note in passing, this detailed construction is not unlike the specification of a prior in the Bayesian context—in this instance, the specification is strongly based on an expectation of the shape of the true underlying distribution.

[9]For example, the usual specification in terms of a normal distribution is to define $\beta_i = \beta + \beta\nu_i$ where ν_i is the random variable. The constrained specification would be $\beta_i = \beta + \beta\nu_i$ when the standard deviation equals the mean or $\beta_i = \beta + h\beta\nu_i$ when h is the coefficient of variation taking any positive value. We would generally expect h to lie in the $[0, 1]$ range since a standard deviation greater than the mean estimate $typically$ results in behaviorally unacceptable parameter estimates.

Table 2.5. Summary of Empirical Results for Commuter Trips

Attribute	*Alternatives*	*Multinomial Logit*	*Mixed Logit*
New light rail constant	New light rail	2.451 (7.17)	2.048 (4.33)
New busway constant	New busway	1.500 (4.13)	1.093 (2.23)
New train constant	New train	2.282 (6.94)	1.844 (3.99)
Existing bus constant	Bus	1.996 (6.21)	1.533 (3.38)
Train constant	Existing and new Train	1.609 (4.78)	1.056 (2.25)
Existing busway constant	Busway	1.836 (5.49)	1.338 (2.86)
Destination inside study area	New light rail	–.9885 (–3.09)	–1.156 (–3.27)
Destination inside study area	New heavy rail	–1.301 (–3.52)	–1.552 (–3.82)
Gender (male = 1)	Public transport	1.353 (7.56)	1.899 (7.09)
Personal income	Public transport	–.0077 (–3.97)	–.0126 (–4.17)
Random Parameters in Mixed Logit (Std Dev = mean for constrained triangular distribution)			
Main mode in–vehicle cost	All public transport	–.1970 (–14.1)	–.2609 (–13.1)
Main mode in–vehicle cost	Car	–.1191 (–4.17)	–.1630 (–3.51)
Car parking cost	Car	–.0161 (–2.19)	–.0349 (–2.73)
Main mode in–vehicle time	Car	–.0333 (–8.50)	–.0767 (–6.59)
Main mode in–vehicle time	All public transport	–.0521 (–24.4)	–.0680 (–19.9)
Access plus wait time	All rail modes	–.0372 (–6.49)	–.0452 (–6.27)
Access time	All bus modes	–.0603 (–7.15)	–.0737 (–6.81)
Wait time	All bus modes	–.0727 (–3.01)	–.0822 (–2.90)
Egress travel time	Car	–.0533 (–3.87)	–.0855 (–3.68)
Egress travel time	All public transport	–.0122 (–2.59)	–.0145 (–2.60)
Access bus mode fare	Where bus is access mode	–.0911 (–2.81)	–.0950 (–2.62)
Log–likelihood at zero		–3580.47	–3580.47
Log–likelihood at convergence		–2463.46	–2435.75
Pseudo–R^2		0.31	0.32
Sample Size		1,840	1,840

Note: All public transport = (new heavy rail, new light rail, new busway, bus, train, busway);
time is in minutes and cost is in dollars ($2003). t-values in brackets in columns 3 and 4.
* The access mode travel time relates to the chosen access mode associated with public transport main.

4.3 *WTP* Derived from Individual Parameters

Of particular interest is the derivation of the conditional individual-specific parameter estimates and the associated values of travel time savings for each individual (see von Haefen, 2003 for a discussion of individual-specific welfare measures in RUM). As described in Train (2003), we can obtain the conditional estimator for any individual by using Bayes Theorem. The estimator for a specific parameter would be

$$\begin{aligned} E[\beta_q|data_q] &= \int_{\beta_q} \beta_q \Pr(\beta_q|data_q)d\beta_q \\ &= \int_{\beta_q} \beta_q \frac{\Pr(data_q|\beta_q)\Pr(\beta_q)}{\Pr(data_q)}d\beta_q \\ &= \int_{\beta_q} \beta_q \frac{\Pr(data_q|\beta_q)\Pr(\beta_q)}{\int_{\beta_q}\Pr(data_q|\beta_q)\Pr(\beta_q)d\beta_q}d\beta_q \\ &= \frac{\int_{\beta_q} \beta_q \Pr(data_q|\beta_q)\Pr(\beta_q)d\beta_q}{\int_{\beta_q}\Pr(data_q|\beta_q)\Pr(\beta_q)d\beta_q}. \end{aligned} \tag{4.1}$$

This is the classical counterpart to the posterior mean derived from (3.1). The marginal density, $\Pr(\beta_q)$ is implied by the distribution of ν_q in (3.2) where the distribution is induced by the stochastic specification of ν_q. The conditional density is the contribution of individual q to the likelihood function. The denominator in the conditional mean is the theoretical contribution of individual q to the likelihood function for the observed data. That is, the choice probability defined in (3.7). The numerator of the expectation is a weighted mixture of the values of β_q over the range of β_q where the weighting function is, again, the likelihood function. Since the integrals cannot be computed analytically, we compute them, once again, by simulation. The simulation estimator of the conditional mean for β_q is

$$\hat{E}_s[\beta_q|Individual_q] = \frac{1/R\sum_{r=1}^{R}\beta_{q,r}L(\beta_{q,r}|data_q)}{1/R\sum_{r=1}^{R}L(\beta_{q,r}|data_q)} \tag{4.2}$$

where the weighting function in each case is the contribution to the likelihood function (not its log), computed at the r^{th} draw of q, r in the simulation (see equation (3.2)). The approach in (4.2) can also be used to estimate the conditional variance or standard deviation of β_q by estimating the expected square and subtracting the square of the mean. This estimated conditional variance will be smaller than the average variance obtained simply by computing the sample variance of the estimated conditional means, as the latter is averaged over all the data in the sample while the former is averaged with respect only to the data for individual q. (Scarpa *et al.* 2005 have done a similar analysis of individual specific welfare measures based on mixed logit with both finite and continuous parameter distributions).

VTTS values are computed using the ratios, β_k/β_{Cost} for the various elements of travel time and in vehicle cost. Thus, the individual estimator of

Table 2.6. Behavioural Values of Travel Time Savings ($/Person Hour) Based on Individual Parameters: Mixed Logit Model, Commuter Trips (Mean Gross Personal Income Per Hour = $32.05)

Willingness to Pay Attribute	*VTTS ($/person Hour)*	
	Mean	*Range*
Main mode in-vehicle time – car	28.20	(6.37–39.80)
Egress time – car	31.61	(15.80–75.60)
Main mode in-vehicle time – public transport	15.71	(2.60–31.50)
Waiting time – all bus	19.10	(9.90–40.30)
Access time – all bus	17.10	(8.70–31.70)
Access plus wait time – all rail	10.50	(96.60–22.90)
Egress time – all public transport	3.40	(2.0–7.40)

VTTS, based on the theory of (4.1) and (4.2) is, for example, for the main mode in vehicle time for car:

$$\hat{E}_s[VTTS_q] = \frac{1/R\sum_{r=1}^{R}\hat{\beta}_{q,INVT,r}/\hat{\beta}_{q,INVCost,r}L(\hat{\beta}_{q,r}|data_q)}{1/R\sum_{r=1}^{R}L(\hat{\beta}_{q,r}|data_q)}. \quad (4.3)$$

Behavioural values of travel time savings (VTTS) based on individual parameters are summarised in Table 2.6 for the mixed logit commuter trip model. All of the VTTS have a distribution in the positive range. The mean and range of VTTS's are all intuitively plausible. We expect car VTTS to be higher than public transport with out-of vehicle (i.e. egress) time being higher than for the main mode in-vehicle time. For public transport, we normally expect out-of-vehicle times (especially wait time) to be valued higher than in-vehicle time. This is true when we contrast the VTTS for bus wait time with the other values. Access time has a significant element of in-vehicle time (e.g., bus to rail station) and this explains it similarity to the main mode in-vehicle time for bus. The two values that are relatively low are egress time for public transport (essentially walking) and the access plus wait time for all rail modes.

4.4 *WTP* Derived from Population Moments

Behavioural values of travel time savings (VTTS) for the mixed logit model using the population parameter estimates (i.e., marginals) are summarised in Table 2.7. These are computed simply by averaging over the entire population of individuals. For the example suggested above, the estimator would be:

$$\hat{E}_s[VTTS_q] = \frac{1}{R}\sum_{r=1}^{R}\frac{\hat{\beta}_{q,INVT,r}}{\hat{\beta}_{q,INVCost,r}}. \quad (4.4)$$

Table 2.7. Behavioural Values of Travel Time Savings ($/Person Hour) Based on Population Moments: Mixed Logit Model, Commuter Trips (Mean Gross Personal Income Per Hour = $32.05)

Willingness to Pay Attribute	*VTTS ($/person Hour)*	
	Mean	*Range*
Main mode in-vehicle time – car	36.04	(1.01 – 197)
Egress time – car	40.18	(1.12 – 219)
Main mode in-vehicle time – public transport	19.53	(0.204–99.9)
Waiting time – all bus	23.50	(0.25 – 120)
Access time – all bus	21.15	(0.18 – 108)
Access plus wait time – all rail	12.96	(0.14 – 66)
Egress time – all public transport	4.17	(0.04 – 21.3)

What is particularly noteworthy is the huge and unacceptable range of all *WTP* estimates in the positive domain. A closer look at this range shows a real problem with the upper maximum on all attributes. Approximately 12% of these values are above the maximum of the range in Table 2.6 derived from the individual parameters. This is expected given that the numerator and denominator are randomly drawn from the full distribution for each observation without the benefit of the objective priors available to the posterior approach. Eliminating approximately 12% of the upper tail and 8% of the lower tail appears to remove the differences in the mean between the two approaches. This is not the solution however. Fortunately it appears that taking ratios of individual-specific parameters resolves this.[10]

Train (2003) suggests that one possible way to control the values of *WTP* to a plausible range is to have either the numerator or denominator fixed when deriving *WTP* estimates. However, this commentary appears to be within the context of using population moments without the benefits of parameter mapping associated with individual-specific parameters conditioned on prior information on actual choices made. With individual-specific parameters it seems that the problem disappears.

5. Conclusions

The key result from this study is that it is very straightforward to use the classical inference paradigm to derive *WTP* estimates from individual-specific parameters for the attributes defining the numerator and denominator of the

[10]Such evidence has been found repeatedly by the authors using many data sets.

valuation expression.[11] We have illustrated the simplicity of estimating individual level parameters in the random parameters discrete choice model.

These *WTP* estimates are behaviourally meaningful at the mean across a sample and throughout the distribution that represents the preference heterogeneity of the sampled population. Establishing such behavioural sense requires judicial selection of analytical distributions for each parameter estimate (such as the constrained triangular distribution used in this paper), as is required for all methodologies, Bayesian or classical.

[11] This can be applied to measures of *WTP*, or other functions of the model components (see Greene (2003) for another example). The computation of individual level functions such as *WTP* or part-worths is a simple by product of the computation of the simulation estimator (and is already incorporated in NLOGIT 3.0).

Chapter 3

THE COST OF POWER OUTAGES TO HETEROGENEOUS HOUSEHOLDS

An Application of the Mixed Gamma-Lognormal Distribution

David F. Layton
Daniel J. Evans School of Public Affairs
University of Washington
dflayton@u.washington.edu

Klaus Moeltner
Department of Resource Economics
University of Nevada, Reno
moeltner@cabnr.unr.edu

Abstract We use a repeated dichotomous choice contingent valuation survey to elicit households' willingness to pay to avoid unannounced interruptions in electricity service. The data pose multiple econometric challenges including: correlated responses for a given household, heteroskedastic errors, and a willingness to pay distribution with large mass near zero. We address these issues by combining a gamma distribution for outage costs with a lognormally distributed scale parameter defined as a function of household characteristics, outage attributes, outage history, and random coefficients. The model is estimated through simulated maximum likelihood. We demonstrate that cost estimates are sensitive to the interaction of attributes of previously experienced and hypothetical interruptions.

Keywords: Power outage costs, non-market valuation, gamma distribution, random coefficients, maximum simulated likelihood.

1. Introduction

Sudden power outages can greatly disrupt social and economic activities of residents and firms in affected areas. Historically, the vast majority of interruptions in electricity supply in the U.S. occurred due to damages to the distribution network of a utility, usually during inclement weather conditions. Outages related to generation or transmission failures have been less common, as regulators required vertically integrated utilities to maintain a generating margin in electricity production and reliability reserves in their transmission system. However, with the advent of deregulation of the electric power sector in many states during the last decade, service interruptions caused by disparities between supply and demand for electricity have become more frequent. As became evident during the 2000/2001 power crisis in California, newly implemented deregulation schemes may lack the right incentive structure for generators to maintain or expand capacity reserves. This shortcoming, in combination with congested transmission grids and rigidities in retail pricing can result in market failures, supply shortages, and—ultimately—widespread blackouts (Faruqui *et al.*, 2001; Joskow, 2001; Borenstein, 2001). Since most existing deregulation frameworks are still in their infancy and thus may be susceptible to similar design problems, the risk of generation and transmission type outages is likely to remain higher in deregulated states compared to regulated markets in the near future.

Regardless of the specific cause of a service interruption and the regulation status of the utilities involved, the development of efficient policies to reduce the risk of blackouts requires knowledge of the economic costs they cause to customers. In a traditional power market regulators can induce utilities to step up reliability efforts (for example by replacing distribution lines with underground connections) by allowing them to recoup their increased cost of service through higher electricity rates. Such rate changes will only be acceptable to end-users if they are proportionate to their value of improved service reliability. In restructured markets, power distributors (retailers) are separate entities and generally remain under some degree of regulation. Many states are considering incentive frameworks that link retailers' rate of return to actual service performance (Energy Information Administration, 1998). Outage prevention and service restoration times are commonly used yardsticks in such performance-based contracts (Warwick, 2000). Naturally, an understanding of costs incurred by customers during service interruptions will be vital to set performance criteria and design economically efficient incentive structures in this context. Markets for generation and transmission in process of deregulation generally rely on an independent system operator (ISO) to ensure acceptable reliability levels. These control units usually apply a mix of economic and administrative tools to maintain adequate generation and transmission levels,

such as transmission tariffs, defaulting fines, and regulations governing capacity markets. Again, economic efficiency dictates that the costs of system failure, i.e. service interruptions, ought to enter into the design of these instruments.

The focus of this study is on outage costs to residential customers. This segment comprises over 85% of retail customers in the U.S. and contributes more to retail sales and revenues than any other user group (35% and 43%, respectively; Energy Information Administration, 1998). Also, households are subjected to the highest risk of interruptions, as they rely on a more extensive infrastructure of distribution lines, substations, and transformers than larger commercial and industrial users (Warwick, 2000). We propose an innovative survey and estimation framework to elicit costs to households associated with specific power outages. Our model allows for the inclusion of both household characteristics and outage features, while capturing unobserved heterogeneity in household preferences for service reliability. In addition, we extend the existing outage cost literature by explicitly analyzing the joint effect of experienced and hypothetical interruptions on welfare losses. Due to the presence of high-dimensional integrals, the estimation of this model requires the application of simulated maximum likelihood techniques.

In the next section we provide a brief discussion of previously used approaches to estimating residential outage costs and motivate our new approach based on repeated dichotomous choice valuation questions. In section we develop an econometric model designed for the repeated dichotomous choice data. In section we discusses the data, estimation results, and policy implications. Section concludes the paper.

2. Approaches to Estimating Residential Outage Costs

Despite the importance of this topic in a time of continued growth in electricity reliance (Energy Information Administration, 2001), there exist only a few studies on outage costs to residential customers in the published literature. Three general methodologies to derive cost estimates have been proposed in existing work. In the first approach households are asked directly their willingness to pay (*WTP*) to avoid a specific outage type (Woo *et al.*, 1991; Beenstock *et al.*, 1998). The second method, as applied in Wacker *et al.* (1985) and Doane *et al.* (1988a) is based on households' direct estimates of itemized costs associated with a given menu of mitigating actions during a service interruption. The third methodology is anchored in a discrete choice framework where households are asked to select one or rank all of several outage scenario/payment options. Examples are Goett *et al.* (1988), Doane *et al.* (1988b) and Beenstock *et al.* (1998).

The first two methods lead to regression models with continuous dependent variables (cost or *WTP* in dollars) that are either estimated through simple OLS (Wacker *et al.*, 1985, Woo *et al.*, 1991), Tobit (Beenstock *et al.*, 1998), or a 2-Stage Heckman model (Doane *et al.*, 1988b). The latter two estimation techniques take account of the fact that some residents report zero costs or *WTP* for a given power interruption. Studies following the third elicitation approach apply variants of the conditional logit model (McFadden, 1974) to generate cost estimates.

Each approach has its benefits and drawbacks. Models based on open-ended *WTP* reports, while computationally convenient, are susceptible to strategic response and non-response bias (Arrow *et al.*, 1993; McFadden, 1994; Beenstock *et al.*, 1998). Asking respondents to itemize costs may mitigate these problems to some extent. However, as such cost menus only capture outlays for a limited number of actual market transactions (purchase of candles and batteries, dining out, etc.), there is a risk of missing non-market welfare losses associated with blackouts, such as health and safety concerns, disruption of work or study, and interference with social events and past-time. Thus, cost estimates from such lists can only be interpreted as lower bounds for actual welfare losses, assuming truthful responses to each line item.

Discrete choice elicitation methods based on conditional logit analysis, in turn, have the theoretical ability to capture both market and non-market values associated with specific outage types. In the residential outage costs literature there is some evidence that such multi-choice models may trigger 'status quo' bias (i.e. a household's inherent resistance to any changes in service provision) and asymmetry effects (i.e. the value of service deteriorations categorically exceeds the value of service improvements), as shown in Doane *et al.* (1988b), Hartman *et al.* (1991), and Beenstock *et al.* (1998).[1]

In this study, we promote the use of repeated dichotomous choice questions to elicit the cost of power outages to residential customers. The dichotomous choice, or referendum-style, format has been found to provide a more familiar decision making context to respondents, and to largely avoid creating strategic response incentives (Arrow *et al.*, 1993; Hanemann, 1994). In our application, each respondent is presented with a series of hypothetical outage scenarios, differing in length and time of occurrence. For each scenario, households have to decide if they would be willing to pay a given amount to avoid a specific interruption, or tolerate the outage with no change in electricity costs. This format collects a large amount of information from each respondent and allows

[1]Furthermore, unlike in most of the studies that have used multi-choice conjoint experiment-designs for their valuation questions, the policy objective for this study was NOT an examination of the potential for price-differentiated service packages. Thus there was no need to present more than two relevant "choices". In addition, as the power interruptions we model in this study are by definition unannounced random events, it would appear counterintuitive to ask a given respondent to "choose" from a set of different outages.

for the valuation of attributes that describe outages. As the econometric model provides household-level estimates of willingness-to-pay-to-avoid outages, or, equivalently, outage cost, as a function of outage and household characteristics, it allows for forecasts of costs to residential customers of new types of outages not explicitly included in the experimental design.

There are a number of econometric challenges associated with the repeated dichotomous choice approach. In the next section we develop an econometric model appropriate for using repeated dichotomous choice data for estimating the costs of residential power outages.

3. The Econometric Model

The econometric model is designed to handle four important features in the data. First, many households may have a near-zero cost attributable to a power outage, especially outages of short or momentary length. Second, it can reasonably be assumed that no household obtains positive value from a power outage. Considering these two points together suggests a distribution of *WTP* that is non-negative but can allow for substantial mass near zero.[2] Third, the survey data consists of up to four responses per respondent, and so we anticipate that intra-respondent *WTP* responses will exhibit correlation due to common unobservables for a given household. Fourth, the interaction of household unobservables with outage attributes will likely lead to heteroskedastic errors. As is well known, ignoring such heteroskedasticity in a discrete choice context will lead to inconsistent parameter estimates, especially if elements of the error variance are correlated with regressors (e.g. Hanemann and Kanninen, 1999).

The model we develop incorporates these four features. We follow Cameron (1988) and specify directly a probability density function (pdf) for latent individual *WTP*. Specifically, we choose a gamma kernel for the distribution of *WTP*. The gamma distribution for dichotomous choice data has been previously considered by McFadden (1994) and Werner (1999). It constrains *WTP* to be non-negative, but is also flexible enough to allow for exponential or normal like behavior, with much of the mass near or far away from zero as implied by the data.[3] As in McFadden (1994) and Werner (1999) we express the scale

[2] See Haab and McConnell (1998) for a detailed discussion of willingness to pay distributions.

[3] McFadden (1994) and Werner (1999) estimate models with a discrete-continuous distribution for *WTP*. The population is modeled as having two components, those with a zero *WTP*, and those with a positive *WTP*. There is a discrete probability of a respondent having a zero *WTP*. The positive component of the *WTP* distribution is modeled using a continuous distribution. This model has come to be called a "spike" model, given the discrete spike at zero in the *WTP* distribution modeled with an extra parameter. See Hanemann and Kanninen (1999) for a more detailed discussion. This approach has not been applied to repeated dichotomous choice valuation with multiple scenarios. The spike model is difficult to generalize to multiple response valuation for a number of reasons. First, the spikes for each scenario are unlikely to be equal requiring additional parameters, or need to be modelled as functions of the attributes of the scenario, along with the continuous portion of the *WTP* distribution which may make identification in

parameter of the gamma distribution as an exponential function of explanatory variables. We then extend this specification by modeling some of the coefficients associated with the regressors as random parameters. In addition to introducing the desired intra-household correlation across choice occasions and heteroskedasticity, this specification allows for an explicit analysis of the interactive effect of various outage attributes on *WTP*. We describe below the econometric model and its estimation via simulated maximum likelihood.

First we consider the *WTP* model without unobserved preference heterogeneity. Respondents are presented with a question that asks whether they would be willing to pay $\$B$ (B stands for bid) to prevent a power outage. Each power outage is described by a set of characteristics, as is each respondent. Denoting the vector of power outage and household characteristics by *x*, we model each respondent i's *WTP* for a given outage, j , as a function $WTP_{ij}(x_{ij}, \theta)$ where θ is a set of parameters. Bids can vary across outage types and respondents, and respondents will answer that they will pay B_{ij} if B_{ij} is less than their WTP_{ij}. Formally, we follow standard practice, (see Cameron and James 1987) and denote a "yes" response by $Y_{ij} = 1$, and a "no" response by $Y_{ij} = 0$. Then:

$$Y_{ij} = \begin{cases} 1, & \text{if } B_{ij} < WTP_{ij}(x_{ij}, \theta) \\ 0, & \text{if } B_{ij} > WTP_{ij}(x_{ij}, \theta) \end{cases} \tag{3.1}$$

We take a fully parametric approach. To ensure a positive *WTP*, one can either formulate the model in terms of a transformed Bid variable, usually by taking logs, and assuming that an additive error term is from a normal or logistic distribution (for example, Cameron and James, 1987). The alternative approach we follow is similar to the suggestions outlined in Haab and McConnell (1998). Specifically, we consider equation (3.1) as written with the bid, B_{ij}, in levels, but utilize a distribution for WTP_{ij} that takes only non-negative values. We begin by assuming that the WTP_{ij} are independently distributed across households and outages as $Gamma(b_{ij}, c)$, so that the density function $f(WTP_{ij})$ is:

practice difficult. Second, one would expect substantial correlation across scenarios as those who have a zero *WTP* for one scenario are far more likely to have a zero *WTP* in another scenario. This would necessitate the use of a multivariate distribution over the "spikes" which is difficult to implement. On an intuitive level, it is not unreasonable in our context to assume that every household experiences at least incremental costs or disutility from even the shortest outage. In both studies mentioned above, the item to be valued are wilderness areas or wildlife sanctuaries. In those applications, it seems more likely that some respondents' *WTP* is truly nonpositive, and that this subpopulation is distinctly different from other stakeholders.

$$f(WTP_{ij}) = \frac{\left(WTP_{ij}b_{ij}^{-1}\right)^{c-1}\left[\exp\left(-WTP_{ij}b_{ij}^{-1}\right)\right]}{b_{ij} \times \Gamma(c)} \tag{3.2}$$

for $0 \leq WTP_{ij} < \infty$, with $b_{ij}, c > 0$, and $\Gamma(c)$ is the gamma function evaluated at c (e.g. Evans *et al.*, 2000). Following McFadden (1994) and Werner (1999) we model the scale parameter, b_{ij}, as an exponential function of a linear combination of explanatory variables x_{ij} and associated coefficient vector θ_i, i.e. $b_{ij} = \exp(x'_{ij}\theta_i)$. This ensures that $b_{ij} > 0$, as required. As indicated above, to capture household heterogeneity and to introduce correlation across intra-household responses we model these coefficients as stochastic terms. Specifically, we let θ_i follow a multivariate normal distribution with mean vector μ and variance-covariance matrix Ω, i.e. $\theta_i \sim mvn(\mu, \Omega)$. This specification allows the elements of θ_i to – a priori – have unrestricted sign and magnitude. As illustrated in Moeltner and Layton (2002) the estimated covariance terms of Ω can provide additional information on joint effects of different regressors on the dependent variable.

The assumption of multivariate normality for θ_i implies a lognormal-gamma mixture distribution for WTP_{ij}. Note that in contrast to b_{ij} we treat the shape parameter c as common to all respondent-scenario combinations. The mean for this distribution can be conveniently expressed as cb_{ij} (Evans *et al.*, 2000). Note that expressing the scale parameter, b_{ij}, as an exponential of a linear function of covariates and outage attributes will make expected outage costs for a given respondent a non-linear function of all of the covariates and outage attributes interacted together. We will graphically illustrate the rich results this feature of the model yields when we discuss our application.

Since each household responded to up to four outage scenario/bid combinations, computation of the joint probability of observing a specific choice sequence for a given respondent requires computing a multi-dimensional integral. We approximate these probabilities using the random parameter simulator as described in Revelt and Train (1998), McFadden and Train (2000), and Layton and Brown (2000). To simulate the probability of each respondent's set of responses, we first compute the gamma probability conditional on the θ_i, then we simulate the unconditional probability using draws from θ_i's multivariate normal distribution. Considering (3.1), conditional on θ_i, the probability that a respondent says "yes" to a particular valuation question is $1 - F(B_{ij})$, and the probability of a "no" is $F(B_{ij})$ where:

$$F(B_{ij}) = \int_0^{B_{ij}} f(WTP_{ij})dWTP_{ij},$$

and $f(.)$ is the gamma pdf shown in (3.2). $F(.)$ is not closed form, but is readily computed in standard statistical packages. Denote the appropriate "yes"

or "no" conditional probability for a particular valuation scenario by $P_{ij}|\theta_i$. Under our assumptions, the $P_{ij}|\theta_i$ are statistically independent for person i, across all j. Thus the probability, $P_i|\theta_i$, of a series of responses conditional on θ_i is

$$P_i|\theta_i = \prod_{j=1}^{j=m} P_{ij}|\theta_i, \tag{3.3}$$

where m indexes the number of WTP questions. The unconditional probability for person i, is:

$$P_i = \int_{-\infty}^{\infty} \left(\prod_{j=1}^{j=m} P_{ij}|\theta_i \right) f(\theta_i) d\theta_i, \tag{3.4}$$

where the dimension of the integration is equal to the number of random parameters in θ_i. Simulation of P_i is straightforward following Brownstone and Train (1999) and McFadden and Train (2000). At each iteration of the maximum likelihood routine we draw R sets of θ_i as $MVN(\mu, \Omega)$, and compute the simulated P_i, $\tilde{P}_i$, as the average over the R draws:

$$\tilde{P}_i = \frac{1}{R} \sum_{r=1}^{r=R} \left(\prod_{j=1}^{j=m} P_{ij}|\theta_{ir} \right). \tag{3.5}$$

The elements of μ and Ω are updated throughout the optimization process.

4. Empirical Analysis

4.1 Data

The data are from a fall 1998 survey of residential customers implemented by a U.S. utility. The main objective of the survey was to identify priority neighborhoods for reliability improvements in power distribution based on the *WTP* to avoid an outage. Each household was presented with four outage scenarios. For each scenario, households could avoid the outage by use of a preinstalled backup generator for which they would pay a specific fee every time the generator was activated by a power interruption. The selection of scenarios was based on Sullivan *et al.* (1996) and was subjected to further pre-testing using focus groups. Each scenario differed in terms of season (summer versus winter), outage timing and duration, and corresponding bid amounts. The timing and duration were chosen in consultation with the utility. Given the duration of the outages, bid amounts were based on the open-ended *WTP* data from Sullivan *et al.* (1996) in conjunction with results of the focus groups.

The following additional considerations guided the experimental design. First, the bids are such that for a given respondent a one hour outage never costs less than a momentary outage, a four hour outage never costs less than a one hour outage, and so on. Given that the open-ended data previously collected by Sullivan *et al.* (1996) for another utility revealed a fairly long right tail in the distribution of *WTP*, we allowed for a number of fairly high bid levels to be able to adequately model the skewed distribution. A wide range of bids was used to account for the fact that the location (mean, median) of the *WTP* distribution for the consumers in question might be significantly higher or lower than in the previously available data from another utility. Finally, the survey versions were carefully designed to avoid any implicit ordering of the bids through the four scenarios.

The mail survey yielded a 63% response rate. After elimination of protest responses and observations with missing household characteristics, 4,528 observations from 1,421 households were retained for this analysis. Seven of the eight administered scenarios were for winter time outages which is our focus here. Table 3.1 summarizes the seven scenarios we utilize, bid ranges, and sample counts. Household characteristics were collected as part of the survey and are supplemented with information available from customer accounts. Table describes a set of variables that relate to the types of electricity needs a household may have, which we utilize in our model estimation.

Table 3.1. Scenario and bid design

Scenario	*Duration (hrs)*	*Time*	*Bid Levels*		*No. of Obs.*
			lowest	highest	
1	1	7 pm	0.5	30	652
2	4	7 pm	1.0	50	642
3	1	8 am	0.5	40	656
4	Moment (1-2 sec.)	7 pm	0.5	30	654
5	1	midnight	0.5	40	665
6	12	7 pm	15.0	100	623
7	1	3 pm	0.5	40	636
				Total:	4,528

Note: All outages occur on a winter weekday and are unannounced.

4.2 Model Estimation

Our model specification includes a dummy variable for evening outages (*evening*), the log of the duration of a given outage scenario in minutes (

ln_dur), and the household attributes listed in Table 3.2. In various combinations, most of these variables have been considered as determinants of residential outage costs in existing studies (e.g. Doane *et al.*, 1988a; Beenstock *et al.*, 1998). We add to this traditional set of regressors a dummy variable for mobile home residences and a dummy variable taking the value of one if a given household has access to non-electric power sources for heating. The last two variables in Table 3.2, the number and log of total duration of outages during the preceding 12 months, are included in the model to measure the effect of outage history on *WTP* (or cost) estimates. While other studies have captured the impact of past outage occurrences on households' *WTP* to avoid future interruptions (Doane *et al.*, 1988b; Hartman *et al.*, 1991; Beenstock *et al.*, 1998) the separate inclusion of historic outage counts and combined duration appears to be novel. As we will show, these two indicators have significant and offsetting effects on cost estimates.

Table 3.2. Household Characteristics

Variable	*Description*	*Mean*	*Std. Dev.*
generate	1 = home has generator	0.17	0.37
business	1 = business at home	0.13	0.34
medical	1 = medical need at home	0.03	0.17
home	1 = someone at home most of the time	0.60	0.49
hh_ size	household size (persons)	2.60	1.45
over64	number of persons over 64	0.36	0.70
inc000	annual income, $1000	53.02	27.83
mobile	1 = mobile home	0.10	0.29
other_ heat	1 = secondary heating source available	0.49	0.50
ln_ cons	log of avg. monthly electricity consumption in kwh	6.75	0.71
num_ out	number of outages in past 12 months	5.58	8.36
out_ past	log of total duration of outages in past 12 months (hours)	1.17	1.87

We model the outages in the dichotomous choice scenarios as consisting of two components: A short momentary component of less than a minute, and then any additional duration beyond one minute. Momentary outages that last for less than a minute have a particular suite of impacts on some households but not on others. Sensitive electrical equipment such as medical devices and home office systems may fail, but most other home electricity uses will not be greatly affected. Longer outages share this initial effect and as duration increases other costs begin to mount. The literature suggests that the impact of outage duration increases, but at a decreasing rate, so we model the effect of duration beyond the first minute in log form. Following Moeltner and Layton (2002) we include an intercept term in our model while setting the value of log

duration for a momentary (1 minute or less) outage scenario to zero. Thus the intercept term is the effect of a momentary outage on *WTP* (*moment*), and the coefficient on log duration (*ln_dur*) measures the impact of a duration length longer than a minute.

Specifying all k =15 elements of θ_i as correlated random coefficients would require estimation of k elements of μ plus $k(k+1)/2$ elements of Ω for a total of 135 parameters. Such a large number of parameters are not likely to be identified without a prohibitively large data set. Further, estimation is not computationally feasible given the need to simulate the response probability at each function evaluation (Keane, 1997). We thus restrict randomness to variables of primary interest with likely heterogeneity in preferences. These are past outage duration (*out_past*) and occurrence (*num_out*), as well as the two main attributes of the hypothetical outage scenarios, *ln_dur* and *moment*.[4] Adding the resulting ten variance-covariance terms in Ω and the gamma shape parameter c to the 15 elements of θ_i yields a total number of 26 model parameters. We estimate this model through simulated maximum likelihood using R =1,000 repetitions for the simulated probabilities described in (3.5).

Table 3.3 summarizes the estimation results from the mixed Gamma-Lognormal model. Generally, the model exhibits a reasonably good fit with the underlying data with a pseudo-R^2 of 0.25. The majority of the coefficient estimates are significant at the 5% level or higher. Specifically, the gamma shape parameter, c, is estimated with high precision. A value of c less than one indicates a high probability mass near zero (Evans *et al.*, 2000). This result is compatible with similar findings by Doane *et al.*, (1988a) and Beenstock *et al.* (1998), who report a preponderance of zeros in their open-ended *WTP* elicitations even after purging their data of potential strategic and protest responses. Evidently, a large share of residential customers in our sample does not consider the bulk of the power outages described as especially bothersome. We discuss in turn the results for the outage attributes, household characteristics, and past outage history before elaborating on the covariance estimates. We conclude this section with a comparison of our results to those available in the literature.

4.3 Outage Attributes

The effect of evening outages emerges as insignificant compared to a combined baseline of afternoon and morning interruptions. A possible explanation for this finding may be that the period of daylight during winter is relatively short in the survey region. Accordingly, electricity needs for lighting are re-

[4]The evening dummy is specified as a fixed coefficient in part on the basis of preliminary work, which suggested it had little impact in our data set.

Table 3.3. Estimation Results

Parameters	*Coeff.*	*Stand. err.*	
c	0.232	(0.023)	***
evening	–0.015	(0.132)	
generate	–0.494	(0.174)	***
business	0.338	(0.200)	*
medical	0.304	(0.407)	
home	0.658	(0.148)	***
hh_ size	–0.124	(0.048)	**
over64	–0.215	(0.109)	**
inc000	0.026	(0.003)	***
mobile	0.457	(0.222)	**
other_ heat	–0.467	(0.134)	***
ln_ cons	0.133	(0.101)	
ln_ dur	0.455	(0.056)	***
num_ out	–0.033	(0.011)	***
out_ past	0.260	(0.066)	***
moment	–0.933	(0.808)	
Variance and Covariance Terms			
ln_ dur	0.077	(0.062)	
ln_ dur / num_ out	0.001	(0.006)	
num_ out	0.000	(0.000)	a
ln_ dur / out_ past	0.138	(0.069)	**
num_ out / out_ past	0.002	(0.010)	
out_ past	0.257	(0.154)	*
ln_ dur / moment	–0.899	(0.514)	*
num_ out / moment	–0.015	(0.065)	
out_ past / moment	–1.647	(0.651)	**
moment	10.648	(4.075)	***
Log-likelihood	2,356.100		
Pseudo-R^2 = 1-[–2,356.1/ln(0.5)]	0.250		

Note: Standard Errors in parentheses. a = rounded to zero;
*significant at 10% level; ** significant at 5% level;
*** significant at 1% level.

quired for much of the day. In addition, many businesses in the particular metropolitan area that generated this sample offer staggered work shifts, which distributes electricity needs more evenly over a 24 hour time period. An alternative explanation is that the limited number of outage scenarios that could be valued in the survey did not permit sufficient contrast between duration and time of day. For instance a 12 hour evening outage would cover evening, late night, and morning, thus mitigating much of the time of day effect. This

suggests that when time of day is a variable of important policy interest in conjunction with outages of long duration, many survey versions will be required – perhaps prohibitively many.

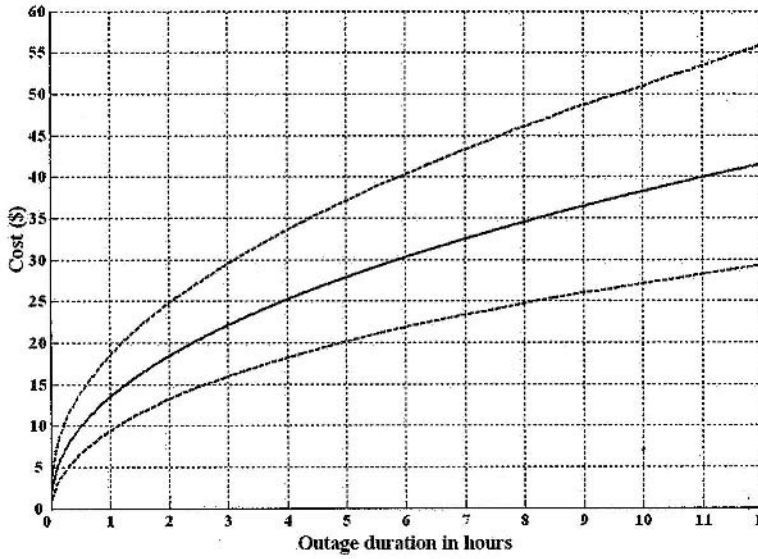


Figure 3.1. Outage Costs Versus Duration (mean and 95% confidence intervals averaged over all households). Outage costs increase at a decreasing rate with outage duration.

As reflected by the insignificant coefficient for moment, a purely instantaneous interruption does not cause any sizeable costs to the average household. This is consistent with findings reported in Caves *et al.* (1990). Outage costs and associated *WTP* values do, however, increase with the duration of an interruption as indicated by the positive sign and high level of significance for *ln_dur*. Figure 3.1 depicts the resulting duration-cost function for a prototypical household and an interruption starting at 7pm. The 95% confidence intervals are based on the empirical distribution of household-specific estimated *WTP* averaged over all respondents. Given the distributional assumptions in our model, the expectation of WTP_{ij} is itself a random variable following a lognormal distribution with mean $E(b_{ij}c) = c \times \exp\left(x'_{ij}\mu + 0.5x'_{ij}\Omega x_{ij}\right)$. Due to some outliers, this expression generates excessively high values for some households. We therefore use the median, $c \times \exp(x'_{ij}\mu)$, as the basis for our point and interval estimates of WTP_{ij}.

Figure 3.1 depicts how median costs change over outage duration. Consistent with results reported by Doane *et al.* (1988a), outage costs increase at a decreasing rate with increasing duration. This is intuitively sound as longer outages give households more time to take countervailing measures. At the same time, the variability of outage damages to individual households increases with duration as indicated by the widening spread of the confidence interval. For example, for a one-hour evening interruption our point estimate is \$13 with a 95% confidence interval of \$9 to \$18. At a duration of 12 hours the point estimate is \$42, with a 95% confidence interval of \$29 to \$56.

4.4 Household Characteristics

Turning to the effect of household characteristics, we note from Table 3.3 that the presence of medical needs and annual electricity consumption do not significantly affect *WTP* values. As expected, the presence of business activities run from home, the presence of residents at home during most of the day, and income have a positive and significant effect on cost estimates. Similarly, households residing in mobile homes have a significantly higher sensitivity to power interruptions. This is an anticipated result given the reduced insulation of such dwellings and the corresponding higher reliance of their occupants on uninterrupted heating. As expected, the availability of non-electric heating significantly reduces the *WTP* to avoid a specified interruption. The negative and significant coefficients for household size and the number of persons over age 64 are probably indicative of reduced disposable income for such families.

4.5 Past Outage History

One of the key insights provided by this analysis flows from our specification of a rich structure for past outage history. By using two components of past outage history, log duration and number of occurrences, we show that past outage history is not a uni-dimensional concept, but instead illustrate that different components of outage history have contrasting effects on *WTP*. These contrasting effects derive from the significant and opposite signs for the number of past outages during the preceding 12 months (*num_out*) and the log of combined duration of such outages in hours (*out_past*). This implies that an increase in historic outage frequency, ceteris paribus, decreases a household's *WTP* to avoid further interruptions. This could be indicative of a learning-to-cope, or preparedness effect induced by frequent outage events. In stark contrast, however, estimated *WTP* increases with the combined duration of recent interruptions. Evidently, one or more longer blackouts in the recent past stir up decidedly unpleasant memories in affected respondents and seem to induce substantially higher costs than they generate learning gains.

These results may explain the contradicting findings in the existing outage cost literature on the role of outage history on cost estimates for a specified future interruption. Specifically, Doane, Hartman *et al.* (1988b) find that households that traditionally experience a larger number of outages have a decreased *WTP* to avoid additional blackouts, while Doane *et al.* (1988a) and Beenstock *et al.* (1998) reach the exact opposite conclusion. This apparent discrepancy could be a result of different average length of past outages in each of these cases. None of these studies incorporate measures of historic outage duration in their estimation models.

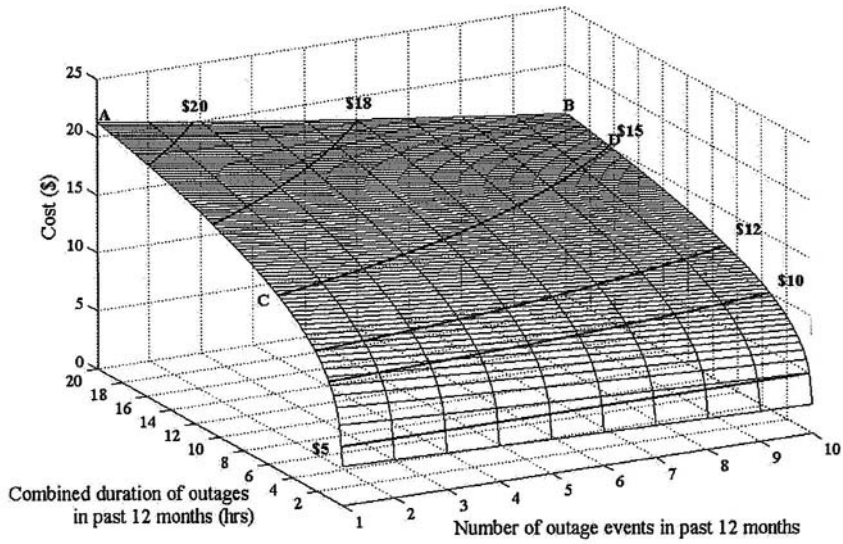


Figure 3.2. *WTP* for a One-Hour Outage: Past Duration vs. Past Events. The iso-*WTP* lines show that *WTP* to avoid this particular outage is the same for households that have experienced one long outage or several shorter interruptions. Alternatively, *WTP* is lower if a given past duration is distributed over several outage occurrences. This may be indicative of two countervailing forces: a cost-awareness factor versus a learning-to-cope effect.

Figure 3.2 illustrates these offsetting effects. The surface plane of figure 3.2 connects simulated cost estimates for a hypothesized evening outage of one-hour duration at different combinations of number and duration of past interruptions. In each case, cost estimates were first generated for each respondent and then averaged over households. For example, a prototypical household that experienced one 20-hour outage in the past would be willing to pay approximately $22 to avoid the stipulated interruption (point A). If the com-

bined duration of 20 hours is distributed over, say, 10 individual events, *WTP* decreases to approximately \$16 (point B). The darker lines crossing the cost surface represent iso-*WTP* curves for several dollar amounts. For instance, a *WTP* value of \$15 could be reached with one 6-hour interruption (point C), ten outages with a combined duration of 16 hours (point D) or any of the frequency / duration pairs along the line C-D.

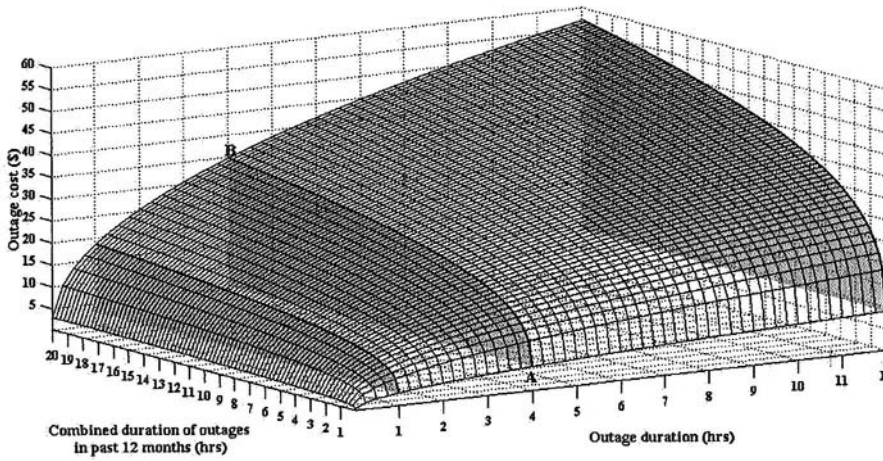


Figure 3.3. Outage Costs by Scenario Duration and Past Duration. The WTP to avoid future outages of a given duration increases with the combined duration of past outages. This effect is relatively stronger for longer hypothetical future outages.

Figure 3.3 shows the importance of capturing historic duration effects when analyzing the value of electric reliability to residential customers from a different perspective. The figure depicts cost estimates associated with a new unannounced future outage as a function of combined duration of past interruptions. For ease of interpretation, cost estimates for one, four, and 12-hour duration are highlighted through cross-section planes. For example, a four-hour evening interruption causes costs of approximately \$7 to a household that has not experienced any blackouts over the preceding 12 months (point A). In contrast, *WTP* to avoid a four-hour outage is about five times higher for a household with a combined past duration of 20 hours (point B). Comparing the vertical height of the cross-sections at any historic duration value, one can

also note that combined past duration affects *WTP* estimates relatively more for longer proposed interruptions.

The recognition of these interactive effects of frequency and duration of past blackouts may offer additional guidance to utilities in identifying residential neighborhoods with relatively high sensitivity to power interruptions. Clearly, a sole focus on the number of past outage events in this context may lead to sub-optimal allocation of reliability efforts.

4.6 Covariance Parameters

The bottom half of Table 3.3 shows estimation results for the random elements of Ω. Based on a likelihood ratio test, we strongly reject the null hypothesis that all elements of Ω are zero, i.e. that outage features and outage history have an equal effect on *WTP* for all customers. The variance term for moment is relatively large and highly significant. This indicates strong heterogeneity in costs from an instantaneous interruption across individual households. Thus, even though the costs caused by momentary outages are negligible for the average household, there are some families that experience considerable damage even from a very short blackout. This mirrors the results for costs of momentary outages to commercial/industrial firms in Moeltner and Layton (2002). The remaining elements of Ω that emerge as significant at the 5% level or higher are the covariance terms for *out_past* with *ln_dur* and *moment*, respectively. The first term is positive and lends itself to an intuitively sound interpretation: households that are relatively more affected by combined past duration are also more sensitive to duration as specified in the hypothetical scenarios. The negative sign on the second term suggests that households whose *WTP* values depend more strongly on the combined duration of experienced outages are relatively less sensitive to instantaneous interruptions. Conversely, households whose *WTP* is affected less strongly by the duration of past interruptions are relatively more concerned about instantaneous blackouts. Presumably, such residents experience the bulk of outage costs during the initial period of a blackout and are relatively more capable to cope with prolonged duration. The negative sign for the covariance term between *moment* and scenario duration (*ln_dur*) supports this hypothesis even though this term is only significant at the 10% level.

4.7 Comparison with Previous Estimates

Our results are compared to those stated in other studies in Table 3.4. As recommended by Caves *et al.* (1990) and Woo and Pupp (1992) we report outage costs in both absolute dollars and in terms of dollars per kwh unserved, using additional information provided by the Utility on energy consumption by a prototypical household for the season and time period of consideration.

Table 3.4. Cross-study Comparison of Cost Estimates

	This Study	Doane *et al.* (1988)	Doane *et al.* (1988b)	Woo *et al.* (1991)
Data Year:	1998	1986	1986	1989
Timing:	winter evening	winter evening / morning	winter eve./ mor.	winter
Method:		2-stage Heckman	Self-stated	OLS
Duration		Cost (1998 $)		
1 hr	13.45	16.33	13.66	9.83
4 hrs	25.17	29.16	26.79	13.10
8 hrs	34.49	N/A	N/A	19.65
12 hrs	41.51	49.39	58.11	30.13
		Cost ($/kwh unserved)		
1 hr	5.34	14.61	N/A	12.71
4 hrs	2.66	5.29	N/A	7.34
8 hrs	2.29	N/A	N/A	4.98
12 hrs	2.06	3.38	N/A	3.28

As can be seen from the table, our cost estimates in absolute dollars are reasonably close to those found in Doane *et al.* (1988a), and the results based on self-stated costs in Doane *et al.* (1988b). The estimates by Woo *et al.* (1991) are clearly lower than those produced by this analysis and the other two comparison studies for all listed outage durations. To some extent, this may be related to the fact that Woo *et al.* (1991) use OLS regression to generate these cost estimates. This is likely to place the mean of the resulting underlying cost distribution closer to zero than would be the case in models that impose non-negativity constraints on outage costs, as applied in the other three studies. At the same time, these lower estimates may simply indicate a relatively lower reliance on electric power of the particular population captured in that analysis. When expressed in terms of dollars per kwh unserved Woo *et al.* (1991)'s cost estimates are close to the ones reported in Doane *et al.* (1988a), which would lend support to the latter hypothesis. While following the same general decrease-with-duration pattern, our cost estimates in dollars per kwh unserved are about 50% smaller in magnitude than those generated by the two comparison sources. This suggests that cumulative energy consumption by a representative household from our population during the listed outage periods is about twice as high as underlying consumption for the sample considered in Doane *et al.* (1988a), and approximately three to four times higher than for households in Woo *et al.* (1991). This relatively pronounced difference may

be attributable to the different geographic location, differing sample characteristics, and changes in electricity consumption during the intervening time periods that separate the three studies.

5. Discussion and Conclusion

We have developed an econometric model that captures the essential features of repeated dichotomous choice non-market valuation data. By using a gamma distribution for the kernel of *WTP*, the model allows for the distribution of *WTP* to have large amount of mass near zero while still constraining WTP to be non-negative. This is crucial for estimating *WTP* for goods that may not be worth much to many households such as preventing a momentary electricity outage. Our model allows for heterogeneity in *WTP* by specifying the scale parameter of the gamma distribution to be lognormally distributed in the population. The lognormal distribution for the scale parameter captures both heteroskedasticity and within-subject correlation in responses to the multiple dichotomous choice valuation questions. This models important features of the data. For example, as shown by the small mean but large variance for a momentary outage, a momentary outage imposes little cost on average but it imposes large costs on *some* households.

It appears that whether deregulation of the retail electricity market continues or not, with higher electricity usage and insufficient construction of new generating capacity in some areas of the country, rational management of the risks of power outages will become more, not less important in the coming years. Rational management requires an understanding of whether the benefits of reliability improving actions outweigh the costs. Given the complexities of managing a transmission and distribution network, it is crucial that utilities or regulators be able to disaggregate their costs and benefits as much as possible so that they can effectively target projects to those that would most benefit. This is critical as in real world applications available budgets are likely to be exhausted before all beneficial projects have been implemented. Understanding how marginal *WTP* behaves as a function of the attributes of an unannounced outage and past outage history are crucial determinants of the relative benefits of mitigating outages.

Our model provides a rich analysis by using a number of covariates that are typically observable by utilities, such as the availability of non-electric heating, or whether the customer lives in a mobile home. The specification of the covariates in exponential form allows for non-linear surfaces describing the *WTP* to avoid different types of outages. Overall, the results conform to expectations based on the outage costs literature and experience. These factors can be used by a utility or a regulatory authority to better target reliability improvements to neighborhoods (or even at the circuit level) that would most benefit. For

example, by conditioning *WTP* on both the number of previous outages the household has experienced and the duration of previous outages we can show that they have differential effects on welfare losses. As illustrated in figure 3.2, ceteris paribus, households that experience one long outage are willing to pay more to prevent *any* kind of future outage than households that have experienced a number of shorter outages. This illustrates another margin on which we can compare the relative benefits of different reliability improvements.

Chapter 4

CAPTURING CORRELATION AND TASTE HETEROGENEITY WITH MIXED GEV MODELS

Stephane Hess
Centre for Transport Studies, Imperial College London and RAND Europe, Cambridge
stephane.hess@imperial.ac.uk

Michel Bierlaire
Institute of Mathematics, École Polytechnique Fédérale de Lausanne
michel.bierlaire@epfl.ch

John W. Polak
Centre for Transport Studies, Imperial College London
j.polak@imperial.ac.uk

Abstract Research in the area of discrete choice modelling can be split into two broad categories; applications accounting for the prevalence of unobserved inter-alternative correlation, and applications concerned with the representation of random inter-agent taste heterogeneity. The difference between these two is however not as clear-cut as this division might suggest, and there is in fact a high risk of confounding between the two phenomena. In this article, we investigate the potential of mixed Generalised Extreme Value (GEV) models to simultaneously account for the two phenomena, using a Stated Preference (SP) dataset for mode-choice in Switzerland. Initial results using more basic modelling techniques reveal the presence of both correlation and random taste heterogeneity. The subsequent use of mixed GEV models on this dataset leads to important gains in performance over the use of the more basic models. However, the results also show that, by simultaneously accounting for correlation and random taste heterogeneity, the scope to retrieve the individual phenomena is reduced. This shows that a failure to account for the potential impacts of either of the two phenomena can lead to erroneous conclusions about the existence of the other

phenomenon. This is a strong indication that the use of mixed GEV models to jointly explain random taste heterogeneity and inter-alternative correlation in a common modelling framework should be encouraged in the case where the nature of the error-structure is not clear a priori.

Keywords: Mixed GEV, random taste heterogeneity, inter-alternative correlation, simulation-based estimation

1. Introduction

Two main streams of model structures can be identified from the existing body of literature on discrete choice models; models concerned with representing the correlation between alternatives in the unobserved utility components, and models concerned with allowing for random variations in tastes across decision-makers.

An appropriate treatment of the correlation structure is crucial especially in the case where a model is used for forecasting of market shares after hypothetical changes to the market structure. In this case, the unrealistic substitution patterns of the Multinomial Logit (MNL) model can lead to very misleading forecasts of demand in the case where heightened correlation exists between some of the alternatives. Acknowledging the potential existence of random taste heterogeneity is similarly important. Indeed, although for reasons of interpretation, it is always preferable to as much as possible attempt to explain the variation in decision-makers' behaviour as a function of socio-demographic characteristics, the limitations of the data (along with inherent randomness involved in decision-making) mean that there is usually some remaining non-quantifiable (random) variation. By not explicitly accounting for such heterogeneity in a model, researchers not only discard valuable information about variations in choice-behaviour, but are also at risk of reaching false conclusions, most notably in the form of biased trade-offs between coefficients (c.f. Hensher and Greene, 2003; Hess and Polak, 2004a).

While the two phenomena of inter-alternative correlation and inter-agent taste heterogeneity have usually been treated in quite separate ways, it should be noted that the differences between these two phenomena are not necessarily that *clear-cut*, and that there is a significant risk of confounding. As an example, in the classic red bus/blue bus problem (c.f. Train, 2003), the correlation in the unobserved utility of the two different bus types could in fact be a reflection of the existence of random taste heterogeneity in the preference for buses. Such random differences would clearly induce correlation in the unobserved utility components. As such, accounting for (arbitrary) correlation in the unobserved utility components without acknowledging the potential effects of random taste heterogeneity can mask the presence of the latter phenomenon. The converse can also be the case; as an example, Hess, Bierlaire &

Polak (2005) have recently shown that the presence of unexplained correlated attributes across alternatives can lead to the erroneous conclusion that there are random variations in tastes across decision-makers.

The discussion presented in this article looks at the issues researchers are faced with in the case of choice scenarios where the two phenomena of unobserved inter-alternative correlation and random inter-agent taste variation potentially both have an effect on decision-making behaviour. It is in this case crucial to disentangle the two effects. The discussion also applies to the case where only one of the two phenomena is present, but where it is not clear *a priori* whether the error term reflects the presence of random taste heterogeneity or *simple* inter-alternative correlation, as caused for example by unobserved shared attributes.

Two different approaches have classically been used in the joint analysis of these two phenomena; the Multinomial Probit (MNP) model (c.f. Daganzo, 1979), and more recently, the Error Components Logit (ECL) formulation of the Mixed Multinomial Logit (MMNL) model (c.f. McFadden and Train, 2000). The MNP model rapidly becomes computationally intractable in the case of complex model structures; the ECL model has similar problems, and can also become difficult to formulate due to important identification issues. In this article, we illustrate the potential of an alternative approach, based on the integration of GEV-style choice probabilities over the distribution of taste coefficients, leading to a mixed GEV model (c.f. Chernew *et al.* 2001, Bhat and Guo, 2004). This model form not only reduces the number of random terms in the models to the number of random taste coefficients, but also avoids some issues of identification that are specific to the *ECL* formulation (c.f. Walker, 2001).

The remainder of this article is organised as follows. In the following section, we give an overview of the theory, looking first at closed-form GEV models, and then at mixed GEV models. Section 3 presents a summary of the empirical analysis conducted to explore the potential of mixed GEV models and to highlight the issues of confounding discussed above. Finally, the fourth section gives a summary of the findings and presents the conclusions of the research.

2. Methodology

A random utility model is defined by a choice set $\mathcal{C}$ containing J alternatives, and a vector of J random utility functions

$$U = \begin{pmatrix} U_1 \\ \vdots \\ U_J \end{pmatrix} = \begin{pmatrix} V_1 \\ \vdots \\ V_J \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_J \end{pmatrix} = V + \varepsilon, \tag{2.1}$$

where U and ε are random vectors and $V \in \mathbb{R}^J$. Each V_i is defined by

$$V_i = f(\beta, x_i), \tag{2.2}$$

where x_i is a vector combining attributes of alternative i and socio-economic attributes of the decision-maker, and β is a vector of (taste-)parameters estimated from the data.

2.1 Closed-form GEV models

The family of Generalised Extreme Value (GEV) models was derived from the random utility paradigm by McFadden (1978). This family of models comprises the basic MNL model (McFadden, 1974), as well as the much-used Nested Logit (NL) model (Williams, 1977; McFadden, 1978; Daly and Zachary, 1978).

In a GEV model, the random vector of variables ε in (2.1) has a Cumulative Distribution Function (CDF) given by

$$F_{\varepsilon_1,\ldots,\varepsilon_J}(x_1,\ldots,x_J) = e^{-G(e^{-x_1},\ldots,e^{-x_J})}, \tag{2.3}$$

which is such that the marginal distribution of the individual ε terms is Gumbel (type I extreme-value). The choice of functional form for the generating function $G()$ determines the correlation structure in place between the individual ε terms, where $G()$ needs to satisfy four main conditions, as set out by McFadden (1978), and later revised by Ben-Akiva and Francois (1983).

The probability of choosing alternative i within the choice set $\mathcal{C}$ for a given decision-maker is given by

$$P(i|V,C) = \frac{y_i G_i(y_1,\ldots,y_J)}{\mu G(y_1,\ldots,y_J)} = \frac{e^{V_i+\log G_i(\ldots)}}{\sum_{j=1}^{J} e^{V_j+\log G_j(\ldots)}}. \tag{2.4}$$

where J gives the number of available alternatives, $y_i = e^{V_i}$, V_i is the deterministic part of the utility function associated with alternative i, and $G_i = \partial G/\partial y_i$. The factor μ is the scale parameter, which, in the absence of separate population groups, is generally constrained to be equal to 1.

With the most basic choice of generating function

$$G(y) = \sum_{j\in\mathcal{C}} y_j^{\mu}, \tag{2.5}$$

we obtain the MNL model, in which the substitution patterns are governed by the Independence of Irrelevant Alternatives (*IIA*) assumption.

The corresponding generating function for an NL model with M nests is given by:

$$G(y) = \sum_{m=1}^{M} \left(\sum_{j \in \mathcal{C}_m} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}, \tag{2.6}$$

where $\mathcal{C}_m$ gives the set of alternatives contained in nest m (with $m = 1, ..., M$), μ_m is the structural parameter for nest m, and where, with this notation, μ_m is constrained to be greater than 1, with the correlation between the unobserved utilities of alternatives sharing nest m being given by $1 - \frac{1}{\mu_m^2}$. In the NL model, the nests are mutually exclusive; each alternative belongs to precisely one nest.

An extension of the NL generating function (equation 2.6) leads to a model form allowing for cross-nesting, whose generating function is given by:

$$G(y_1, \ldots, y_J) = \sum_{m=1}^{M} \left(\sum_{j \in \mathcal{C}_m} (\alpha_{jm} y_j)^{\mu_m} \right)^{\frac{\mu}{\mu_m}}, \tag{2.7}$$

where α_{jm} is the allocation parameter for alternative j and nest m.

The history of cross-nested Logit (CNL) models reaches back to the initial developments of the GEV family; first discussions of this structure were given by Williams (1977) and McFadden (1978). This model form has been used and analysed under different names by a number of authors, including Small (1987), Vovsha (1997), Vovsha & Bekhor (1998), Koppelman & Wen (2000), Wen & Koppelman (2001), Ben-Akiva & Bierlaire (2003), Daly & Bierlaire (2003), Bierlaire (2004), and Papola (2004). CNL models allow for ambiguous allocation of alternatives to nests, hence reflecting the different degrees of similarity between them. There are many problems in which this extra flexibility has the potential to offer considerable improvements, even in the case of a relatively low number of nests or alternatives, as illustrated for example by Bierlaire *et al.* (2001).

2.2 Mixed GEV models

In a *mixed* GEV model, the vector V in equation (2.1) is itself a random vector. In this case, the probability of choosing alternative i within the choice set $\mathcal{C}$ for a given decision-maker is given by

$$P(i \mid \mathcal{C}) = \int_V P(i|V, \mathcal{C}) \mathrm{d}V, \tag{2.8}$$

where $P(i|V, \mathcal{C})$ is defined as in equation (2.4).

Historically, the GEV model used inside the integral in equation (2.8) has been of MNL form, leading to the MMNL model. Two conceptually

different, yet mathematically identical (as illustrated namely by Ben-Akiva and Bierlaire 2003) modelling approaches can arise from this notation; the Random-Coefficients Logit (RCL) model, and the Error-Components Logit (ECL) model.

In the RCL model, some entries of the vector β in equation (2.2) are specified to be random variables, capturing taste heterogeneity in the population. The choice probability of alternative i is then given by:

$$P(i \mid \mathcal{C}) = \int_{\beta} P(i \mid \mathcal{C}, \beta) f(\beta, \theta) \, \mathrm{d}\beta, \tag{2.9}$$

where $P(i \mid \mathcal{C}, \beta)$ is the MNL choice-probability of alternative i, conditional on β, and where θ is a vector of parameters of the distribution of the elements contained in the vector β, giving for example the mean and standard deviation across the population. Recent examples of this approach are given by Revelt & Train (1998), Bhat (2000), Hess & Polak (2004a,b) and Hess, Train & Polak (2004).

In the ECL model, the vector V in equation (2.1) is defined as

$$V = V(\beta, x) + \xi, \tag{2.10}$$

where $V(\beta, x) \in \mathbb{R}^J$ and ξ is a random vector of disturbances. In this case, the error term is composed of two parts, and the utility function is given by

$$U = V(\beta, x) + \xi + \varepsilon, \tag{2.11}$$

where the vector ξ is generally assumed to follow a multivariate Normal distribution, with mean zero and covariance matrix Ω, where Ω is usually constrained to be diagonal (Walker, 2001). By allowing some alternatives to share the same error-components, correlation between these alternatives is introduced into the unobserved part of utility. This approach can thus be used to relax the *IIA* property of the MNL model, and it has been shown that, with an appropriate specification of error-components, the ECL structure can theoretically approximate any random utility model (and thus also any GEV-style nesting structure) arbitrarily closely (McFadden and Train, 2000). Another major advantage of this model structure is that it can be specified so as to allow for heteroscedasticity. For recent applications of the *ECL* formulation, see for example Bhat (1998) and Brownstone & Train (1999).

The two approaches (RCL and ECL) can be combined straightforwardly, allowing for the joint modelling of random taste heterogeneity and inter-alternative correlation. However, while the MMNL model is very flexible (and more so than the MNP model), important issues of identification need to be dealt with in the specification of the error-component structure (Walker, 2001). Furthermore, although the MMNL model has the theoretical property of being

able to approximate other random utility models arbitrarily closely, this may not always be as straightforward in practice (Garrow, 2004). Finally, depending on the correlation structure, the high number of error-components required can lead to high simulation costs. Indeed, the integral in equation (2.8) does not generally have a closed form, and numerical techniques, typically simulation, are required during the estimation and application of MMNL models (and Mixed GEV models by extension). The development of ever more powerful computers and recent improvements in the efficiency of simulation techniques (Bhat, 2001, Hess *et al.* 2003, Hess, Train and Polak, 2004) have significantly reduced the computational overheads of this process, and the number of applications using the RCL model especially has increased rapidly over recent years. Nevertheless, the computational cost of estimating and applying mixed GEV models remains high, when compared to their closed-form counterparts.

While integration over mixing distributions is necessary in the representation of random taste heterogeneity, this is not strictly the case for inter-alternative correlation. Indeed, just as, conditional on a given value of the taste-coefficients, a model allowing for random taste heterogeneity reduces to an MNL model, a model allowing for inter-alternative correlation in addition to random taste heterogeneity can in this case be seen to reduce to a given GEV model (assuming that an appropriate GEV model exists). As such, the correlation structure can be represented with the help of a GEV model, while the random taste heterogeneity is accommodated through integration over the assumed distribution of β. The use of a more complicated GEV model as the integrand leads to a more general type of a mixed GEV model, of which the RCL model is simply the most basic form. Applications of this approach include, for example, Chernew *et al.* (2001) and Bhat & Guo (2004). In such a mixed GEV model, the number of random terms, and hence the number of dimensions of integration (and thus simulation) is limited to the number of random taste coefficients, whereas, in the ECL model, one additional random term is in principle needed for the representation of each separate nest. It should be noted that the potential runtime-advantage resulting from this difference in dimensions of integration only manifests itself beyond a certain number of nests, as the more complicated form of the integrand in mixed GEV models initially gives the ECL model a computational advantage. The use of mixed GEV models does however have another advantage over the use of the ECL model in that it avoids the issues of identification that are specific to this latter model form, although additional GEV-specific formulation and identification issues apply.

Finally, it should be noted that while the error-components method has historically only been used with an MNL model as the basis, the approach can theoretically also be used when ε is GEV distributed, for example in the case where some correlation is to be captured by the GEV structure, with a remaining amount of correlation (or indeed heteroscedasticity) to be explained by the

error-components. This can be useful in the case where existing GEV structures are incapable of capturing the full array of correlation in the data (GEV models are homoscedastic and do not allow to capture all types of correlation structure, c.f. Abbé, 2003), while the exclusive reliance on error-components would lead to excessive computational cost or issues of identification. This approach would thus lead to an error-component GEV model. In this article, we concentrate on the use of the random-coefficients GEV model, the analysis of the potential of advanced error-components GEV models (not based on MNL) is an important area for further research.

3. Empirical analysis

The data used for our empirical analysis form part of the survey data collected to estimate the hypothetical demand for a new high-speed transit system in Switzerland; the Swiss Metro (Abay, 1999; Bierlaire et al., 2001). The aim is to build a mag-lev underground system operating at speeds up to 500 km/h in partial vacuum, connecting the major urban centres along Switzerland's *Mittelland* corridor; St. Gallen, Zurich, Bern, Lausanne and Geneva[1]. Aside from the problems of funding, technological feasibility and commercial viability, there is an important question about the impact that the development of such a system would have on the environment. Even though the construction of the Swiss Metro (SM) is thus rather unlikely in the near future, the data collected to estimate the demand for the system can give important insights into respondents' evaluation of hypothetical choice alternatives in general, and transport modes in particular. Furthermore, the SM alternative can be seen as a proxy for a high-speed rail alternative; in the face of increasingly congested roads and skies, the analysis of the potential demand for such advanced public transport modes is a topic of great interest.

A combined Revealed/Stated Preference (RP/SP) approach was used to collect the data (Abay, 1999). Initial interviews about a specific observed trip were followed by a set of SP experiments based on this specific trip, where both car-travellers and rail-travellers were used in the survey. The SP surveys comprised 9 hypothetical choice scenarios, using the three alternatives of car, rail and SM, where car was only available to car-owners. The main explanatory variables used to describe the alternatives were travel-time, cost/fare and headway (for train and SM alternatives). Two different seating arrangements were used for SM alternatives, corresponding to *1^{st} class rail-travel*, and *business class aircraft seats*. Fares for SM services were obtained by multiplying rail fares by a factor of 1.2, while car running costs were set to 1.20 CHF/km.[2]

[1] For details, see www.swissmetro.com
[2] 1 CHF$\approx$ \$0.8

The aim of the present article is to illustrate the potential of Mixed GEV models in practice, rather than making policy implications *per se*. As such, only the SP survey was used, whereas a more policy-oriented analysis would have had to make use of the combined RP/SP survey. Also, the potential scale differences in the error-term between car users and train users were not directly taken into account, where a treatment of these differences would again have been important in a more policy-oriented analysis. A separate analysis revealed some differences in scale between the two groups; allowing for these differences did however not significantly affect the conclusions with regards to the nesting structure or the presence of random taste heterogeneity. Software limitations meant that it was not possible to jointly accommodate scale differences and correlation across repeated choice observations; the latter phenomenon was in this case judged to be more important (c.f. section 3.4). Finally, it should be noted that the sample used in this analysis can be seen as being choice-based (given the selection of respondents on the basis of RP choices). As it was not possible to properly take into account the effects of sampling (the population weights were clearly only known for the two existing modes), the results of this analysis must be seen as applying to the present sample only.

Only commuters and business travellers were included in the analysis, and no distinction was made between these two groups of travellers at this stage, leading to a sample size of 6,870 observations. The explanatory variables used in the model fitting exercise included cost, travel-time, and headway. Additionally, the impacts of seating arrangements for SM, age for rail travellers (divided into 5 roughly equally sized discrete groups), and season ticket ownership for rail-based alternatives were taken into account in the model. While separate travel-time coefficients coefficients were used for the three different modes, it was not possible to identify significantly different cost-coefficients for the three modes. Similarly, the differences between the estimated season ticket constants for rail and SM were not significant, such that a common coefficient was used. Attempts to account for possible further interactions between socio-demographic variables and taste coefficients were not successful. Additionally, some effort went into experimenting with non-linear specifications for the marginal utilities of the various explanatory variables; however, this did not lead to any significant gains in model performance. For reasons of identification, the ASC of rail was set to zero, in all model types used in the analysis. No significant random heterogeneity was identified in any of the models for either of the three ASCs, such that the three possible normalisation approaches are equivalent in the mixed models, just as they are in the closed-form models. Finally, aside from the season ticket variable, other inertia variables, such as car ownership, license holding and past choices can be expected to have a sig-

nificant effect on choice behaviour; the analysis of these effects was however beyond the scope of the present analysis.

For the calibration of the various models discussed in this article, the estimation software BIOGEME (Bierlaire, 2003) was used[3]. This estimation tool can be used for all types of closed-form as well as Mixed GEV model structures. Furthermore, the program can accommodate non-linear utility functions, and the estimation can be performed so as to account for correlation across repeated choice observations for the same individual.

When estimating models based on mixing distributions, it is of interest to attempt to minimise the computational overhead of the calibration process (Hess et al., 2004). This is especially crucial in the case of mixed GEV models that are based on a more complicated integrand than the simple MNL formula. One such improvement that can lead to important gains in simulation efficiency is the use of quasi-random number sequences instead of pseudo-random number sequences as the basis of the simulation process (Train, 2003). In the present application, one such quasi-random approach, known as the Halton sequence (Halton, 1960), was used in conjunction with an iterative drawing procedure. This procedure is based on the notion that the first few iterations of the maximisation process are *rough* steps in the general direction of the maximum of the log-likelihood function, requiring a lower degree of precision in the simulation. As such, a comparatively lower number of draws can be used for these initial steps, leading to important reductions in computation time. To this extent, the model was first estimated to a preset convergence level using a very low number of draws. This number of draws was then increased, and the final estimates from the preceding run were used as starting values. This process was repeated until the preset number of 1,000 draws (per dimension of integration, and per individual) was reached; a sensitivity analysis showed this to be sufficient to obtain stable estimates. At each step in this iterative process (increase in the number of draws), the sequences of Halton draws were newly generated, so as to obtain as uniform a spread as possible with the given number of draws used in a specific run. A trust-region algorithm was used in the estimation, and at each step, a more stringent convergence criterion was used. Overall, this approach is very similar to that proposed by Bastin (2004), except that in our approach, the change in the number of draws is controlled externally (and set prior to estimation), rather than being controlled internally. Furthermore, the approach of Bastin (2004) allows for occasional decreases in the number of draws during the estimation process.

[3]The estimation software, together with examples, and documentation, is available from http://roso.epfl.ch/biogeme; the data and model files for the application presented in this article are available from http://roso.epfl.ch/mbi/biogeme/swissmetro

3.1 Multinomial logit model

As a basis for comparison, a simple MNL model was first fitted to the data; the estimation results for this model are reported in the first part of table 4.1. As expected, the results show negative marginal utilities for increases in travel-time on all three modes, with similar conclusions for cost and headway increases. The model further shows that older people are relatively more likely to choose rail, while season-ticket holders are more likely to choose rail and SM (when compared to car). Finally, the results show that in terms of the seating arrangements for SM, respondents have a preference for first-class rail seats over business class aircraft seats. In terms of the implied willingness to pay for travel-time reductions, the results show significantly higher values of travel-time savings (VTTS) for rail, while the value for SM is only marginally higher than that for car.

3.2 Nested Logit model

To account for the potential existence of heightened correlation between some of the alternatives, three separate NL structures were estimated on the data; grouping together car and rail, car and SM, and rail and SM respectively. Only the nesting of car and rail, i.e. the grouping of existing modes versus the hypothetical SM alternative, resulted in a structural parameter that is greater than 1 (using the notation from section 2.1). The results of this estimation are shown in the second part of table 4.1[4]. With this model structure, the nesting parameter (μ_{CR}) takes a value of 2.23, implying a high correlation between the unobserved utilities of the car and rail alternatives of around 0.8. Aside from a difference in scale, the substantive results of the two models are very similar, although the VTTS measures are lower than in the corresponding MNL model, especially so for the car and rail alternatives. This also implies that the results show a clearer difference between the VTTS for car and SM. Finally, in terms of model fit, the results show a very significant increase in Log-Likelihood (LL) by 122.42 units, with one additional parameter. This leads to a likelihood-ratio test value of 244.84, which has an associated χ^2_1 p-value that is identical to zero (3^{-55}).

3.3 Cross-Nested Logit model

As reported in section 3.2, significant correlation could only be retrieved between the car and rail alternatives, leading to a nesting of existing versus hypothetical alternatives. It is however conceivable that such correlation also exists between the rail and the SM alternatives, given that they have the com-

[4]The t-test for μ_{CR} is expressed with regards to a base-value of 1.

Table 4.1. Estimation results for MNL, NL and CNL models

	MNL		NL		CNL_A		CNL_B	
Parameter	*Estimate*	*Asy. z-value*	*Estimate*	*Asy. z-value*	*Estimate*	*Asy. z-value*	*Estimate*	*Asy. z-value*
ASC Car	0.5731	2.88	0.3013	2.56	-0.3472	-3.93	-0.2832	-2.71
ASC SM	0.8661	4.89	0.5965	4.31	0.0071	0.09	0.047	0.42
Age for Rail	0.2639	6.15	0.178	6.3	0.1538	9.17	0.1652	6.71
Airline seats for SM	-0.435	-4.35	-0.2743	-3.05	-0.1725	-3.65	-0.2309	-3.61
Season ticket for Rail/SM	0.9718	5.31	0.7595	7.17	0.6893	7.75	0.6715	6.92
Headway (minutes)	-0.0055	-5.55	-0.0037	-5.5	-0.0028	-5.94	-0.0031	-5.74
Cost (CHF)	-0.0098	-13.98	-0.0071	-11.75	-0.0065	-14.91	-0.007	-13.53
Travel time Car (minutes)	-0.0116	-10.66	-0.0077	-6.67	-0.0062	-10.98	-0.0067	-5.64
Travel time Rail (minutes)	-0.0152	-14.03	-0.0106	-10.13	-0.0079	-12.56	-0.0089	-8.7
Travel time SM (minutes)	-0.0119	-6.55	-0.0085	-5.06	-0.0067	-9.62	-0.0073	-5
μ_{CR}	-		2.23	7.76	2.78	9.27	2.67	7.48
μ_{SR}	-		-		5.91	6.61	4.68	6.9
$\alpha_{R,CR}$	-		-		0.4293	3.76	0.5	-
$\alpha_{R,SR}$	-		-		0.5707	3.76	0.5	-
VTTS Car (CHF/hour)	71.02		65.07		57.23		57.43	
VTTS Rail (CHF/hour)	93.06		89.58		72.92		76.29	
VTTS SM (CHF/hour)	72.86		71.83		61.85		62.57	
Final LL	-5328.24		-5205.82		-5162.03		-5167.12	
$\rho^2(0)$	0.2471		0.2644		0.2706		0.2698	

mon aspect of being public transport modes. To test for the presence of such correlation, a CNL model was fitted to the data, allowing the rail alternative to belong to a rail-SM nest as well as to the car-rail nest. The results of this estimation process are reported in the third part of table 4.1 (CNL$_A$)[5]. The results show that in addition to high correlation between the unobserved utilities of the two existing modes of car and rail, there is also very high correlation between the unobserved parts of the utilities for rail and SM. The allocation parameters $\alpha_{R,CR}$ and $\alpha_{R,SR}$ show the degree of membership of the rail alternative to the nests it shares with car and SM respectively, where the estimates are very similar, with slightly higher allocation to the public transport nest.

The CNL model reduces to the NL model described in section 3.2 when $\alpha_{R,CR} = 1$ and $\alpha_{R,SR} = 0$. In this scenario, the nesting parameter μ_{SR} becomes obsolete. When μ_{CR} further becomes equal to 1, the model reduces to the MNL model described in section 3.1. Likelihood-ratio tests can thus be used to compare the CNL model to the MNL and NL models, with 3, respectively 2 degrees of freedom (only one α is actually estimated, given that $\alpha_{R,CR} = 1 - \alpha_{R,SR}$). The resulting likelihood ratio test values are 332.42 and 87.58, both of which have p-values that are indistinguishable from zero, for χ^2_3 and χ^2_2 tests respectively. This shows that important gains in model fit can be obtained by accounting for the correlation between the two public transport alternatives; interestingly, this was not possible in the NL model, suggesting that this correlation can only be explained simultaneously with the correlation between the car and rail alternatives.

In terms of substantive results, the estimated coefficients are again all of the expected sign. However, the implied VTTS measures are significantly lower than those reported with the MNL and NL structures, where a similar observation can be made for the willingness to pay for headway reductions. This is the result of an increase in the relative weight of the marginal utility of cost when compared to the MNL and NL structures, and shows the impact of model structure on the relative scale of the various coefficients. This thus suggests that, by accounting for the correlation structure, the cost attribute gains in weight when compared to the other attributes.

The interpretation that should be given to the allocation parameters in a CNL model is not clear, although intuitive interpretations exist for example in the case of route-choice. In the present application, the allocation of the rail alternative was split almost evenly between the car-rail and the rail-SM nest. To establish the impact of these parameters, the model was re-estimated, with both allocation parameters constrained to a value of 0.5. The results of this process are reported in the fourth part of table 4.1 (CNL$_B$). The use of constrained al-

[5]The t-tests for μ_{CR} and μ_{SR} are expressed with regards to a base-value of 1, while for $\alpha_{R,CR}$ and $\alpha_{R,SR}$, a base-value of 0.5 is used.

location parameters leads to a slight drop in the estimated correlation in the two nests. Furthermore, it leads to a 4.6% increase in the estimated VTTS for the rail alternative. Aside from these two changes, the substantive impacts of the additional constraint are relatively minor. The constraint leads to a statistically significant drop in LL by 5.09 units, equating to a likelihood ratio test-value of 10.18, with an associated p-value of 0.0062.

To conclude this section, it should be noted that similar experiments where conducted with a structure allowing car to belong to a car-rail and a car-SM nest; no extra correlation between car and SM could however be identified.

3.4 Mixed Multinomial Logit model

As discussed in the introduction to this article, not allowing for potential random variations in tastes across respondents puts researchers at risk of producing seriously biased results. With this in mind, several experiments were conducted to explore the potential prevalence of random taste heterogeneity in the population of decision-makers. The repeated choice nature of the data was taken into account in these experiments, such that tastes vary across individuals, but not across observations for the same individual (Train, 2003). This leads to efficient estimates, whereas the purely cross-sectional leads only to consistent estimates. Attempts were also made to accommodate other SP panel effects, such as inertia, but none of these was found to have a significant effect.

Significant random taste heterogeneity was identified for five coefficients; the three travel-time coefficients, in addition to the dummy coefficients for age for rail users, and for seating type for SM users. For reasons of simplicity, a Normal distribution was used for all five coefficients. This is a valid assumption for the two dummy coefficients, but can lead to problems with the three travel-time coefficients. Indeed, by using a Normal distribution, researchers in effect make an a priori assumption that the coefficient takes a positive value for some of the respondents. The use of bounded distributions is in this case preferable (7, Hess, Bierlaire and Polak, 2005). However, in the present application, the Normal distribution led to very good performance, while problems in estimation were encountered when using alternative distributions. Furthermore, with the estimated distributional parameters, the probability of a wrongly signed coefficient was always at an acceptable level.

The results of the estimation are summarised in the first part of table 4.2 (MMNL$_A$). The first observation that can be made is that the MMNL model leads to an improvement in LL over the MNL model by 229.63 units, with 5 additional parameters. This equates to a likelihood-ratio test-value of 459.26, giving a χ^2_5 p-value of 0. This illustrates the important gains in model fit that result from accommodating random variations in respondents' tastes. The re-

sults further show that the effect of using a Normal distribution for the three travel-time coefficients is benign, with probabilities of a *wrongly* signed coefficient of 1%, 0% and 2% for car, rail and SM respectively. Finally, it should be noted that, with the MMNL model, the estimates of the two ASCs, as well as that of the mean for the age-dummy for rail-travellers, are not significant at the usual 95% level of confidence.

In terms of actual estimation results, the model shows that, while age still has a positive mean effect on the utility of the rail alternative, for about 30% of respondents, this effect is now negative. Tests with bounded distributions led to poor results, suggesting that these results do indeed signal the existence of travellers for which this dummy variable is negative, rather than being simply an effect of using the Normal distribution. A similar observation can be made for the coefficient associated with the type of seating, where the results now indicate that almost 42% of travellers have a preference for aircraft-type business-class seats over first-class rail-seats. These results illustrate the potential of the MMNL model; the closed-form models falsely suggest a consistent positive effect of age and rail-type seats across the population.

In terms of the implied willingness to pay for travel-time reductions, the results show consistently higher VTTS measures for all three modes than was the case in the closed-form models. This shows the important bias that can result from not accounting for random variations in the coefficients involved in trade-off calculations. Although it was not possible to estimate such a coefficient in the present analysis, it should be stressed that the risk of bias becomes even greater in the case of a randomly distributed cost-coefficient. Again, like in the MNL model, the estimated VTTS for car and SM are very similar, while the corresponding measure for rail is significantly higher. It is important to note that the use of fixed coefficients not only leads to a risk of biased results, but also leads to a loss of all information about the variation in the VTTS across respondents. The standard deviations reported in table 4.2 for the travel-time coefficients are very high, and lead to very wide confidence intervals for the VTTS. As an illustration, the lower and upper 80% quantiles were calculated, leading to lower limits of 42.13, 82.83, and 36.08 CHF/hour for car, rail and SM respectively, with corresponding upper limits of 145.12, 144.32 and 156.10 CHF/hour respectively. This shows that while rail has got the highest associated mean VTTS, it has the narrowest confidence interval, followed by car and SM. The variation in the VTTS for SM is so important that, while the mean VTTS for SM lies in between those for car and rail, its lower and upper limits are more extreme than those of car and rail respectively. This could be seen as a reflection of the uncertainty involved with the evaluation of a hypothetical mode.

Table 4.2. Estimation results for MMNL, Mixed NL and Mixed CNL models

	MMNL$_A$		Mixed NL		MMNL$_B$		Mixed CNL	
Parameter	*Estimate*	*Asy. z-value*	*Estimate*	*Asy. z-value*	*Estimate*	*Asy. z-value*	*Estimate*	*Asy. z-value*
ASC Car	0.2758	0.92	0.1807	0.89	0.3268	1.11	–0.2222	–1.71
ASC SM	0.4455	1.63	0.3765	1.81	0.49	1.81	–0.0196	–0.16
Age for Rail (μ)	0.1382	1.4	0.0869	1.29	0.0496	0.53	0.1612	5.94
Age for Rail (σ)	0.2579	6.25	0.2553	7.43	0.3064	8.37	-	-
Airline seats for SM (μ)	–0.3142	–2.09	–0.3376	–2.82	–0.3374	–2.17	–0.2247	–3.54
Airline seats for SM (σ)	1.5553	6.48	0.7717	2.23	1.6191	6.94	-	-
Season ticket for Rail/SM	1.2126	3.66	0.8555	4.42	1.3097	3.88	0.6099	4.56
Headway (minutes)	–0.0072	–5.48	–0.0052	–5.3	–0.0071	–5.43	–0.0031	–5.94
Cost (CHF)	–0.0138	–11.02	–0.0103	–9.07	–0.0139	–11.06	–0.0076	–11.05
Travel time Car (minutes) (μ)	–0.0215	–14.99	–0.016	–11.1	–0.0216	–14.5	-0.0089	–8.77
Travel time Car (minutes) (σ)	0.0092	11.25	0.0074	4.1	0.0094	7.36	0.0039	8.10
Travel time Rail (minutes) (μ)	–0.0261	–15.13	–0.0175	–12.18	–0.0229	–15.09	–0.0100	–10.28
Travel time Rail (minutes) (σ)	0.0055	6.65	-	-	-	-	-	-
Travel time SM (minutes) (μ)	–0.0221	–13.41	–0.0175	–11.89	–0.0221	–13.43	–0.0086	–6.32
Travel time SM (minutes) (σ)	0.0108	8.77	0.0104	8.07	0.012	10.37	-	-
μ_{CR}	-	-	2.28	5.08	-	-	4.10	4.27
μ_{SR}	-	-	-	-	-	-	4.26	6.03
VTTS (CHF/hour)	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Car	93.48	40	93.2	43.11	93.24	40.58	70.25	30.45
Rail	113.48	23.91	101.94	-	98.85	-	78.96	-
SM	96.09	46.96	101.94	60.58	95.4	51.8	67.45	-
Final LL	–5098.61		–5026.23		–5105.77		–5086.16	
$\rho^2(0)$	0.2795		0.2898		0.2785		0.2813	

3.5 Mixed Nested Logit model

The results in sections 3.2, 3.3 and 3.4 have shown that important gains in model performance can be obtained both by accounting for the presence of inter-alternative correlation in the unobserved utility terms, and by allowing for a random distribution of tastes across decision-makers. However, as highlighted in the introduction and theoretical part of this article, it is not clear a priori whether these results actually signal the presence of separate phenomena, or whether the two approaches simply explain the same phenomenon in different ways. The aim was now to attempt to jointly model the two phenomena, hence reducing the risk of confounding. For this, a mixed NL model was fitted to the data.

Whereas, with the MMNL model described in section 3.4, it was possible to retrieve significant random variation for five taste coefficients, this number was reduced to four in the mixed NL model. Indeed, the standard deviation associated with the marginal utility of travel time for rail alternatives was no longer statistically significant at any reasonable level of significance. This was already the coefficient with the smallest variation in the MMNL model (c.f. table 4.2), and by accounting for inter-alternative correlation, the error-term in the model decreases, reducing the scope for retrieving random taste heterogeneity further. This signals possible confounding in the simple MMNL model presented in section 3.4. The final estimates for the mixed NL model are reported in the second part of table 4.2[6]. The results show that, compared to the NL model reported in table 4.1, the use of the mixed NL model leads to a gain in LL by 179.59 units, with 4 additional parameters. This equates to a likelihood ratio test of 359.18, with an associated χ^2_4 p-value of 0. Similarly, the mixed NL model leads to an improvement in LL by 72.38 units over the MMNL model from section 3.4. To allow for the use of a nested log-likelihood ratio comparison between the mixed NL and MMNL structures, the MMNL model from section 3.4 was re-estimated with a fixed coefficient for rail travel-time. The results of this re-estimation are reported in the third part of table 4.2, showing that, as expected, the use of a fixed coefficient leads to a significant drop in LL by 7.16 units. The use of the mixed NL model leads to a highly significant improvement in LL by 79.54 units when compared to this re-estimated MMNL model, with a single additional parameter. These results reflect the importance of jointly accommodating the two phenomena of correlation and random taste heterogeneity.

In terms of actual estimation results, the values in table 4.2 show that the mixed NL model retrieves a correlation structure between car and rail alternatives that is virtually indistinguishable from that obtained when using the

[6] Again, the t-test for μ_{CR} is expressed with regards to a base-value of 1.

simple NL model reported in table 4.1. However, the significance level of the nesting parameter is markedly lower. A similar observation can be made for all but one of the standard deviations of the randomly distributed coefficients (when compared to the two MMNL models). This drop in significance levels is to be expected, given that the mixed NL model decomposes the error-term further than the NL and MMNL models. It can also be noted that in the mixed NL model, the mean VTTS measure for rail and SM are now indistinguishable, whereas, in the NL and MMNL models, the VTTS for rail was markedly higher. This could be seen as an effect of using a fixed travel-time coefficient for rail, when compared to the MMNL model; however, the re-estimated MMNL model uses the same restriction, yet still yields a slightly higher mean VTTS for rail than for SM. Any other remaining differences between the two models are largely down to a difference in scale.

3.6 Mixed Cross-Nested Logit model

The final model fitted during the analysis was a mixed CNL model, using the same nesting structure as the CNL model described in section 3.3. In section 3.5, we observed that, by accounting for the correlation between the car and rail alternatives, the scope for retrieving significant amounts of random taste heterogeneity is reduced. When fitting the mixed CNL model, serious estimation problems were encountered. These related specifically to the ability to retrieve random taste heterogeneity, especially when also accounting for the repeated choice nature of the dataset. These problems reflect the complexity of the model, but could also be a sign of a lack of explanatory power in the data, in such that the error-term cannot be partitioned enough to reproduce a mixed CNL structure with a high number of random taste coefficients. Eventually, it was possible to estimate a mixed CNL model with a single randomly distributed taste coefficient, namely the marginal utility of travel-time for the car-alternative. For estimation purposes, the allocation parameters were both constrained to be equal to 0.5. The results of this estimation process are reproduced in the fourth part of table 4.2.[7]

The first observation that can be made from table 4.2 is that, with one additional parameter, the mixed CNL model leads to a very significant improvement over the constrained CNL model reported in the second part of table 4.1; the difference in LL is 80.96, leading to a likelihood-ratio test-value of 161.92, which has an associated χ^2_1 value of zero. This shows that even a single randomly distributed taste coefficient leads to important gains in explanatory power. The mixed CNL model also has a higher LL than the two MMNL models, although no nested likelihood-ratio test can be performed for these

[7]The t-tests for μ_{CR} and μ_{SR} are again expressed with regards to a base-value of 1.

differences. On the other hand, the LL of the mixed CNL model is inferior to that of the mixed NL model by 59.93 units. This, in combination with the NL results, suggests that the increased partitioning of the error term resulting from allowing for cross-nesting adds less explanatory power than the partitioning resulting from accounting for the additional levels of random taste heterogeneity in the mixed NL model. Efforts to explain a larger part of the error term by accounting for further levels of taste heterogeneity in a mixed CNL framework are ongoing.

The other main observation that can be made from table 4.2 is that, while the VTTS measures produced by the mixed CNL model are closer in scale to those produced by the closed-form models than those produced by the other mixed models, the VTTS of SM is now lower than the mean VTTS of the car alternative. This can however be seen as an effect of using a randomly distributed coefficient for the marginal utility of travel-time for car, while a fixed coefficient is used for the other two modes. Finally, it should be noted that, while the estimated value for μ_{SR} is very similar to that obtained with the constrained CNL model in section 3.3, the value estimated for μ_{CR} is markedly higher. This shows that the estimation of the structural parameters is affected by the use of a utility function containing random coefficients. This in turn again suggests some interaction between the part of the error-term linked to random taste heterogeneity and the part linked to inter-alternative correlation.

4. Summary and Conclusions

In this article, we have discussed the issues arising with model specification in the case of a non-trivial error-structure. We have focussed on two separate ways of partitioning the error-term; accounting for (arbitrary) correlation between alternatives in the unobserved utility components, and allowing for a random distribution of tastes across decision-makers. The theoretical discussions presented in this article have highlighted the fact that the distinction between these two phenomena is not clear-cut, and that there exists a significant risk of confounding in the case where researchers account for only one of the two phenomena.

Our empirical analysis has shown that while it is possible to separately model the prevalence of correlation in the choice-set and random taste heterogeneity in the population of decision-makers, and while both approaches lead to very significant gains in model fit, the joint modelling of these two phenomena can be more problematic. Indeed, while the mixed NL model described in section 3.5 retrieves a near identical nesting structure to that obtained with the simple NL model in section 3.2, random taste heterogeneity can only be retrieved for four taste coefficients, as opposed to five in the simple MMNL model (c.f. section 3.4). Even more severe problems were encountered when

using a mixed CNL model, where random taste heterogeneity could only be retrieved for a single coefficient. Although, in the mixed CNL model, these problems were at least party due to model complexity, the overall results do highlight the issue of confounding of taste heterogeneity and correlation, complementing similar observations made by Cherchi and Ortuzar (2004) with regards to the ECL model. It should also be noted that the various model fitting exercises described in this article have highlighted the fact that the assumptions made with regards to the error-structure can have significant impacts on substantive results, such as willingness-to-pay indicators.

It should be stressed that the failure to simultaneously account for all heterogeneity and correlation should not be seen as a deficiency of the model, but rather as a sign that the error-term in the model has decreased. Indeed, by accounting for either of the two phenomena, the modeller explains processes that take place in the unobserved part of utility of the alternatives. This is analogous to the case where the specification of the utility function in the most basic of discrete choice models is improved by the inclusion of more explanatory variables. If it were possible to improve the utility specification to the point were all correlation across alternatives is explained in the observed part of utility, the errors would become independent, and it would no longer be possible to explain inter-alternative correlation with the help of a nesting structure. As such, it can often be observed that, while inter-alternative correlation can be retrieved in models using a very basic specification of the observed utility, further refinement of the utility function will lead to problems with retrieving significant nesting effects. This should clearly be seen as desirable, as any correlation is now explained in a deterministic way, through the observed utility function. A similar process occurs in models jointly allowing for random taste heterogeneity and correlation. When only allowing for either of the two phenomena in a model, the impact of the unrepresented phenomenon will at least be partly carried over into the other phenomenon. This in turn shows that, by simultaneously accounting for the two phenomena, the scope for retrieving apparent significant effects of either of the two phenomena is reduced. On the other hand, this however also means that the risk of falsely explaining correlation by random taste heterogeneity, or vice-versa, is reduced. As such, researchers should always strive to simultaneously account for the potential prevalence of both random taste heterogeneity and unexplainable inter-alternative correlation, in the case where the observed utility function is incapable of explaining sufficient amounts of choice behaviour for the remaining error-term to be distributed purely iid type I extreme-value.

Acknowledgements

The authors would like to express their thanks to Georg Abay and Kay Axhausen for making the Swiss Metro data available for the present analysis. The authors would also like to thank Elisabetta Cherchi, Andrew Daly, Laurie Garrow and Kenneth Train for helpful comments on an earlier version of this article.

Chapter 5

ANALYSIS OF AGRI-ENVIRONMENTAL PAYMENT PROGRAMS

A Joint Framework

Joseph Cooper *
Deputy Director
Resource Economics Division,
Economic Research Service, USA.
jcooper@ers.usda.gov

Abstract The chapter presents an approach for simultaneously estimating farmers' decisions to accept incentive payments in return for adopting a bundle of environmentally benign best management practices. Using the results of a multinomial probit analysis of surveys of over 1,000 farmers facing five adoption decisions in a voluntary program, we show how the farmers' perceptions of the desirability of various bundles change with the offer amounts and with which practices are offered in the bundle. We also demonstrate an estimator for the mean minimum willingness to accept for adoption of a practice conditional on the cost share offers for other practices.

Keywords: best management practices, EQIP, incentive payments, multinomial probit, maximum simulated likelihood estimation, simulated multivariate normal, WTA.

1. Introduction

Agri-environmental payment programs play an important part in improving the environmental performance of agriculture (Claassen and Horan 2000; Batie 1999; Lynch and Smith 1994; Smith 1992; Feather and Cooper 1995; Claassen *et al.* 2001). Federal-level interest in developing these programs is currently strong. For example, the 2002 Farm Act called for a five-fold increase in funding for the USDA's Environmental Quality Incentives Program (EQIP). This chapter focuses on voluntary programs designed along the lines of the EQIP, which provides incentive payments to encourage producers to adopt environ-

*The views expressed are the author's and do not necessarily represent policies or views of the Economic Research Service or the U.S. Department of Agriculture.

mentally benign land management practices such as nutrient management, manure management, and integrated pest management.

For policymaking purposes, it would be useful to know the sensitivity of the producer's decision to enroll in response to a schedule of potential incentive payments and to which practices are bundled together. Such information can be used to assess the costs of encouraging farmers to try various environmentally benign management practices (commonly known as Best Management Practices, or BMPs).

EQIP offers the farmer a suite of BMPs to choose from. Existing published research (Cooper and Keim, 1996) modeled the probability of farmer adoption of BMPs as a function of the incentive payment, with each practice being modeled independently in a bivariate probit analysis of actual adoption and hypothetical adoption. In Cooper and Keim, the bivariate normal was used to jointly model actual and hypothetical use of the same practice as part of a self-selection procedure. Khanna (2001) also conducts a bivariate probit analysis of technology adoption, but between two technologies at a time.

There is no compelling reason to assume that the farmer's decisions to adopt each of these practices should be independent of each other; these BMPs should be considered as a bundle of inter-related practices (Amacher and Feather, 1997). If each adoption decision is treated independently in estimation, then valuable economic information may be lost. If the available set of BMP options does indeed influence the farmer's decision as to which practices to adopt, then the set of adoption decisions follow a multivariate distribution. The multinomial probit (MNP) model, which makes use of the multivariate normal (MVN) distribution, is the appropriate econometric tool for modeling multiple adoption decisions in a joint fashion such that the correlations of the error terms across the practices are nonzero.

In the numerical illustration, a dataset drawn from surveys of over 1,000 farmers in four U.S. regions is used to simultaneously model five discrete choices in an EQIP-like cost sharing program. This program offers cost-shares only for practices that the farmer does not currently use. In the model presented here, farmers who do not use a desired practice are asked whether or not they would accept a hypothetical cost share offer to adopt the practice, and each hypothetical adoption decision is treated jointly. By modeling the decision making process jointly across the offered BMPs, the resulting estimate of the correlations across the decisions allows us to examine which BMPs the farmers consider to be bundles, and to calculate conditional probabilities and summary statistics. This information can be of policy significance in the design of the type of agri-environmental payment program discussed here. Before turning to the econometric model and then to the numerical illustration of the approach, in the next section we provide the theoretical basis for addressing the incentive payment program as a bundle of technologies to be adopted.

2. The Theoretical Model

Consider a farmer who is faced with a set of decisions on what combination of $j = 1, \ldots, J$ BMPs to choose from under a incentive payment program. The farmer's discrete decision to accept incentive payments in exchange for adopting the BMPs can be modelled using the random utility model (RUM) approach (e.g. Hanemann, 1984).[1] From the utility theoretic standpoint, a farmer is willing to accept a cost share A_j per acre to switch to a new BMP j if the observable portion of the farmer's indirect utility with the new practice and incentive payment, $V_{lj}(\mathbf{s}, A_j, \varepsilon_1; \theta)$, is at least as great as at the initial state, $V_0(\mathbf{s}, \varepsilon_0; \theta)$, i.e., the farmer's decision to adopt the practice can be expressed as $V_{lj \geq V_0}$ where 0 is the base state, 1 is the state with the green practice j adopted, s is a vector of explanatory variables, and θ is a vector of the parameters of the functions. Say that C_j is the cost share value that solves $V_{1j}(\mathbf{s}, C_j, \varepsilon_1; \theta) = V_0(\mathbf{s}, \varepsilon_0; \theta)$, then $C_j = C(\mathbf{s}, \varepsilon; \theta)$ is the minimum willingness to accept (WTA) for adopting green practice j. Assume, as is commonly done, that V is a simple linear functional form,$V = \mathbf{s}\theta$,where $\mathbf{s} = \{constant, y\}$ and y is income, then $V_0(\mathbf{s}, \varepsilon_0; \theta) = \theta_{01} + \theta_{02} y + \varepsilon_0$ and $V_{1j}(\mathbf{s}, C_j, \varepsilon_1; \theta) = \theta_{11} + \theta_{12}(y + A_j) + \varepsilon_1$.

In practice, V_{1j} and V_0 are generally not separably identifiable, but their difference ($\Delta V = V_{ij} - V_0$) is. If V has the simple linear functional form above, then $\Delta V = \theta + \theta_2 A_j + \varepsilon$. This difference can be expressed in a probabilistic framework as

$$\begin{aligned} Pr\{response\ is\ \text{“}yes\text{”}\} &= Pr\{A_j \geq C_j(\cdot)\} \\ &= Pr\{V_1 \geq V0\} \\ &= Pr\{\Delta V \geq 0\}, \end{aligned} \tag{2.1}$$

and hence, the parameters necessary to calculating C_j can be estimated through maximum likelihood. If $\Delta V_\varepsilon = \theta + \theta_2 A_j$, then $\Pr\{A_j \geq C_j(.)\} = \Pr\{\varepsilon_0 - \varepsilon_1 \leq \Delta V_\varepsilon\}$. The probability of farmer adoption at C_j is $F_\varepsilon[\Delta V_\varepsilon(C_j)]$, where F_ε is a cumulative density function (CDF), and $\varepsilon = \varepsilon_0 - \varepsilon_1$. Given that changes in profit associated with adoption, as well as any nonfinancial motivations for adoption, are unlikely to be known to the researcher, survey approaches (such as those that explicitly ask the farmer whether or not she would adopt for a given incentive payment A) are needed to estimate the parameters of F_ε (Cooper, 1997; Cooper and Keim, 1996; Khanna, 2001). Now suppose that three BMPs can be cost-shared, and supposed that the farmers answers "no" to cost share offers for practices 1 and 3, but "yes" to practice 2.

[1] In theory, the farmer's utility maximization process is a combination of the discrete decision to adopt as well as the continuous decision of how many acres to adopt the BMPs on. We address only the former, which was the main focus of our survey questions.

Extending the equation (2.1) above and denoting the joint density by g_C:

$$\begin{aligned}\Pr\{no \text{ to 1 \& 3}, yes \text{ to 2}\} &= \Pr\{C_1 \geq A_1, C_2 \leq A_2 \text{ and } C_3 \geq A_3\} \quad (2.2)\\ &= \int_{A_1}^{\infty}\int_{0}^{A_2}\int_{A_3}^{\infty} g_C\left(C_1, C_2, C_3\right) dc_1 dc_2 dc_3.\end{aligned}$$

As per Hanemann and Kanninen (2001), but applied to the *WTA* case, let $G_C(C_1, .., C_J)$ be the joint distribution function associated with the density $g_C(C_1, .., C_J)$ and let $G_{(j)}(C_1, .., C_J)$ denote the partial derivative of this joint distribution with respect to the j^{th} argument:

$$G_{(j)}(C_1, .., C_J) \equiv \partial G_C(C_1, ..., C_J)/\partial C_j.$$

Then, an equivalent way to express equation (2.2) is (ibid):

$$\begin{aligned}\Pr\{no \text{ to 1 \& 3}, yes \text{ to 2}\} &= \Pr\{C_1 \geq A_1, C_2 \leq A_2 \& C_3 \geq A_3\} (2.3)\\ &= \int_{0}^{A_2} G_{(2)}\left(A_1, C_2, A_3\right) dc_2\end{aligned}$$

Assuming the $\Delta V_{\varepsilon}\left(C_j\right)$ are distributed normally but are correlated through the error terms, then the multivariate distribution needs to account for the correlations, where the $(J \times 1)$ vector ΔV_{ε} is distributed as $\Delta V_{\varepsilon} \sim F\left(\mu^1, \ldots, \mu^J; \Sigma\right)$, where Σ is the $(J \times J)$ correlation matrix between the practices. The next section presents the empirical model for estimating the parameters of such a distribution.

3. Econometric Model

Assume that N farmers choose among a set of J practices. The farmer's RUM associated with the incentive payment offer to adopt the BMP is [2]

$$\Delta V_{\varepsilon ij} = \mathbf{x}'_{ij}\beta_j (j = 1, ..., J; i = 1, ..., N), \quad (3.1)$$

where $\mathbf{x}_{ij}$ is a vector of explanatory variables for choice j for farmer i and β_j the vector of coefficients associated with choice j. The MNP model assumes that the correlations between the practices occur through the error terms in the equations, which are distributed $\varepsilon_i \equiv (\varepsilon_{i1}, ..., \varepsilon_{iJ})' \sim IIDN(0, \Sigma), \Sigma =$

[2] A full MNP model would have variables in the RUMs in equation (2.2) whose values vary across the J choices. While such variables are possible for some datasets, such as those used in recreational site choice, such variables are unlikely to be available to researchers modeling the farmer's technology adoption process. However, convergence of a MNP model with such variables generally requires restrictions on the correlation matrix, such as normalizing it along one row

$\left[\sigma_{ij}\right]$. The MNP log-likelihood function to be estimated is an expanded version of the bivariate model (Greene 1997):

$$L(\beta, \Sigma) = \sum_{i=1}^{N} \log F(\omega_i, \Sigma^*), \tag{3.2}$$

where $\omega_i \equiv (q_{i1} * \Delta V_{\varepsilon i1}, \ldots, q_{iJ} * \Delta V_{\varepsilon iJ})'$ and, for the model for current nonuser users only, $x_{ij} = \{x_{ij1} * r_{ij}, \ldots, x_{ijP} * r_{ij}\}$, where dummy variable $r_{ij} = 1$ if i is a current nonuser of j, and 0 otherwise, $p = 1, \ldots, P$ variables, and

$$q_{ij} = \begin{cases} 1 \text{ if farmer } i \text{ adopts practice } j \\ -1 \text{ if farmer } i \text{ does not adopt practice } j \end{cases} \tag{3.3}$$

and $\sum^* = T_i \sum T_i$, where T_i is a $J \times J$ diagonal matrix with $T_i \equiv (q_{i1}, \ldots, q_{iJ})'$, and where the unrestricted $J \times J$ covariance matrix has $(J-1) \times J$ free elements (after imposing symmetry conditions).

Leaving out the subscript i, the multivariate cumulative normal density function in equation (3.2) is

$$F\left(\vec{w}, \Sigma^*\right) = \frac{1}{\sqrt{|\Sigma^*| (2\pi)^J}} \int_{-\infty}^{w_1} \int_{-\infty}^{w_2} \cdots \int_{-\infty}^{w_J} e^{-\frac{1}{2}\theta' \Sigma^{*-1} \theta} d\theta \tag{3.4}$$

where $w_j = (\omega_j - \mu_j)/\sigma_j$, and where σ_j and μ_j are the standard deviation and mean of ω_j, respectively.

As noted earlier, the computational intractability of the MVN density in equation (3.4) accounts for the fact that it is rarely used in dimensions higher than $J = 2$ (bivariate), or increasingly, $J = 3$ (trivariate). The traditional numerical quadrature methods to calculating $F(\cdot)$ tend not only to be unacceptably slow in more than three or four dimensions, they also suffer from serious shortcoming in numerical accuracy as J increases (e.g., Horowitz, Sparmon, and Daganzo, 1981). An alternative to quadrature methods, namely Monte Carlo methods, is necessary to estimate this CDF. Simulation of standard normal variables is a well-studied problem (see Stern for an overview of simulation-based methods), although applications in the applied economics area exist but are rare (e.g. the trivariate model in Dorfman, 1996). To some extent this state is due to desktop computers only recently having the computational speed to perform this analysis and to a lack of available software. For this study, the GHK (Geweke-Hajivassiliou-Keane, 1997) importance sampling technique (Stern 1997) and a similar technique proposed by Genz (1992) were both tried and gave similar results.

Since the Monte Carlo simulator can approximate the probabilities of the MVN density in equation (3.4) to any desired degree of accuracy, the corre-

sponding maximum simulated likelihood estimate (SMLE) based on the simulated MVN can approximate the MLE estimator (Hajivassiliou, McFadden, and Ruud, 1996). For the results to be consistent, the number of simulations must increase with the sample size at a sufficiently rapid rate (Newey and McFadden 1993). One hundred repetitions are used here (as suggested by Geweke, Keane, and Runkle (1997) for their simulated MNP model).

A potential drawback of the econometric model presented above (or any other multivariate probit applications that the authors are aware of) is that it could potentially be subject to biases associated with incorrect specifications of functional form of the RUM and of the normality assumption. We extend Creel and Loomis' (1997) semi-nonparametric (SNP) distribution-free approach for the univariate discrete choice case to our multivariate discrete choice model. This approach uses the Fourier functional form (e.g., Fenton and Gallant, 1996) as a substitute for the parametric functional form of the RUM in equation (3.1).The Fourier functional form is one of the few functional forms known to have Sobolev flexibility, which means that the difference between a function $\Delta V(\mathbf{x}, \theta)$and the true function $f(x)$ can be made arbitrarily small for any value of $\mathbf{x}$ as the sample size becomes large (Gallant, 1987). Creel and Loomis' specification of ΔV modified for the MNP model is:

$$\begin{aligned} &\Delta V_{\varepsilon F}\left(\mathbf{x}_{ij}, \boldsymbol{\theta}_{kj}\right) = \mathbf{x}'_{ij}\boldsymbol{\beta}_j + \\ &\sum_{m=1}^{M}\sum_{l=1}^{L}\left(v_{lmj}\cos\left[l\mathbf{r}'_{mj}s\left(\mathbf{x}_{ij}\right)\right] - w_{lmj}\sin\left[l\mathbf{r}'_{mji}s\left(\mathbf{x}_{ij}\right)\right]\right) \end{aligned} \tag{3.5}$$

where the $p \times 1$ vector $\mathbf{x}_{ij}$ all arguments of the utility difference model, k is the number of coefficients in $\boldsymbol{\theta}_j$,which consists of the $\boldsymbol{\beta}_j$, v_{lmj}, and w_{lmj} coefficients to be estimated, M and L are positive integers, and $\mathbf{r}_{mj}$ is a $p \times 1$ vector of positive and negative integers that forms indices in the conditioning variables and that determine which combinations of variables in $\mathbf{x}_{ij}$ form each of the transformed variables, and $j = 1, \cdots, J$ BMPs.[3] The integer m is the sum of absolute value of the elements in the multi-indexes in vector $\mathbf{r}_m$ and L is order of the transformation, and is basically the number of inner-loop transformations of $\mathbf{x}_i$ (ignoring the j subscript for clarity of exposition). For example, if $\mathbf{x}_i$ contains 3 variables and M = L= 1, then the $\mathbf{r}_m$ vectors are (1,0,0), (0,1,0), and (0,0,1), resulting in k = 9 (not counting the constant). The $p \times 1$ function $s(\mathbf{x}_i)$ prevents periodicity in the model by rescaling $\mathbf{x}_i$ so that it falls in the range [0, 2p–0.000001] (Gallant, 1987). This rescaling of each element in $\mathbf{x}_i$ is achieved by subtracting from each element in $\mathbf{x}_i$ its minimum

[3]In addition to appending x to the Fourier series in equation (3.5), Gallant suggests appending quadratic terms when modeling nonperiodic functions. Our experiments suggest that inclusion of the quadratic terms in the specifications had little impact on the *WTA* estimates. Hence, we leave them out for the sake of efficiency.

value (from across the sample), then dividing this difference by the maximum value (from across the sample), and then multiplying the resulting value by $[2\pi - 0.000001]$. For example, if bid is the only explanatory variable, then $\mathbf{r}_m$ is a (1×1) unit vector and max (M) equals 1. If furthermore, $M = L$ and bid offer A_j has more than three unique values, then

$$\Delta V_{\varepsilon}\left(A_j, \theta_{kj}\right) = \beta_{1_j} + \beta_{2j} A_j + v_{jv} \cos\ s(A_j) + w_{wj} \sin\ s(A_j) \qquad (3.6)$$

If a variable has only three unique values, then only the v or w transformations may be performed. In practice, the level of transformation in (3.6) generally adds sufficient flexibility to the model. To apply this approach to the multivariate discrete model, the $\Delta V_{\varepsilon ij} = \mathbf{x}'_{ij}\boldsymbol{\beta}_j$ terms in the MLE in (3.2) is replaced with that in equation (3.6).

The SNP functional form for the RUM adds substantial flexibility to the model, and if the assumption of the normal distribution is inappropriate, such an effect should be seen through significant coefficients on the higher ordered terms, noting that the parametric model (3.1) is nested in the SNP model (3.5). Of course, statistical differences between the SNP and the parametric-based MNP approaches may be due to incorrect specification of the functional form or the density function, but these cannot be separably identified.

Table 5.1. Descriptions of the Farm Management Practices Presented in the Survey

Conservation Tillage (CONTILL) - Tillage system in which at least 30% of the soil surface is covered by plant residue after planting to reduce soil erosion by water; or where soil erosion by wind is the primary concern, at least 1,000 pounds per acre of flat small grain residue-equivalent are on the surface during the critical erosion period.
Integrated Pest Management (IPM) - Pest control strategy based on the determination of an economic threshold that indicates when a pest population is approaching the level at which control measures are necessary to prevent a decline in net returns. This can include scouting, biological controls and cultural controls.
Legume Crediting (LEGCR) - Nutrient management practice involving the estimation of the amount of nitrogen available for crops from previous legumes (e.g. alfalfa, clover, cover crops, etc.) and reducing the application rate of commercial fertilizers accordingly.
Manure Testing (MANTST) - Nutrient management practice which accounts for the amount of nutrients available for crops from applying livestock or poultry manure and reducing the application rate of commercial fertilizer accordingly.
Soil Moisture Testing (SMTST) - Irrigation water management practice in which tensiometers or water table monitoring wells are used to estimate the amount of water available from subsurface sources.

Table 5.2. Frequency of Occurrence of Actual Adoption of Various BMP Combinations[a]

CONTILL	IPM	LEGCR	MANTST	SMTST	Freq (%)
1	0	0	0	0	36.66
0	0	0	0	0	15.32
1	0	1	0	0	9.41
1	1	0	0	0	8.92
1	1	1	0	0	7.37
0	1	0	0	0	3.78
1	0	0	0	1	3.30
1	1	1	1	0	2.33
0	0	1	0	0	1.94
1	1	1	0	1	1.94
1	0	1	1	0	1.36
Total Use of each BMP in the sample (percentage)					
74.88	29.78	27.74	7.57	9.70	

Notes: Only bundles with a reported frequency of one percent or greater are listed above. The bundles above represent those reported by 92.33 percent of the farmers in the sample. Sample size = 1,031. Only 0.87% of sample reported using all five practices.
[a] Coding: "1" = BMP is used; "0" = BMP is not used.

4. Numerical Illustration

The data used for the numerical illustration are taken from a data collection and modeling effort undertaken jointly by the Natural Resource Conservation Service (NRCS), the Economic Research Service (ERS), the U.S. Geological Survey (USGS), and the National Agricultural Statistical Service (NASS).[4] Data on cropping and tillage practices and input management were obtained from comprehensive field and farm level surveys of about 1,000 farmers apiece for cropping practices in each of four critical watershed regions. None of the respondents indicated that they were enrolled in WQIP (the EQIP-like program in existence at the time of the survey). Table 5.1 describes the five BMPs that were addressed in the survey instrument. Table 5.2 presents the frequency of occurrence of actual adoption of various BMP combinations in the sample. As one would expect, the choice of bundles is clearly not distributed equaled across the possible sets. Out of 32 possible combinations (including the null set), over 92% of the farmers are accounted for by 11 combinations. However, Table 5.2 tells us nothing about what the socially optimal bundles are; the farmer's choice of bundle is largely a business decision, while the socially optimum choice balances economic and environmental costs and benefits. Here

[4]As the data is discussed in detail in Cooper and in Cooper and Keim (1996), for brevity and to avoid repetition, we do not discuss the data in detail here.

we focus on the hypothetical adoption decision by current nonusers of the practice, with the appendix presenting the analysis of the actual (current) adoption decision. In the survey, current nonusers of each practice (i.e., those who said that they did not currently use the practice) were asked if they would adopt the BMP with an incentive payment of $\$[X]$ per acre, a value which was varied across the respondents in the range 2 to 24.

For any one respondent however, to avoid anchoring biases across the responses, the offered bid was the same for each practice. Each of the five adoption questions was placed on the same page so that the respondent was concurrently aware of all five. As the bid variable (cost share) is uncorrelated with any other variables by the design of the survey instrument, it is the only relevant variable for examining the relationship between the bid and probability of acceptance and for calculating the mean benefit for the sample (McFadden 1994), with additional explanatory variables serving largely to stratify the sample. For brevity then,we present the results for the specifications with just the bid variable.[5] The appendix presents econometric results that relate a number of explanatory variables to the actual decision to adopt, i.e., the analysis of adoption where $A_j = \$0$.

As only the adoption decision of current nonusers of the BMPs is analyzed in the main body of this paper, the estimated probabilities are conditional probabilities, which are conditional on nonadoption, i.e., PrFarmer accepts bid A in turn for adoption of the BMPfarmer is not current user of the BMP. This conditional probability is appropriate to the study of this policy goal of examining the USDA's Environmental Quality Incentives Program (EQIP). This is because EQIP offers cost shares only to current nonusers of the BMPs. Hence, the policy relevant density function for EQIP is PrFarmer accepts bid A in turn for adoption of the BMP—farmer is not current user of the BMP, and not the unconditional PrFarmer accepts bid A in turn for adoption of the BMP. In other words, concern over potential sample selection bias in examining only current nonusers is eliminated if our policy interest is EQIP-like cost share programs. In fact, for the purposes of these programs, we are only interested in the subsample of farmers who do not currently use the BMPs.

The likelihood function and maximization routines were programmed by the author in GAUSS.[6] The estimation results are presented in Table 5.3 (correlation coefficients are presented in Table 5.6 in the appendix). The second and third columns in Table 5.3 are the results using the parametric RUM spec-

[5]Results for the SNP model with multiple regressors is too lengthy to present here, but are available upon request from the authors.

[6]The only commercially available program that the authors are aware of that performs the MNP using the simulated normal is an optional package in Limdep. However, the author found that modeling the data on the five BMPs with the commercially available software was too computationally burdensome to be practical.

Table 5.3. Restricted and Unrestricted Multinomial Probit Estimates

		Parametric		SNP	
		Restricted	*Unrestricted*	*Restricted*	*Unrestricted*
Practice	*Variable*	*Coefficient Estimates (Asymptotic z-values)*			
CONTILL	CONST	–0.7815	–0.8199	–0.7783	–0.7520
		–(4.252)	–(5.750)	–(2.445)	–(2.962)
	BID	0.0221	0.0187	0.0217	0.0150
		(1.834)	(1.838)	(1.024)	(.834)
	sin s(BID)			0.0217	0.0738
				(.287)	(1.427)
	cos s(BID)			0.0106	0.0179
				(.105)	(.239)
IPM	CONST	–1.0979	–1.0729	–1.1157	–1.0971
		–(10.61)	–(12.11)	–(7.488)	–(8.110)
	BID	0.0325	0.0256	0.0344	0.0273
		(3.970)	(3.770)	(2.839)	(2.542)
	sin s(BID)			–0.0209	0.0004
				–(.537)	(.011)
	cos s(BID)			0.0037	0.0087
				(.069)	(.212)
LEGCR	CONST	–1.7462	–1.4099	–1.5381	–1.3265
		–(10.93)	–(15.46)	–(7.452)	–(10.86)
	BID	0.0469	0.0301	0.0283	0.0234
		(4.118)	(4.682)	(1.678)	(2.533)
	sin s(BID)			0.0460	–0.0069
				(.744)	–(.215)
	cos s(BID)			–0.1208	–0.0646
				–(1.683)	–(1.156)

ification and the last two columns represent the SNP RUM model results. The "restricted" columns present the results for the model where the off-diagonal terms in the correlation matrix of the five practices are restricted to equal zero. In this case, the estimated coefficients and standard errors are equivalent to those from separate probit specifications for each practice. The coefficient on the offer amount (BID) is of the expected sign and significant to at least the 10% level, and for most cases, the 1% level, except for those for CONTILL in the SNP models, perhaps due to some collinearity in that case of BID with the higher order terms. In fact, the bid offer for CONTILL was $2 per acre lower, and hence, the bid range narrower, than for the other practices (which all had the same bid offers), as pre-testing of the survey suggested that farmers expected conservation tillage to receive a lower cost share than the other practices. Note that the only two cases where one of the higher order terms is significant is that on "cos s(BID)" and "sin s(BID)" in the restricted case for adoption of LEGCR and SMTST, respectively.

Table 5.3 (continued)
Restricted and Unrestricted Multinomial Probit Estimates

		Parametric		SNP	
		Restricted	*Unrestricted*	*Restricted*	*Unrestricted*
Practice	*Variable*	*Coefficient Estimates (Asymptotic z-values)*			
MANTST	CONST	–1.5757	–1.3729	–1.6911	–1.4134
		–(12.15)	–(13.97)	–(8.562)	–(10.77)
	BID	0.0334	0.0226	0.0442	0.0264
		(3.445)	(3.033)	(2.823)	(2.530)
	sin s(BID)			–0.0817	–0.0482
				–(1.634)	–(1.382)
	cos s(BID)			0.0233	0.0003
				(.361)	(.007)
SMTST	CONST	–1.4575	–1.3253	–1.4840	–1.3403
		–(12.32)	–(15.79)	–(9.445)	–(10.1)
	BID	0.0327	0.0239	0.0311	0.0229
		(3.661)	(3.938)	(2.496)	(2.210)
	sin s(BID)			0.0802	0.0364
				(1.726)	(.912)
	Cos s(BID)			0.0339	0.0115
				(.547)	(.238)
	ln L	–2,511.60	–2,099.42	–2,505.85	–2,106.99

Notes: The *unrestricted* multinomial probit model estimates the correlation between the five practices. The *restricted* model assumes the cross practice correlations are zero, and hence, its coefficients and standard errors are the same as in individual standard results for each practice. For each practice, probit the dependent variable = "1" if the farmer agrees to adopt the practice at the offered bid (cost share), and "0" otherwise.

As the restricted model (i.e., each adoption function is independent of the other) is nested within the unrestricted model for both the parametric and SNP cases, the likelihood ratio statistic, namely $LR = -2(\ln L_r - \ln L_u)$, can be used to test the null hypothesis that farmers consider each BMP adoption decision independently from the other decisions. Given the log-likelihood values at the bottom of Table 5.3, this hypothesis is not accepted for any reasonable level of significance in either the parametric or SNP cases. As the unrestricted RUM is nested within the unrestricted SNP RUM, a likelihood ratio test can be used to test the null hypothesis that the BMPs are distributed normally with a linear RUM. This hypothesis is also rejected, but the critical value of 15.14 is much lower than those comparing the restricted and unrestricted models. In fact, a simple visual comparison of the coefficients on BID between the restricted (unrestricted) parametric and restricted (unrestricted) SNP models indicates no major differences. The appendix provides the results for the MNP analysis of current users versus nonusers of the BMPs.

Table 5.4. Definitions of the Explanatory Variables

TACRE - Total acres operated (1123/2034).
EDUC - Formal education of operator (3.194/2.314).
EINDEX - Sheet and rill erosion index.
FLVALUE - Estimated market value per acre of land (1383/1023).
EXPER - Farm operator's years of experience (24.83/20.15).
BPWORK - Number of days annually operator worked off the farm (42.71/99.15).
NETINC - Operation's Net farm income in 1991 (24620/26890).
TISTST - Tissue test performed in 1992 (dummy)(0.029/0.149).
CTILL - Conservation tillage used in 1992 (dummy)(0.174/0.457).
PESTM - Destroy crop residues for host free zones (dummy)(0.163/0.355).
ANIMAL - Farm type-beef,hogs,sheep (dummy)(0.207/0.522).
ROTATE - Grasses and legumes in rotation (dummy)(0.049/0.239).
MANURE - Manure applied to field (dummy)(0.147/0.430).
HEL - Highly erodible land (dummy)(0.174/0.457).
IA - Sample located in the Eastern Iowa or Illinois Basin Area (dummy)(0.729/0.721).
ALBR - Sample located in Albermarle-Pamlico Drainage Area (dummy)(0.088/0.209).
IDAHO - Sample located in the Upper Snake River Basin Area (dummy)(0.123/0.341).

Note: values in brackets are mean/standard error.

Two basic conclusions follow from this analysis. One is that none of the available variables stand out as an important predictor of current use (the SNP specification was not practical with this larger data set).[7] But the most relevant result for this study is that the restricted MNP model is rejected at any reasonable level of significance, given that the log-likelihood for the unrestricted MNP model of –2,176, and the log-likelihood value for the (nested) restricted model (not shown) is –2,223. This result can be interpreted as empirical evidence that the correct decision was made for the hypothetical survey questions to make the respondent concurrently aware of each of the possible BMPs.

For the analysis of hypothetical adoption, the correlation coefficients between the practices are significant to at least the 1% level as well, regardless of whether they are estimated for the parametric or SNP models (Appendix 5.6, second and third set of numbers). The correlation coefficients for the models predicting current use (first set of numbers in Table 5.6) tend to be less significant than the correlations between the hypothetical use results.

This difference in significance is to be expected; whether or not the farmer is a current user of the BMPs is a result of an evolutionary process, while the

[7] 7 The decision on which variables to include in the specifications for each of the practices was based on whether or not the variables appear justified from a farm management standpoint.

Table 5.5. Multinomial Probit Regression Results Predicting Actual BMP Use/Nonuse

	CONTILL	IPM	LEGCR	MANTST	SMTST
Variable	Coefficient Estimates (Asymptotic z-values)				
CONST	–0.0005	–0.7960	–0.9896	–1.5657	–0.8799
	–(.002)	–(3.123)	–(2.743)	–(2.934)	–(1.995)
EDUC	–0.0072	0.1713	0.0928	0.0444	0.0158
	–(.198)	(4.973)	(2.838)	(.814)	(.226)
CTILL	0.3638	–	–	–	–
	(3.583)				
TISTST	–	–	-0.1174	-1.9290	–
			-(.402)	-(2.107)	
HEL	–0.0665	–	–	–	–
	–(.536)				
EXPER	0.0018	–0.0027	–0.0015	–0.0064	–0.0053
	(.489)	–(.706)	–(.424)	–(1.034)	–(.736)
PESTM	–0.0046	0.3862	–	–	–
	–(.032)	(2.998)			
ROTATE	0.0442	–0.0041	0.2687	–	–
	(.196)	-(.018)	(1.500)		
MANURE	–0.1153	–0.1821	0.0957	0.3512	–
	–(.954)	–(1.336)	(.828)	(2.167)	
ANIMAL	–0.0074	–0.2869	–0.0071	0.1424	–0.2030
	–(.062)	–(2.246)	–(.068)	(.868)	–(.860)
TACRE	7.38E-06	5.66E-05	–3.07E-06	–1.85E-05	3.66E-05
	(.235)	(1.513)	–(.088)	–(.466)	(.886)
FLVALUE	–2.46E-05	5.24E-05	–0.0001	–7.22E-05	–1.86E-05
	–(.310)	(.739)	–(1.636)	–(.579)	–(.137)
IA	0.4343	–0.0815	0.8586	0.7698	–0.4328
	(2.017)	–(.398)	(2.862)	(1.792)	–(1.498)
ALBR	0.4087	–0.1323	–0.3991	–0.1349	–1.6877
	(1.478)	–(.479)	–(1.113)	–(.261)	–(4.01)
IDAHO	0.2278	–0.3957	0.5796	0.4910	0.2289
	(.917)	–(1.667)	(1.805)	(1.021)	(.795)
BPWORK	–0.0002	–0.0003	–0.0008	0.0002	–0.0004
	–(.400)	–(.534)	–(1.541)	(.218)	–(.382)
NETINC	1.02E-06	6.21E-07	–3.23E-06	–2.19E-06	7.53E-06
	(.394)	(.286)	–(1.524)	–(.655)	(1.772)

Note: For each BMP, the dependent variable = "1" if the farmer currently uses the BMP and "0" otherwise. Log-Likelihood = –2,175.86.

hypothetical adoption decisions are made over a bundle of practices offered to the farmer at one point in time in a survey instrument.[8]

[8] An EQIP contract application is submitted at a particular point in time, when the proposed practices have not yet been adopted. The farmer is paid per year for each of the practices he agrees to adopt over the life of the contract.

Table 5.6. Estimates of Correlations Between the BMPs.

Practice	CONTILL	IPM	LEGCR	MANTST	SMTST
	Regression Predicting Actual BMP Use/Nonuse				
CONTILL	–				
IPM	0.1230 (1.6)	–			
LEGCR	0.2020 (2.7)	0.4010 (7.2)	–		
MANTST	0.0451 (0.4)	0.4170 (5.2)	0.5310 (6.7)	–	
SMTST	0.1860 (1.8)	0.2040 (2.1)	0.1240 (1.3)	0.3050 (2.7)	–
	Regression for the Hypothetical Adoptors Only - Parametric				
CONTILL	–				
IPM	0.7379 (16.1)	–			
LEGCR	0.7584 (17.7)	0.8151 (22.4)	–		
MANTST	0.5295 (7.2)	0.7341 (14.7)	0.8936 (28.1)	–	
SMTST	0.6052 (9.7)	0.6700 (12.3)	0.7857 (15.1)	0.7649(16.9)	–
	Regression for the Hypothetical Adoptors Only - SNP				
CONTILL	–				
IPM	0.7400 (15.7)	–			
LEGCR	0.7776 (18.2)	0.8188 (22.5)	–		
MANTST	0.5545 (7.6)	0.7508 (15.1)	0.8957 (28.3)	–	
SMTST	0.6120 (9.7)	0.6722 (12.1)	0.7906 (14.8)	0.7749(17.4)	–

Next, given that the restricted correlation model is not accepted (i.e., the BMP adoption decisions are not independent across BMPs), we turn to an evaluation of how the unrestricted MVN results can be used for analysis of bundling. The basic value of the multivariate analysis is it allows us to calculate the joint probabilities as a function of the incentive payments.

Figures 5.1 through 5.4 provide examples of how the joint probability changes as a function of the incentive payment offers for four of the five BMPs analyzed here, i.e., the curves represent $\partial G_C\left(C_1, ..., C_J\right)/\partial C_j$ calculated across a wide range of cost share offers. For example, Figure 5.1 plots the probability of non-acceptance of the conservation tillage cost share as a function of the cost share offer amount, given the value of the cost share offers for the other four BMPs. In Figure 5.1, four scenarios (numbers 2 through 5) with different fixed offers for BMPs other than CONTILL are presented. For comparison, scenario 1 is the predicted probability for the standard univariate normal density function that does not explicitly account for the other bid offers.

Given the estimates of the CDFs generated from the analysis of discrete responses in the figures, the question may arise of how to summarize these distributions of *WTA* for practical purposes. In the discrete choice contingent valuation (CV) literature, the most common summary statistic is the mean of

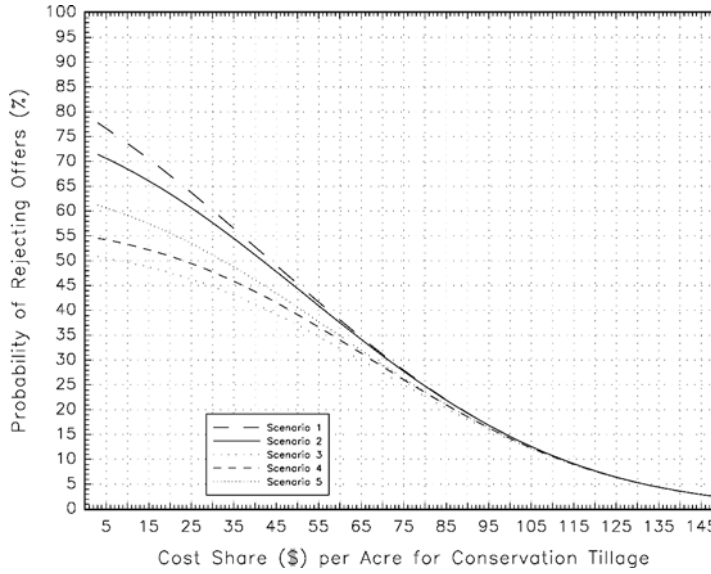


Figure 5.1. Conservation tillage. Probability of rejecting offers at different cost share per acre.

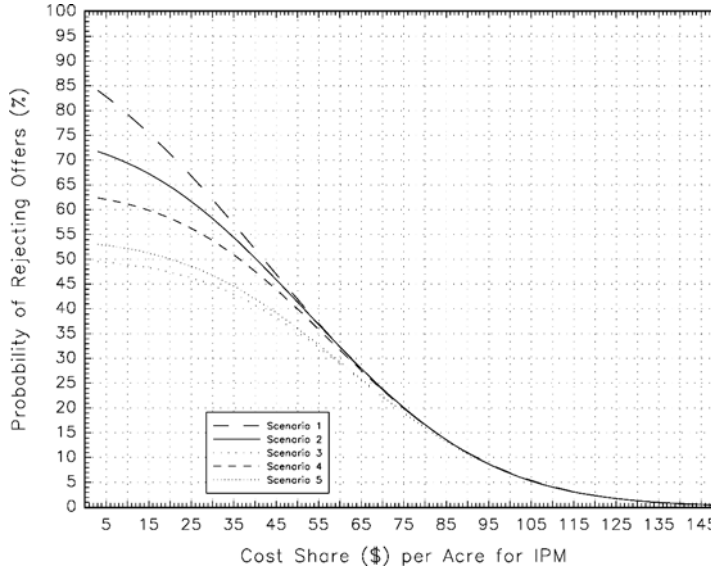


Figure 5.2. IPM. Probability of rejecting offers at different cost share per acre.

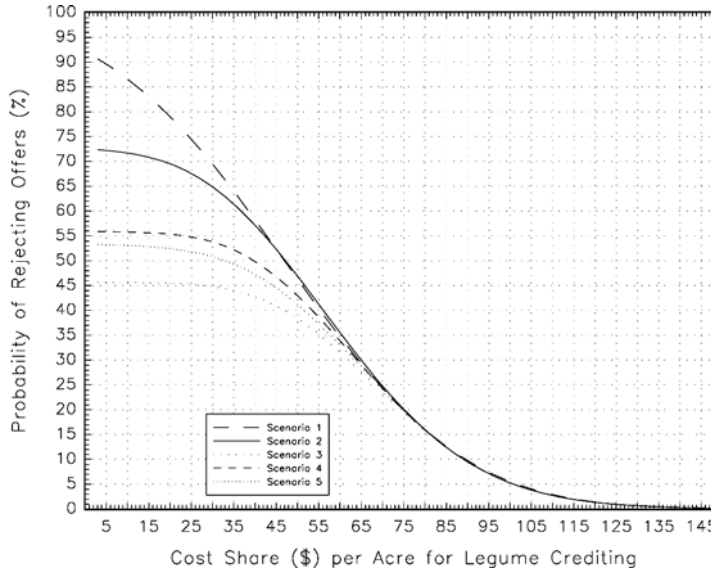


Figure 5.3. Legume crediting. Probability of rejecting offers at different cost share per acre.

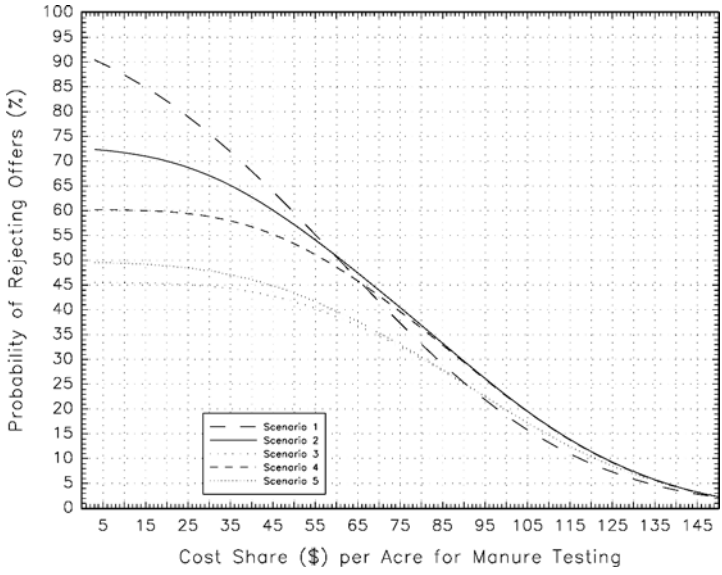


Figure 5.4. Manure testing. Probability of rejecting offers at different cost share per acre.

the estimated *WTP* or *WTA* distribution. Given the estimated coefficients from the multivariate probit model, it is possible to calculate measures of conditional mean *WTA*. Hanemann (1984) notes that in the case where the benefit measure C_R is restricted to the non-negative range, its mean value can be found using the following formula for the mean of a random variable:

$$C_R = \int_0^\infty F(C)dc, \tag{4.1}$$

where $F(C) = F[-\Delta V_\epsilon(\mathbf{s}, C, \theta)]$ is the cumulative density function for WTP.[9][10] Here we present a new application, in which the mean of the benefit measure for one good is calculated conditional on the bid offers made for other goods, an application made possible by the estimation of the multivariate normal CDF. Given a five dimensional version of the multivariate CDF in equations 2.3 and 3.4, one can take the integral under F with respect to one of the BMPs, thereby yielding a mean *WTA* for that BMP, conditional on the cost share offer amounts for the other BMPs. For example, for BMP $j = 2$ the mean *WTA*, $C_{R(2)}$, is

$$C_{R(2)} = \int_0^\infty F_{(2)}\left(A_1, C_2, A_3, A_4, A_5\right) dc_2 \tag{4.2}$$

In other words, Equation (4.2) corresponds to the area under the curves in Figures (5.1) through 5.4. Of course, given that a numerical approach is used to calculate the multivariate CDF, the mean value must also be calculated numerically. The lower right hand corner of each figure presents the mean *WTA* values for each scenario, with the value associated with scenario 1 calculated using Equation 4.1 and the values associated with scenarios 2 through 5 calculated using appropriate variations of Equation (4.2).

In Figures 5.1 through 5.4, the probability functions and mean *WTA* values differ little for scenarios 1 and 2, with the latter being consistently lower, although little perhaps can be said of this comparison as the results are generated from two different density function assumptions. Bigger differences tend to occur between scenario 2 and the other joint density scenarios (scenarios 3 through 5 in each figure). As is expected, the lowest minimum *WTA* in each figure is associated with scenario 3, the one in which all the BMPs except the one in question (e.g. CONTILL in Figure (5.1)) are offered at a cost share

[9]In this study, we are interested in comparisons of E(WTA) between the scenarios and not what the ideal benefit measure should be. For a discussion of the pros and cons of various benefit measures and issues of consistency of the estimated distribution with the benefit measure, see Hanemann and Kanninen (2001).
[10]WTP, the integral is taken over $1 - F(C)$, where $F(C) = F[-\Delta V_\epsilon(\mathbf{s}, C, \theta)]$

of $30 per acre. In other words, given that adoption of the practices is positively correlated (Table (5.6)), one would expect that more a farmer is offered to adopt a set of related practices, the less she will be willing to accept for the practice in question. Hence, for comparisons of scenarios 2 through 5, *WTA* should be highest under 2 and lowest under 3. This relationship holds for each of the four figures. However, what is interesting from the standpoint of bundling the BMPs is that only one or two of the BMPs in scenario 3 in each figure may be driving much of the reduction in *WTA* over scenarios 2, 4, and 5. First take Figure 5.1. In this figure, *WTA* under scenario 4 is not much higher than under scenario 3 even though only two of the other costs are being offered at nonzero cost shares. In fact, as shown in Table 5.2, the bundle CONTILL, IPM, LEGCR in scenario 4 is used by 7.37% of the actual users, while only 0.19% use the bundle CONTILL, MANTST, SMTST in scenario 5.[11] In Figure 5.2, the *WTA* for IPM in conjunction with CONTILL and SMTST cost shares at $30 (scenario 5) was almost as low as that with the four BMPs being offered at $30 in scenario 3 (scenario 3). For Figure 5.3, no pairs of BMPs offered at $30 in conjunction with LEGCR seemed to offer the bulk of the decrease in *WTA* associated with that in moving from scenario 2 to 3. In Figure 5.4 however, offering CONTILL and SMTST at $30 (scenario 5) yielded a *WTA* for MANTST almost as low as that for scenario 3.

5. Conclusion

This chapter develops an econometric model based on the multivariate normal distribution that identifies producer tendencies to bundle types of management practices that may be covered under an incentive payment system. Identifying producer tendencies to bundle these types of practices may increase adoption and lower the costs of voluntary adoption programs. Although the scenario examined here relies on payments to encourage adoption, identifying these producer tendencies can also lower the government's costs of voluntary adoption programs that rely on the dissemination of information to encourage adoption. Since a critical component of voluntary adoption is producer perceptions, as in the numerical illustration, identifying and packaging BMPs that are perceived to be jointly beneficial, or bundled, may increase adoption and lower the costs of the programs. Thus, jointly modeling the observed adoption data across the BMPs can indicate which practices should be bundled into composite practices.

[11] Further breakdowns of scenario 4 could be used to test whether IPM or LEGCR are contributing most to reducing *WTA* from the level under scenario 2, but are not considered here for the sake of brevity, given that the target audience for the detailed information on the bundles are managers of the cost sharing program, and perhaps not the general readership of this journal.

Our model can be used to identify programs costs in the currently less than ideal situation facing EQIP (and all similar programs), where the environmental benefits associated with BMP adoption are unquantified. Research is perhaps moving in the direction of quantifying (if not monetizing) the environmental benefits of such practices, e.g. the USGS' Sparrow model may be modified in the future to measure impacts on sediment and nutrient loadings in watersheds that are associated with such practices. Given our estimated model in conjunction with an environmental benefits model, benefit-cost tradeoffs of BMP adoption can be assessed.

Chapter 6

A COMPARISON BETWEEN MULTINOMIAL LOGIT AND PROBIT MODELS

Analyzing the Determinants of Environmental Innovations

Andreas Ziegler
Centre for European Economic Research (ZEW)
Department of Environmental and Resource Economics, Environmental Management Mannheim
Germany
ziegler@zew.de

Abstract Although the estimation of flexible multinomial discrete choice models generally needs the incorporation of simulation methods, their application is recently common in environmental and resource economics such as in many other economic disciplines (e.g. transportation economics). Based on a firm level data set of the German manufacturing sector, this paper examines determinants of environmental innovations by comparing the estimation results in flexible multinomial probit models and restrictive multinomial logit and independent probit models. The analysis of the two latter models implies that some specific environmental organizational measures, technological opportunities, and market pull factors have a significantly positive effect on both environmental product and process innovations. Taking this into consideration, the flexible multinomial probit model analysis provides few new insights since the simulated maximum likelihood estimations are rather unreliable as a consequence of the sole inclusion of firm-specific characteristics as explanatory variables. In this respect, the incorporation of simulation methods into the maximum likelihood estimations is not crucial since the problems do not decrease if the number of random draws in the considered Geweke-Hajivassiliou-Keane simulator rises. Furthermore, the difficulties grow if the number of choice alternatives increases. It can therefore be concluded that the applicability of these flexible multinomial discrete choice models without the incorporation of choice-specific attributes as explanatory variables is rather limited in practice.

Keywords: Environmental innovations, double dividend, multinomial probit.

1. Introduction

Although the estimation of flexible multinomial discrete choice models generally needs the incorporation of simulation methods due to the underlying multiple integrals, their application has only recently become common in environmental and resource economics (see e.g. Bjørner *et al.*, 2004, Rennings *et al.*, 2004) as well as in many other economic disciplines (e.g. transportation economics, see Bolduc, 1999, Brownstone and Train, 1999, Greene and Hensher, 2003). Based on a firm level data set of the German manufacturing sector, this paper examines determinants of different types of environmental innovations as a measure of environmental performance by comparing the estimation results in multinomial probit models (MPM) and multinomial logit models (MLM). The background is that such environmental innovations receive increasing attention from policy makers and scholars since they are expected to produce a double dividend, i.e. limit environmental burden and contribute to the technological modernization of the economy (see Rennings, 2000, Jaffe *et al.*, 2002).

It is well-known that the popular MLM (see McFadden, 1974) have restrictive properties due to the underlying assumptions regarding the stochastic model components. Whereas the independent MPM are still rather restrictive (see Hausman and Wise, 1978) due to the independence assumption in these components, the flexible variants of the MPM (see e.g. Börsch-Supan and Hajivassiliou, 1993) are a general framework since they allow correlations between the choice alternatives of the dependent variable. Indeed, the application of flexible MPM requires the inclusion of simulators into an estimation method due to the underlying multiple integrals in the choice probabilities. In this paper, the maximum simulated likelihood method (MSL), i.e. the simulated counterpart of the maximum likelihood method (ML), incorporating the so-called Geweke-Hajivassiliou-Keane (GHK) simulator (see Börsch-Supan and Hajivassiliou, 1993, Geweke *et al.*, 1994, Keane, 1994) is considered, since its use seems to be advantageous compared with the use of other combinations of classical estimation methods and simulators.[1]

This paper is organized as follows: In the second section, the methodological approach as well as the dependent and explanatory variables for the empirical analysis are explained. The third section provides some details concerning the firm level data set. In the fourth section, the estimation results in the differ-

[1] A general practical advantage is that the MSL estimation of MPM has been implemented directly in some software packages (e.g. GAUSSX and LIMDEP) in contrast to other simulated classical estimations of MPM and also (to my knowledge) in contrast to the MSL estimation of mixed logit models as alternative flexible multinomial discrete choice models (see e.g. Revelt and Train, 1998, Brownstone and Train, 1999, McFadden and Train, 2000, Greene and Hensher, 2003, Bjørner *et al.*, 2004).

ent multinomial discrete choice models are compared. The final section draws some conclusions.

2. Methodological Approach and Variables

2.1 Background on Determinants of Environmental Performance

Environmental product and process innovations as specific kinds of conventional product and process innovations (as defined in the Oslo-Manual of the OECD, 1997) consist of new or modified products and processes to avoid or reduce environmental harms (see e.g. Rennings and Zwick, 2002). It is often argued that self-regulated environmental organizational measures such as the certification for environmental management systems (EMS) are able to promote such innovations (see e.g. Rennings *et al.*, 2003). In this respect, initiatives regarding the encouragement of firms to participate in voluntary pollution prevention programs play a central role in the discussion of so-called soft environmental policy instruments (see e.g. Arora and Cason, 1995, Khanna and Damon, 1999). By analyzing the effect of environmental organizational measures, this paper differs from other econometric studies that examine determinants of such measures (see e.g. Henriques and Sadorsky, 1996, DeCanio and Watkins, 1998, Nakamura *et al.*, 2001, Khanna and Anton, 2002). Concerning the analysis of such measures as determinants of environmental performance, it rather has some analogies to the studies of Khanna and Damon (1999), Dasgupta *et al.* (2000), or Anton *et al.* (2004). But in contrast to these studies, this paper considers environmental product and process innovations as indicators for environmental performance instead of narrower proxies such as toxic chemicals according to the Toxic Release Inventory (TRI) or self-assessed environmental regulatory compliance.

Jaffe and Palmer (1997) examine the effect of environmental regulatory compliance expenditures on innovation activities. Indeed, they consider innovation activities in general, i.e. not specifically environmental activities. To my knowledge, the only econometric analyses of determinants of environmental innovations can be found in Cleff and Rennings (1999), Rennings *et al.* (2003), and Brunnermeier and Cohen (2003). However, the former two studies examine restrictive samples of companies so that the conclusions regarding determinants of environmental innovations are limited. The latter study does not consider the effects of environmental organizational measures. Furthermore, it measures environmental innovations by the number of successful environmental patent applications. Instead, this paper considers the end of the innovation process and thus applies the innovation definition according to the Oslo-Manual of the OECD (1997). But most notably, Brunnermeier and Cohen (2003) only utilize aggregated data at the industry level. This is the reason

why they are not able to examine firm-specific determinants of environmental innovations so that the conclusions based on their analysis are again limited. In contrast, a firm level data set of the German manufacturing sector is analyzed in this paper.

2.2 Multinomial Discrete Choice Models

This paper particularly considers multinomial discrete choice models. Note that not only determinants of environmental innovations in general are examined since the underlying data include information on environmental product and process innovations. This is the reason why binary discrete choice models are not analyzed. In the framework of binary logit or probit models for the explanation of a specific type of environmental innovation (see the results in Ziegler and Rennings, 2004), the basic choice alternative may also comprise another type of environmental innovation. Thus, these simple discrete choice models cannot examine determinants of specific types of environmental innovations compared with the absence of environmental innovations as basic alternative. However, such an analysis is possible with multinomial discrete choice models by constructing suitable mutually exclusive alternatives. This paper compares the estimation results in three-alternative and four-alternative discrete choice models. In this framework, it is assumed that a firm i ($i = 1, \ldots, N$) realizes one specific type j ($j = 1, \ldots, J$ with $J = 3$ in the three-alternative and $J = 4$ in the four-alternative case) of mutually exclusive types of environmental innovations (including the possibility of the realization of no environmental innovation). The underlying (unobservable) latent variables have the following appearance:

$$U_{ij} = \beta_j{}' x_i + \varepsilon_{ij}$$

In this framework, one can imagine U_{ij} of firm i as an attraction measure for the profit with reference to the realization of environmental innovation type j. The vectors $x_i = (x_{i1}, \ldots, x_{i,24})'$ comprise 23 explanatory variables as discussed below and one constant. The corresponding unknown coefficient vectors are $\beta_j = (\beta_{j1}, \ldots, \beta_{j,24})'$. The coefficient vectors β_3 in the three-alternative and β_4 in the four-alternative discrete choice models are restricted to zero to ensure the formal identification of the models. The values of the latent variables U_{ij} cannot be observed and depend on the stochastic components ε_{ij} that summarize all unobserved factors that influence the environmental innovation decision. Observable are the realizations of the following Bernoulli variables:

$$D_{ij} = \begin{cases} 1 & \text{if firm } i \text{ realizes environmental innovation type } j \\ 0 & \text{otherwise} \end{cases}$$

It is assumed that i chooses j if U_{ij} is greater than all other $U_{ij'}$ ($j \neq j'; j, j' = 1, \ldots, J$). In other words, firm i chooses environmental innovation type j if its realization leads to a higher profit than the realization of all other environmental innovation types. The probability for the choice of j by i is therefore:

$$P_{ij} = P(U_{ij} > U_{ij'})$$

The choice probabilities P_{ij} particularly depend on the unknown parameters, summarized in the vector θ. This parameter vector comprises the coefficients of the explanatory variables in β_j and the variance covariance parameters in the flexible MPM as discussed below. By incorporating these choice probabilities $P_{ij}(\theta)$ into the ML estimator and by considering N independent observations, one obtains the specific ML estimator:

$$\hat{\theta} = \arg\max_{\theta} \left[\sum_{i=1}^{N} \sum_{j=1}^{J} D_{ij} \ln P_{ij}(\theta) \right]$$

If it is assumed that the stochastic components ε_{ij} ($\forall i, j$) are independently and identically distributed with Type I extreme value density functions, one obtains the popular MLM(see McFadden, 1974). Since the choice probabilities $P_{ij}(\theta)$ have a simple structure, these multinomial discrete choice models can be easily estimated by the ML, implemented in standard software packages. If it is assumed that the ε_{ij} ($\forall i, j$) are independently and identically distributed standard normal random variables, one comes to the independent MPM. Indeed, both multinomial discrete choice models are based on the restrictive independence assumption in the stochastic components ε_{ij}. On the one hand, this leads to the so-called independence of irrelevant alternatives (IIA) property in the MLM. This property implies that the choice between two alternatives is independent of the existence of other alternatives and thus may be limiting in many cases. On the other hand, the independent MPM are likewise restrictive and have properties similar to the IIA (see Hausman and Wise, 1978).

In contrast, the flexible variants of the MPM (see e.g. Börsch-Supan and Hajivassiliou, 1993) are a general framework since they allow correlations between the choice alternatives of the dependent variable (i.e. the environmental innovation types). One obtains these models if it is assumed that the ε_{ij} are jointly normally distributed:

$$\varepsilon_i = (\varepsilon_{i1}, \ldots \varepsilon_{iJ}) \sim \mathrm{N}_J(0; \Sigma_J)$$

The variance covariance matrices $\Sigma_J = (\sigma_{jj'})$ contain six or ten different variance and covariance parameters in three-alternative or four-alternative probit models. However, not all these parameters are formally identifiable (see e.g. Dansie, 1985, Bunch, 1991). Only two or five variance covariance parameters

can at most be formally identified in these MPM. This paper restricts the variances σ_{JJ} and $\sigma_{J-1,J-1}$ to the value one and the covariances σ_{jJ} $(\forall j \neq J)$ to the value zero. Consequently, only one variance and one covariance in the three-alternative as well as two variances and three covariances in the four-alternative case are freely estimable. Due to practical aspects, the corresponding standard deviations σ_j $(j = 1, \ldots, J-2)$ and correlation coefficients $\text{corr}(\varepsilon_{ij}, \varepsilon_{ij'}) = \sigma_{jj'}/\sqrt{\sigma_{jj}\sigma_{j'j'}}$ $(j, j' = 1, \ldots, J-1; j \neq j')$ are examined in this paper (for details see Ziegler and Eymann, 2001).

The practical disadvantage of the application of flexible MPM is the inconvenient form of the choice probabilities $P_{ij}(\theta)$ that are characterized by $(J-1)$-dimensional integrals. Therefore, the ML estimations of flexible MPM with many choice alternatives are computationally infeasible with deterministic numerical integration methods and thus standard software packages cannot be utilized. But the $P_{ij}(\theta)$ can be quickly and accurately approximated with (unbiased) stochastic simulation methods, i.e. with R repeatedly transformed draws of (pseudo-) random numbers (see e.g. the overviews in Hajivassiliou *et al.*, 1996, or Vijverberg, 1997). In this paper, the GHK simulator (see Börsch-Supan and Hajivassiliou, 1993, Keane, 1994, Geweke *et al.*, 1994) is examined since it outperforms other simulation methods with regard to the approximation of the true probability. For the GHK simulator, repeated sequential (pseudo-) random draws from the truncated standard normal distribution have to be made. By incorporating the simulated choice probabilities into the ML estimator, the MSL estimator (see e.g. Gouriéroux and Monfort, 1993) is obtained.

Note that the ML estimations in the present case of three-alternative and four-alternative probit models would be possible with deterministic numerical integration methods. Indeed, the conventional ML estimations are not generally better than the MSL estimations if multiple integrals arise. With regard to the necessary incorporation of simulation methods into the ML estimations of MPM with more than four choice alternatives, the influence of the GHK simulator on the reliability of the estimation results is also examined in this paper. Note furthermore that the independent MPM are in principle computationally not problematic due to the independence assumption in the stochastic components ε_{ij} as aforementioned.[2] The choice probabilities $P_{ij}(\theta)$ are only characterized by one-dimensional integrals even if the number J of choice alternatives is large. With regard to the comparison to the MSL estimations of the flexible variants of the MPM, the GHK simulator is also incorporated into the ML estimations of the independent MPM.

[2]One obtains these restrictive variants of the MPM if it is assumed that the variance covariance matrix Σ_J of ε_i as discussed above is the identity matrix.

2.3 Dependent and Explanatory Variables

In this paper, environmental product innovations mean the planned introduction of an environmentally improved or a new environmentally friendly product (e.g. solvent-free paints or energy efficient products such as cars or washing machines) to the market by the end of 2005. Environmental process innovations mean the planned realization of a more environmentally friendly composition of one or more firm-internal processes (e.g. water recycling or flue gas desulphurization) in this period (independent of the realization of environmental product innovations). As aforementioned, three-alternative and four-alternative discrete choice models are examined. Concerning the three-alternative case, environmental innovation type $j = 1$ comprises both an environmental product and a process innovation and $j = 2$ contains either an environmental product or a process innovation, but not both types together. The basic choice alternative $j = 3$ comprises neither an environmental product nor a process innovation. Regarding the four-alternative case, $j = 1$ is the same as in the three-alternative case. The environmental innovation type $j = 2$ comprises an environmental product, but not a process innovation, and $j = 3$ comprises an environmental process, but not a product innovation. The basic choice alternative $j = 4$ is identical to the corresponding basic choice alternative $j = 3$ in the three-alternative discrete choice models.

Concerning environmental organizational measures as major explanatory variables, particularly the effects of two different certifications for EMS are analyzed. The environmental audit regulation EMAS (Eco Management and Audit Scheme, enacted by the European Union) and ISO 14001 (published by the International Standard Organization) are the most important standards for environmental management and eco-audit. Compared with the European EMAS standard, ISO 14001 has a world-wide dimension. Both standards are aimed at the improvement of the environmental protection of organizations. Furthermore, it is argued that the certification for EMAS and ISO 14001 is able to promote environmental innovations (see e.g. Rennings *et al.*, 2003). In the following, the dummy variables "ISO14001$_i$" or "EMAS$_i$" take the value one if at least one facility of firm i is presently (in 2003) certified according to ISO 14001 or EMAS. Beyond these certifications for EMS, the effects of some specific environmental organizational or management measures are also examined. The dummy variables "Lifecycle$_i$" or "Disposal$_i$" take the value one if company i presently evaluates products by means of life cycle considerations or if it presently performs measures concerning the disposal or withdrawal of products. Finally, another certified non-environmental organizational measure, namely the certification for the quality management system ISO 9001, is included. The corresponding dummy variable "ISO9001$_i$" takes the value one if one or more facilities of i are presently certified according to ISO 9001.

This paper also takes into account some insights from industrial economics (see e.g. Pavitt, 1984, Harabi, 1997, Ebling and Janz, 1999, Gottschalk and Janz, 2003), i.e. some variables that may influence the realization of innovations in general and thus environmental innovations as specific kinds of innovations are considered. Due to the highest explanatory power in preliminary investigations (see Ziegler and Rennings, 2004), the following explanatory variables are examined: As market pull factors, the dummy variables "Comp-Pressure$_i$", "Comp-Factor-Client$_i$", and "Comp-Factor-Environment$_i$" that take the value one if company i states that the pressure of competition on the most important sales market has increased within the previous three years (from the beginning of 2001 until 2003), if it states that customer satisfaction is an important factor to deliver competitive advantages on the most important sales market, and if it states that environmental issues are similarly an important factor to deliver competitive advantages on this market (in each case within the previous three years). Concerning R&D activities in general as technological opportunities, the corresponding dummy variable "R&D$_i$" that takes the value one if firm i has carried out R&D activities in the previous year (2002) is considered. Regarding exports, the variable "Exports$_i$" that takes the value one if i has exported in the previous year is analyzed.

Concerning firm size as further explanatory firm-specific characteristic, this paper includes the variable "Ln-Employees-Squared$_i$" that indicates the squared logarithm of the number of salaried employees (divided by 10 due to scaling) of i at the end of the previous year. Regarding present firm age (i.e. the years since foundation or since the last organizational modification, i.e. fusion or splitting), the variables "Reciprocal-Age$_i$" and "Reciprocal-Age-Squared$_i$" that indicate the reciprocal of the age of firm i and the squared reciprocal of the age of i (in each case multiplied by ten due to scaling) are considered. Note that it has also been experimented in preliminary investigations with some other variables concerning firm size and firm age, i.e. with the unsquared and/or squared firm size, the unsquared logarithm of firm size, the unsquared and/or squared firm age, or the unsquared and/or squared logarithm of firm age. However, although the inclusion of different firm size and firm age variables leads to qualitatively similar estimation results, the inclusion of the former three variables has the highest explanatory power. Finally, a dummy variable for the present number of facilities, eight industry dummy variables, and a regional dummy variable (for Western Germany excluding Berlin) are also examined. Note that these explanatory control variables are utilized in all ML or MSL estimations of the multinomial discrete choice models, albeit the corresponding coefficient estimates are not displayed below for brevity. In the following, the indices i are neglected in the analysis of the explanatory variables.

3. Data

The utilized data are based on a telephone survey that has been conducted at the Centre for European Economic Research (ZEW) in Mannheim, Germany, in the second half of 2003. The population is the universe of all German manufacturing companies (NACE-Codes 15-37) with at least 50 employees. Based on the database of "Creditreform" (the largest German credit rating agency), 2,998 addresses were drawn from a stratified representative sample considering two firm size classes (less than 200 and at least 200 employees), eleven industries, and two regions (Western and Eastern Germany). It was tried to reach 2,511 of these 2,998 companies. 112 of these 2511 companies could not be reached, 1811 have refused to participate, and 588 and thus 24.5% of the 2,399 reached companies have participated.[3] The person responsible for production in the company has been the target of the interviews. Note that all firm-specific questions and thus all variables discussed above refer to all facilities of a company in Germany. In contrast, the firm-specific information in the database of "Creditreform" e.g. concerning the number of employees refer to all world-wide facilities. As a consequence, the sample comprises some companies with less than 50 (salaried) employees. Altogether, N =390 of the 588 companies that have participated can be examined in the econometric analysis. Note that besides the companies with missing data regarding the aforementioned variables, firms that have been founded or organizationally modified in the years 2002 or 2003 are excluded since many questions refer to the period from 2001 to 2003. With regard to firm size, industries, and regions, it has been tested whether the corresponding shares for these N =390 firms comply with the shares from the population. The appropriate two-tailed tests have shown that the underlying null hypotheses can never be rejected at the 10% level of significance. Thus, sample selection bias should not be a problem. However, only these tests could be performed since such population information is merely available for these variables. In contrast, there exists e.g. no information regarding corporate environmental innovativeness in Germany.

Concerning the environmental product and process innovations, the companies have been questioned whether they have introduced these measures within the previous three years and whether they plan to realize these measures in the next two years, independent of any innovative activity in the past. Note that the past environmental innovations are not utilized in the econometric analysis because e.g. environmental organizational measures may depend on environmental innovations which makes these explanatory variables endogenous. By including explanatory variables surveyed for the past and future environmental innovations or (in other words) by examining lagged explanatory variables,

[3]This participation rate is rather ordinary in telephone surveys in Germany

this econometric problem can be reduced (see also Arora and Cason, 1995). Regarding the different environmental innovation types in the three-alternative and four-alternative discrete choice models of the N =390 companies, 119 (=30.5%) will realize both an environmental product and a process innovation, 164 (=42.1%) will realize either an environmental product or a process innovation, but not both types together, 137 (=35.1%) will realize an environmental process, but not a product innovation, 27 (=6.9%) will realize an environmental product, but not a process innovation, and 107 (=27.4%) will neither realize an environmental product nor a process innovation.

The ML estimations of the MLM have been performed with the software package STATA. In doing so, the so-called robust estimates of the standard deviations of the coefficient estimates (according to White, 1982, for details see the handbooks of STATA) to calculate the z-statistics are considered. The MSL estimations of the independent and flexible MPM have been carried out with a self-developed GAUSS program. For calculating the z-statistics in the framework of these MSL estimations, the standard deviations of the coefficient estimates have also been estimated robustly (by means of the GHK simulator, for details regarding simulated z-tests see Ziegler, 2001). In the following, the effect of a variable is regarded as insignificant if the absolute value of the corresponding z-statistic is smaller than 1.65. Note that for analyzing the MSL estimations, it has been experimented with different numbers R of random draws in the GHK simulator due to the inconsistent findings about the reasonable number in the literature (see e.g. Börsch-Supan and Hajivassiliou, 1993, Geweke *et al.*, 1997, Ziegler and Eymann, 2001).

4. Results

4.1 Multinomial Logit and Independent Probit Analysis

Table 6.1 reports the estimation results in the logit models. According to this, the variables "Lifecycle" and "Disposal" have a positive influence on environmental innovation type j =1 (compared with the absence of environmental innovations as basic choice alternative) in both MLM at the 5% level of significance. Furthermore, measures concerning the disposal or withdrawal of products have a positive influence on j =2 at the same level of significance. Concerning both types of certified EMS, the variable "ISO14001" has a significantly positive effect on j =1 in both MLM, albeit only at the 10% level of significance. At the same level of significance, it also has a positive influence on j =2 in the three-alternative and on j =3 in the four-alternative logit model. In contrast, the EMAS certification has no significant effect in both MLM. Thus, this analysis implies that specific environmental organizational measures may stimulate environmental product and process innovations to a somewhat larger extent than entire EMS. Measures concerning the disposal

or withdrawal of products additionally have a particularly strong effect on the realization of environmental product innovations. Finally, the ISO 9001 certification, like the ISO 14001 certification, has a similar positive influence on $j = 1$ in both MLM and on $j = 3$ in the four-alternative logit model at the 10% level of significance.

Table 6.1. ML Estimates in the Multinomial Logit Models

Explanatory variables	*Three-alternative model*		*Four-alternative model*		
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
Constant	–4.30	–2.30	–4.34	–6.82	–2.19
	(–3.58)	(–2.43)	(–3.61)	(–2.98)	(–2.27)
ISO14001	0.78	0.76	0.78	0.66	0.77
	(1.69)	(1.77)	(1.68)	(0.92)	(1.78)
EMAS	0.24	–0.29	0.24	–0.24	–0.30
	(0.35)	(–0.45)	(0.36)	(–0.21)	(–0.45)
Lifecycle	1.02	–0.06	0.97	–1.45	0.09
	(2.05)	(–0.11)	(1.97)	(–1.19)	(0.17)
Disposal	1.00	0.64	1.02	1.19	0.53
	(2.84)	(2.02)	(2.90)	(2.21)	(1.63)
ISO9001	0.69	0.40	0.65	–0.19	0.54
	(1.87)	(1.32)	(1.77)	(–0.33)	(1.69)
Comp-Pressure	0.23	–0.44	0.24	-0.02	–0.53
	(0.65)	(–1.42)	(0.69)	(–0.04)	(–1.66)
Comp-Factor-Client	1.40	0.21	1.43	0.44	0.16
	(1.99)	(0.48)	(2.02)	(0.50)	(0.37)
Comp-Factor-Environ.	1.06	0.34	1.06	1.34	0.08
	(2.70)	(0.85)	(2.70)	(2.32)	(0.18)
R&D	1.37	0.68	1.35	0.86	0.66
	(3.04)	(2.01)	(3.01)	(1.35)	(1.87)
Exports	0.30	0.43	0.33	1.92	0.26
	(0.57)	(1.14)	(0.61)	(1.94)	(0.67)
ln(Employees-squared)	0.46	0.45	0.47	0.59	0.42
	(2.69)	(2.66)	(2.72)	(2.13)	(2.39)
Reciprocal-Age	–1.11	–1.09	–1.13	–1.74	–0.96
	(–2.66)	(–3.00)	(–2.69)	(–2.72)	(–2.58)
Reciprocal-Age-squared	2.89	2.65	2.96	4.34	2.30
	(3.08)	(3.09)	(3.11)	(3.44)	(2.58)

Remarks.
Basic choice alternative: No environmental innovation. All estimations include a dummy for the number of facilities, eight industry dummies, and a regional dummy.
Asymptotic z-statistics in parentheses. N=390.

Concerning the variables from industrial economics, firm size and firm age have a strong effect in both MLM at the 5% level of significance. In this respect, a U-shaped relationship between firm age and the probability

of the realization of all considered environmental innovation types (compared with the probability of the realization of no environmental innovation) can be identified. The market pull factors "Comp-Factor-Client" and "Comp-Factor-Environment" have a positive effect on j =1 in both MLM and "Comp-Factor-Environment" additionally on j =2 in the four-alternative logit model at the 5% level of significance. This analysis thus implies that customer satisfaction and environmental issues as an important competition factor also stimulate both environmental product and process innovations together. Environmental issues as an important competition factor additionally have a specific effect on the realization of environmental product innovations. Finally, R&D activities strongly stimulate both environmental product and process innovations. The corresponding variable has a positive effect on j =1 in both MLM and on j =2 in the three-alternative logit model at the 5% level of significance as well as on $j = 3$ in the four-alternative logit model at the 10% level of significance.

After this, the ML estimations of the MLM are compared with the MSL estimations of the independent MPM for R =200 random draws in the GHK simulator. Table 6.2 reports the corresponding estimation results in the independent three-alternative and four-alternative probit models. Note that the coefficient estimates in Table 6.1 and Table 6.2 are not directly comparable since the underlying standard normal and Type I extreme value distributions in the corresponding multinomial discrete choice models have different variances. Taking this into consideration, extremely strong analogies between the estimation results appear so that the conclusions from the MLM analysis can widely be maintained. The only qualitative difference is that the level of significance regarding the influence of a few explanatory variables slightly changes. In this respect, the positive effects of the variables "Disposal" and "R&D" on environmental innovation type j =2 in the independent three-alternative probit model have a little higher level of significance than the corresponding positive effects in the three-alternative logit model. In contrast, the positive influence of the variable "Exports" on j =2 in the independent four-alternative probit model has a comparatively lower level of significance. It should however be particularly emphasized that the coefficient estimates and z-statistics in both independent MPM are extremely reliable and independent of the number R of random draws in the GHK simulator. The estimation results for R =1,000, R =500, R =50, and largely even for R =10 are qualitatively nearly identical to those in Table 6.2.

4.2 Flexible Multinomial Probit Analysis

In contrast, the MSL estimations of the flexible MPM are rather unreliable. Variations in R lead to quite different estimation results, particularly concerning the variance covariance parameter estimates and the z-statistics. But an

Table 6.2. SML Estimates in the Independent Multinomial Probit Models ($R = 200$)

Explanatory variables	*Three-alternative model*		*Four-alternative model*		
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
Constant	–3.04	–1.68	–3.06	–4.38	–1.52
	(–3.57)	(–2.25)	(–3.63)	(–3.68)	(–2.05)
ISO14001	0.56	0.52	0.54	0.53	0.52
	(1.70)	(1.66)	(1.68)	(1.31)	(1.66)
EMAS	0.23	–0.18	0.21	–0.34	–0.21
	(0.47)	(–0.37)	(0.44)	(–0.52)	(–0.43)
Lifecycle	0.76	–0.12	0.76	–0.72	0.04
	(2.17)	(–0.32)	(2.17)	(–1.16)	(0.11)
Disposal	0.72	0.44	0.71	0.82	0.35
	(2.87)	(1.87)	(2.86)	(2.71)	(1.48)
ISO9001	0.47	0.32	0.44	–0.08	0.41
	(1.78)	(1.32)	(1.70)	(–0.24)	(1.69)
Comp-Pressure	0.19	-0.37	0.21	0.01	–0.41
	(0.73)	(–1.56)	(0.82)	(0.04)	(–1.71)
Comp-Factor-Client	0.97	0.17	0.97	0.44	0.10
	(2.01)	(0.48)	(2.03)	(0.87)	(0.28)
Comp-Factor-Environ.	0.80	0.29	0.81	0.90	0.08
	(2.82)	(0.97)	(2.88)	(2.57)	(0.27)
R&D	1.01	0.51	0.99	0.54	0.50
	(3.20)	(1.94)	(3.16)	(1.52)	(1.86)
Exports	0.15	0.35	0.17	1.12	0.19
	(0.40)	(1.16)	(0.47)	(2.49)	(0.63)
ln(Employees-squared)	0.36	0.35	0.35	0.41	0.32
	(2.75)	(2.69)	(2.74)	(2.40)	(2.43)
Reciprocal-Age	–0.83	–0.85	–0.79	–1.24	–0.74
	(–2.79)	(–3.14)	(–2.69)	(–3.42)	(–2.72)
Reciprocal-Age-squared	2.17	2.06	2.10	3.05	1.75
	(3.33)	(3.34)	(3.24)	(4.06)	(2.80)

Remarks.
Basic choice alternative: No environmental innovation. All estimations include a dummy for the number of facilities, eight industry dummies, and a regional dummy.
Asymptotic z-statistics in parentheses. N=390.

increase in R does not lead to more reliable estimation results so that the incorporation of the GHK simulator into the ML estimations is not crucial for these fundamental difficulties. These problems are obviously a consequence of the sole inclusion of firm-specific characteristics as explanatory variables (such as firm size or firm age) that do not vary between the choice alternatives as it is typical for some empirical applications. Without the incorporation of choice-specific attributes, the practical identification of MSL or ML estimations of flexible MPM is often difficult (see Keane, 1992). Although such models are

formally identified, they can exhibit very small variations in the log-likelihood function from its maximum over a wide range of parameter values (see also the estimation results in flexible three-alternative probit models in Rennings *et al.*, 2004). Note that not only variations in R, but also other classifications of the environmental innovation types $j = 1, \ldots, J$ (e.g. the change between $j = 1$ and $j = 3$) and thus other variance covariance parameter restrictions lead to quite different estimation results (see Ziegler and Rennings, 2004[4]). Furthermore, different starting values at the beginning of the MSL maximization process for the parameters and different tolerance limits of the gradient of the log-likelihood function as discussed in Rennings *et al.* (2004) have strong effects in this respect.

Table 6.3 compares the estimation results in the flexible three-alternative probit models for $R = 10$, $R = 200$, and $R = 1000$ random draws in the GHK simulator. For $R = 10$, the coefficient estimates of the explanatory variables have the same sign as those in the corresponding independent MPM and furthermore they have a similar size (see Table 6.2). But this is not surprising since the estimates of the variance covariance parameters σ_1 and $\text{corr}(\varepsilon_{i1}, \varepsilon_{i2})$ are not very different from the restrictions σ_1=1 and $\text{corr}(\varepsilon_{i1}, \varepsilon_{i2}) = 0$ in the independent three-alternative probit models. In contrast, the variance covariance parameter estimates are clearly higher for R =200 or R =1,000. As a consequence, the absolute coefficient estimates of the explanatory variables are mostly higher with respect to j =2 and always clearly higher with respect to j =1. Another result is that the estimated standard deviations of the coefficient estimates of the explanatory variables with respect to j =1 are also clearly higher for R =200 or R =1000. Due to the consequential low z-statistics, no explanatory variable has any significant effect on j =1 for R =1,000. For R =10, firm size and firm age have a similar effect on the probability of the realization of all considered environmental innovation types at the 5% level of significance as according to the MLM and independent MPM analysis. Furthermore, the variables "ISO14001", "Disposal", and "R&D" have a positive influence on $j = 1$ and j =2 at the 10% level of significance. However, many other effects have clearly higher levels of significance than in the ML or MSL estimations of the corresponding MLM or independent MPM (see Table 6.1 and Table 6.2).

Despite these obvious indications of fundamental practical identification problems, it should be noticed that the log-likelihood functions in these flexible three-alternative probit models always converge to a maximum, irrespective of the number R of random draws in the GHK simulator (note that it has also

[4]Concerning the comparison of the estimation results in that study with those in Table 6.3 and Table 6.4 in this paper, note that the description of the environmental innovation types in Ziegler and Rennings (2004) wrongly implies the same structure as discussed above.

Table 6.3. SML Estimates in the Flexible Three-alternative Probit Models

Explanatory variables	$R = 10$		$R = 200$		$R = 1000$	
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
Constant	–3.02	–1.80	–6.35	–1.98	–5.75	–2.04
	(–1.43)	(–2.33)	(–0.88)	(–2.35)	(–0.40)	(–2.12)
ISO14001	0.63	0.59	0.90	0.62	0.86	0.63
	(1.65)	(1.76)	(1.13)	(1.78)	(0.83)	(1.82)
EMAS	0.22	–0.14	0.58	0.00	0.49	–0.04
	(0.36)	(–0.26)	(0.47)	(0.01)	(0.25)	(–0.07)
Lifecycle	0.75	–0.00	1.89	0.33	1.56	0.25
	(0.92)	(–0.01)	(0.78)	(0.80)	(0.38)	(0.49)
Disposal	0.74	0.48	1.28	0.56	1.17	0.57
	(1.84)	(1.79)	(1.04)	(2.27)	(0.55)	(2.11)
ISO9001	0.44	0.30	0.99	0.44	0.85	0.40
	(1.14)	(0.99)	(0.99)	(1.71)	(0.52)	(1.50)
Comp-Pressure	0.14	–0.33	0.51	–0.31	0.45	–0.28
	(0.30)	(–0.92)	(0.46)	(–1.21)	(0.20)	(–0.85)
Comp-Factor-Client	0.94	0.26	1.83	0.34	1.61	0.33
	(1.09)	(0.60)	(0.75)	(0.93)	(0.37)	(0.73)
Comp-Factor-Environ.	0.81	0.34	1.60	0.59	1.45	0.56
	(1.34)	(0.62)	(1.08)	(1.72)	(0.49)	(1.36)
R&D	1.05	0.60	1.90	0.67	1.75	0.69
	(1.89)	(1.71)	(1.05)	(2.48)	(0.50)	(2.08)
Exports	0.22	0.38	0.32	0.42	0.34	0.41
	(0.48)	(1.20)	(0.43)	(1.30)	(0.53)	(1.27)
ln(Employees-squared)	0.37	0.35	0.53	0.39	0.52	0.39
	(2.44)	(2.66)	(1.60)	(2.82)	(0.86)	(2.78)
Reciprocal-Age	–0.89	–0.93	–1.16	–0.92	–1.07	–0.91
	(–2.45)	(–3.12)	(–1.72)	(–3.21)	(–1.35)	(–3.24)
Reciprocal-Age-squared	2.34	2.26	3.15	2.29	2.91	2.28
	(2.61)	(3.02)	(1.78)	(3.51)	(1.14)	(3.50)
$\hat{\sigma}_1$	1.12		3.05		2.62	
$\hat{Corr}(\varepsilon_{i1}, \varepsilon_{i2})$	0.37		0.80		0.72	

Remarks.
Basic choice alternative: No environmental innovation. All estimations include a dummy for the number of facilities, eight industry dummies, and a regional dummy.
Asymptotic z-statistics in parentheses. N=390.

been experimented with R =50 and R =500). In this respect, the MSL estimations of flexible four-alternative probit models are even less reliable. It has again been experimented with R =10, R =50, R =200, R =500, and R =1,000 random draws in the GHK simulator. The result is that the log-likelihood functions do not converge to a maximum for R =10, R =200, and R =500. Therefore, an increase in R again does not systematically lead to

more reliable estimation results. Table 6.4 reports the estimation results in the flexible four-alternative probit models for R =20 (this case is examined as an example of a small R) and R =1,000. As in the MSL estimations of the flexible three-alternative probit models according to Table 6.3, the coefficient estimates of the explanatory variables and the variance covariance parameter estimates vary strongly for both R. It seems again that the higher estimate of σ_1 for R =1,000 leads to clearly higher estimated standard deviations of the coefficient estimates of the explanatory variables with respect to j =1.

Altogether, the levels of significance regarding the effects of many explanatory variables are (irrespective of R) clearly higher than in the MSL estimations of the corresponding independent MPM according to Table 6.2. Of all ML or MSL estimations of the four-alternative discrete choice models, the effects of firm size and firm age on environmental innovation types j =1 and j =3 and of the variable "R&D" on j =1 are most robust. It should however be mentioned that the influences of these explanatory variables on j =1 are insignificant in the flexible three-alternative probit model for R =1,000. Therefore, the flexible MPM analysis provides few new insights compared with the MLM and independent MPM analysis. It is particularly not clear whether the different estimation results in the independent MPM and MLM on the one hand and in the flexible MPM on the other hand are due to a potential misspecification of the former models or due to the unreliable estimation results in the latter models. Concerning these unreliabilities, it should be emphasized that they obviously grow if the number J of environmental innovation types increases and that the difficulties do not decrease if the number R of random draws in the GHK simulator rises.

These conclusions can also be confirmed by a small Monte Carlo experiment that has been performed based on the underlying firm level data set. First of all, shortened three-alternative and four-alternative logit and independent probit models have been estimated (with $R = 10$ in the MSL estimations of the independent MPM). After this, different data generating processes (DGP) have been constructed and replicated 20 times based on the ten explanatory variables (and one constant) for the N =390 companies and on the resulting coefficient estimates. Note that those explanatory dummy variables whose shares of ones are smaller than 0.2 or higher than 0.8 have been excluded. The reason for this procedure is that according to preliminary investigations, the inclusion of such explanatory variables leads to a one-to-one linkage between one dummy variable value and $D_{i2} = 1$ or D_{i2} =0 $(\forall i)$ in the four-alternative case in some replications of the DGP.[5] The three-alternative and four-alternative logit DGP

[5]The examined explanatory variables are therefore: "ISO14001", "Disposal", "ISO9001", "Comp-Pressure", "Comp-Factor-Environment", "R&D", "Ln-Employees-Squared", "Reciprocal-Age", "Reciprocal-Age-Squared", and the dummy variable for the number of facilities.

Table 6.4. SML Estimates in the Flexible Four-alternative Probit Models

Explanatory variables	$R = 20$			$R = 1,000$		
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
Constant	–2.09	–2.81	–1.50	–4.27	–3.13	–1.64
	(–2.51)	(–0.96)	(–2.13)	(–1.58)	(–0.80)	(–1.98)
ISO14001	0.46	0.52	0.47	0.73	0.49	0.59
	(1.85)	(1.64)	(1.59)	(1.41)	(1.42)	(1.78)
EMAS	0.11	–0.06	–0.16	0.33	–0.34	–0.19
	(0.29)	(–0.08)	(–0.35)	(0.42)	(–0.46)	(–0.34)
Lifecycle	0.35	–0.14	–0.04	1.31	–0.39	0.25
	(1.17)	(–0.09)	(–0.11)	(1.30)	(–0.27)	(0.45)
Disposal	0.54	0.62	0.34	0.93	0.65	0.39
	(2.78)	(1.65)	(1.46)	(1.64)	(1.47)	(1.47)
ISO9001	0.28	0.07	0.41	0.72	0.04	0.46
	(1.27)	(0.13)	(1.70)	(1.48)	(0.06)	(1.64)
Comp-Pressure	0.05	–0.00	–0.40	0.31	–0.08	–0.39
	(0.21)	(–0.00)	(–1.60)	(0.53)	(–0.20)	(–1.43)
Comp-Factor-Client	0.59	0.49	0.09	1.26	0.30	0.15
	(1.53)	(1.27)	(0.26)	(1.21)	(0.76)	(0.38)
Comp-Factor-Environ.	0.58	0.67	0.04	1.16	0.78	0.23
	(2.04)	(1.64)	(0.13)	(1.40)	(1.16)	(0.52)
R&D	0.71	0.58	0.48	1.42	0.49	0.59
	(2.85)	(2.16)	(1.61)	(1.74)	(1.38)	(1.86)
Exports	0.23	0.63	0.15	0.21	0.88	0.25
	(0.72)	(0.63)	(0.51)	(0.37)	(0.86)	(0.62)
ln(Employees-squared)	0.31	0.34	0.31	0.42	0.34	0.34
	(3.04)	(2.15)	(2.42)	(2.13)	(2.18)	(2.45)
Reciprocal-Age	–0.67	–0.95	–0.68	–0.98	–1.03	–0.79
	(–2.94)	(–1.23)	(–2.53)	(–1.89)	(–2.49)	(–2.43)
Reciprocal-Age-squared	1.76	2.33	1.61	2.61	2.59	1.91
	(3.36)	(1.41)	(2.60)	(2.07)	(2.49)	(2.38)
$\hat{\sigma}_1$		0.30			2.03	
$\hat{\sigma}_2$		0.48			0.20	
$\hat{Corr}(\varepsilon_{i1}, \varepsilon_{i2})$		0.45			0.25	
$\hat{Corr}(\varepsilon_{i1}, \varepsilon_{i3})$		0.15			0.46	
$\hat{Corr}(\varepsilon_{i2}, \varepsilon_{i3})$		-0.21			-0.08	

Remarks.
Basic choice alternative: No environmental innovation. All estimations include a dummy for the number of facilities, eight industry dummies, and a regional dummy.
Asymptotic z-statistics in parentheses. N=390.

or independent probit DGP are thus characterized by the coefficient estimates in the MLM or in the independent MPM and by the corresponding variance covariance matrices of ε_i as discussed in section 2.2. Another DGP is identified

by the coefficient estimates in the shortened independent three-alternative probit model and some correlations between the components of ε_i (with $\sigma_1 = 1.5$ and corr$(\varepsilon_{i1}, \varepsilon_{i2})$ =0.5).[6] Note that the parameter estimates in the shortened flexible MPM have not been utilized since the corresponding MSL estimations are even less reliable for different R (particularly regarding the variance covariance parameters) than the MSL estimations of the flexible MPM as discussed above. Furthermore, the estimation results in the flexible MPM are also analyzed if the DGP are characterized by the independent MPM.

One result (such as in the empirical analysis of the flexible three-alternative probit models discussed above) is that the log-likelihood functions in the MSL estimations of the shortened flexible three-alternative probit models always converge to a maximum over the 20 replications of both DGP for R =10, R =50, and R =200 random draws in the GHK simulator. The most reliable parameter estimates arise for R =10 if the underlying DGP is characterized by the independent MPM. In contrast, there are some outliers for R =50 and R =200 and particularly if the underlying DGP comprises the aforementioned correlations. The spread of the parameter estimates is in this case extremely large, notably for R =200. Furthermore, the parameter estimates often reach regions on the boundary of the parameter space, e.g. the estimates of σ_1 and corr$(\varepsilon_{i1}, \varepsilon_{i2})$ are 411.34 and 1.00 in one of the replications of this DGP for R =200. Such fundamental practical identification problems occur to an even larger extent in the MSL estimations of the flexible four-alternative probit models since the log-likelihood function repeatedly does not converge to a maximum over the 20 replications of the DGP that is characterized by the independent MPM. The numbers of these problems are one, seven, and nine for R =10, R =50, and R =200. Therefore, an increase in R rather leads to a rise of the problems.

In contrast, the MSL estimations of the independent three-alternative probit models based on the DGP that comprises the aforementioned correlations are always very reliable. Compared with the corresponding MSL estimations based on the DGP that is characterized by the independent MPM, the only differences are the somewhat higher biases for the coefficient estimates due to the underlying misspecification. In this respect, it should be emphasized that the MSL estimations of the independent three-alternative and four-alternative probit models lead to extremely reliable and accurate results if the underlying DGP is correctly characterized by the independent MPM. Irrespective of

[6]Note that four-alternative probit models incorporating some (positive) correlations between the components of ε_i could not be analyzed since even the inclusion of only these ten explanatory variables repeatedly leads to a one-to-one linkage between one dummy variable value and D_{i2} =1 or D_{i2}=0 ($\forall i$) in many replications of the DGP. This is due to the small number of realizations of environmental innovation type j =2. Other data sets should therefore be examined in future Monte Carlo experiments to analyze MPM with a higher number J of choice alternatives.

$R = 10$, $R = 50$, or $R = 200$, the average absolute biases over all coefficient estimates and over all 20 replications of the DGP are only 0.06 in the three-alternative and only 0.08 in the four-alternative case. These biases (and also the root mean squared errors) surprisingly have a similar extent as those in the ML estimations of MLM if the underlying DGP is likewise correctly characterized by the MLM. The average absolute biases are 0.09 in the three-alternative and 0.07 in the four-alternative logit model.

5. Conclusions

This paper examines determinants of different types of environmental innovations in the German manufacturing sector by comparing the estimation results in flexible and restrictive multinomial discrete choice models. The MLM and independent MPM analysis implies that some specific environmental organizational measures, technological opportunities, and market pull factors have a significantly positive effect on both environmental product and process innovations. In contrast, the effects of certified EMS are statistically less reliable. While the world-wide ISO 14001 standard has a significantly weak positive influence, the European EMAS standard has no significant effect on environmental innovations at all. Thus, some specific environmental organizational measures obviously may stimulate environmental innovations to a somewhat larger extent than entire EMS. In the framework of the discussion of so-called soft environmental policy instruments such as initiatives concerning the certification for EMS, this means that the contribution of such programs to environmental technological innovation is not completely clear. In the future, it seems to be promising to analyze how far single measures such as life cycle considerations are applied. Concerning some incentives for the realization of environmental organizational measures or more directly for the realization of environmental technological innovations, policy could consider the opportunity of rewards e.g. in public procurement apart from only supporting initiatives concerning the certification for EMS.

Taking the results of the MLM and independent MPM analysis into consideration, the flexible MPM analysis provides few new insights. The MSL estimations of the flexible MPM are rather unreliable due to the arising practical identification difficulties as a consequence of the sole inclusion of firm-specific characteristics as explanatory variables. Therefore, it can be concluded that the applicability of these flexible multinomial discrete choice models without the incorporation of choice-specific attributes is rather limited in practice. It should be emphasized that these estimation problems have only been sparsely noticed in the literature. In this respect, the theoretical advantages of the flexible MPM compared with the restrictive MLM (or also independent MPM) can be questioned in practice. However, it should also be emphasized that this

conclusion can merely be maintained if firm-specific characteristics are only included. If choice-specific attributes are additionally or solely included as explanatory variables, the MSL estimations of flexible MPM (incorporating the GHK simulator) lead to very reliable results (see e.g. the Monte Carlo experiments in Börsch-Supan and Hajivassiliou, 1993, Geweke *et al.*, 1994, Stern, 2000, or Munizaga *et al.*, 2000, for one-period MPM, and in Geweke *et al.*, 1997, or Ziegler and Eymann, 2001, for multiperiod MPM).

Concerning the sole inclusion of firm-specific characteristics as explanatory variables into the flexible MPM, this paper cannot give a conclusive answer on the reliability of the corresponding MSL estimations. The small Monte Carlo experiment only confirms the main conclusions from the empirical analysis in accordance with the study of Keane (1992). According to this, the incorporation of simulation methods is not crucial for the practical identification problems since the difficulties do not decrease if the number R of random draws in the GHK simulator rises. Furthermore, the difficulties grow if the number J of environmental innovation types increases. Note that only three-alternative and four-alternative discrete choice models are analyzed. One could therefore imagine that the application of flexible MPM with a high number J of choice alternatives as it is suitable in other empirical studies can lead to even more practical identification problems, although their use just in these cases is theoretically clearly more advantageous compared with the use of restrictive MLM. Thus, a systematic comparison between the estimation results in flexible MPM with high J and those with rather small J as in this paper would be desirable in the future. In this respect, the analysis of one-period MPM compared with multiperiod MPM based on panel data (see e.g. the application in Börsch-Supan *et al.*, 1992) would also be interesting.

Chapter 7

MIXED LOGIT WITH BOUNDED DISTRIBUTIONS OF CORRELATED PARTWORTHS

Kenneth Train
Department of Economics
University of California, Berkeley *
Kenneth.Train@NERA.com

Garrett Sonnier
Anderson School of Management
University of California, Los Angeles
garrett.sonnier@anderson.ucla.edu

Abstract The use of a joint normal distribution for partworths is computationally attractive, particularly with Bayesian MCMC procedures, and yet is unrealistic for any attribute whose partworth is logically bounded (e.g., is necessarily positive or cannot be unboundedly large). A mixed logit is specified with partworths that are transformations of normally distributed terms, where the transformation induces bounds; examples include censored normals, log-normals, and S_B distributions which are bounded on both sides. The model retains the computational advantages of joint normals while providing greater flexibility for the distributions of correlated partworths. The method is applied to data on customers' choice among vehicles in stated choice experiments. The flexibility that the transformations allow is found to greatly improve the model, both in terms of fit and plausibility, without appreciably increasing the computational burden.

*A Gauss routine and manual to implement the procedures described in this paper are available on Train's website at http:\\elsa.berkeley.edu\ ∼train. We are grateful for comments from Peter Rossi on an earlier version of this paper.

Keywords: Mixed logit, correlated partworths, bounded distributions, choice of vehicle.

1. Introduction

Mixed logit is a flexible discrete choice model that incorporates random variation in partworths.[1] McFadden and Train (2000) show that mixed logit can approximate any random utility choice model to any degree of accuracy through appropriate specification of distributions of the partworths. Procedures for estimating mixed logits have been developed within both the classical (e.g., Revelt and Train, 1998, Brownstone and Train, 1999) and Bayesian (Allenby, 1997; Sawtooth Software, 1999) traditions.

Mixed logit models have been used for environmental analysis in numerous contexts, including: households' response to rebates on energy-efficient appliances (Revelt and Train, 1998); the impact of fish stock, which is affected by water quality, on anglers' choice of fishing site (Train, 1998); the demand for wind, hydro and other forms of renewable power generation (Goett *et al.*, 2000); and consumers' willingness to pay for water service improvements (Hensher *et al.*, 2004). Consumers' choice of vehicle, which is the application in the current paper, is particularly important for environmental analysis since energy consumption and emissions are largely dependent on this choice. Mixed logits of vehicle choice have been previously estimated by Brownstone and Train (1999), Brownstone *et al.* (2000), and Train and Winston (2004).

The distribution of partworths is critical in any application. Normal distributions are relatively easy to implement in both the classical and Bayesian methods. However, since the normal is unbounded on each side of zero, its use in many setting is inappropriate. A normal distribution for a price coefficient implies that some share of the population actually prefer higher prices. Also, since the normal distribution overlaps zero, its use for a price coefficient can preclude the calculation of willingness-to-pay statistics: The willingness-to-pay for an attribute is the partworth of that attribute divided by the price coefficient. Since division by zero is undefined, and division by a number arbitrarily close to zero gives an arbitrarily large result, the mean and variance of willingness-to-pay need not exist when the price coefficient is normal.[2] A normal distribution is also inappropriate for the partworth of a desirable attribute that is valued (or, at worst, ignored) by all customers or an undesirable attribute that is disliked (or ignored) by all customers. Similarly, when an attribute con-

[1]The partworth of an attribute is the coefficient of the attribute in the utility function. The term is used extensively in marketing, and we adopt it here because it is more succinct than "utility coefficient" and more specific than "coefficient."

[2]The applications cited in the previous paragraph specify the price coefficient to be fixed or log-normal in order to avoid this issue. See, e.g., Merrill (1928), Geary (1930), Fieller (1932), Marsaglia (1965), and Hinkley (1969) on the distribution of a ratio of two normally distributed terms.

sists of various levels, the partworth for each higher level must logically be no smaller than the partworth for each lower level; normal distributions do not embody this requirement.

Bounded distributions can and have been used in mixed logits estimated by both the classical and Bayesian procedures (e.g., Bhat, 1998, 2000; Revelt and Train, 1998; Train, 1998; Revelt, 1999; Boatwright *et al.*, 1999; Brownstone and Train, 2000; Johnson, 2000; Train 2001). However, each estimation procedure, while feasible with bounded distributions, entails numerical difficulties that are intrinsic to its form, as described and illustrated by Train (2001). In particular: Classical procedures handle triangular, truncated normal, and similarly bounded distributions easily while Bayesian procedures are relatively slow with these distributions. On the other hand, fully correlated partworths are difficult to handle in classical procedures due to the proliferation of parameters, while the Bayesian procedures accommodate these correlations readily. Obtaining partworths that are bounded *and* correlated has been relatively difficult with either procedure.

Bayesian procedures operate effectively with normals because of the convenient posteriors that arise with normals. In this paper, we build upon the observation in Train (2001) that the Bayesian procedures operate as effectively with log-normals as normals because the log-normal is simply a transformation of the normal that does not entail any other parameters. This concept is expanded by using other transformations that provide various types of bounded distributions. These transformations can operate on correlated normals to provide correlated partworths with bounded distributions. The numerical advantages of the Bayesian procedures with correlated normals are retained while having partworths whose distributions are bounded.

Many useful distributions can be obtained as transformations of normals. Let scalar β be normally distributed with mean b and variance ω. Bounded distributions are obtained through the following kinds of transformations. These transformations are weakly monotonic (non-decreasing in β) and depend only on β without utilizing b and ω.

- Log-normal. The transformation is $c = \exp(\beta)$. The distribution is bounded below by zero. It is useful for the partworths of attributes that are liked by all customers. The sign is reversed for undesirable attributes, such as a price variable, such that the partworth is necessarily negative.

- Normal censored from below at zero. The transformation is $c = \max(0, \beta)$. There is a mass at zero, with the density above zero being the same as the normal density of β. The share at zero is $\Phi(-b/\omega)$, where Φ is the standard normal cumulative distribution. This transformation is useful for partworths of an attribute that some customers do not care about (i.e., are indifferent to its presence and simply ignore) and other

customers find desirable. The estimation of b and ω determines the share massed at zero and the share distributed above zero.

- Johnson's (1949) S_B distribution.[3] The transformation $c = e^{\beta}/(1+e^{\beta})$ creates a partworth that is distributed between zero and one, with mean, variance and shape determined by the mean and variance of β. [4] For a distribution that has support from ℓ to u, the transformation is $c = \ell + (u-\ell) \times e^{\beta}/(1+e^{\beta})$. The S_B distribution is useful for a variety of purposes. S_B densities can be shaped like log-normals but with an upper bound and with thinner tails below the bound. S_B densities are more flexible than log-normals: they can be shaped like a plateau with a fairly flat area between drop-offs on each side (as in Figure 2 for our application) and can even be bi-modal. When a lower bound other than zero is specified, the distribution is useful for an attribute that some people like and others dislike but for which there is a limit for how much the person values having or avoiding the attribute.

For multiple partworths, β is generalized to be a vector with length equal to the number of partworths, with mean vector b and variance matrix Ω. Each partworth is defined as a transformation of the corresponding element of β. The covariance among the elements of β induces covariance among the partworths. As such, the procedure allows for correlated partworths under any combination of the above distributions.

Numerous authors have implemented log-normal distributions within mixed logit, though usually without allowing full correlation; see, e.g., Bhat (1998, 2000), Train (1998), and Revelt and Train (1998).[5] R. Johnson (2000) examined censored normals and found that they provided more reasonable results and better fit than uncensored normals in his application. The use of the S_B distribution seems to be new. We will investigate its usefulness in the context of our application.

The computational advantage of the method rests on the simplicity of the posteriors on b and Ω that arise, as described in the next section, when the transformation of β does not depend on b and Ω. Transformations that depend on b and Ω can be useful in some settings, but do not provide the same simplicity. For example, truncated normals cannot be accommodated within our procedure because the necessary transformation entails b and Ω rather than de-

[3] See also Johnson and Kotz, 1970, p. 23.

[4] As Johnson and Kotz note, the formulas for the moments are very complex. We calculate them through simulation as described section 4. The median is $1/(1+\exp(b/\sqrt{\omega}))$.

[5] Experience indicates that the parameters of log-normal distributions are hard to estimate with classical procedures, due to the fact that the log-likelihood surface is highly non-quadratic. The Bayesian procedure avoids this difficulty.

pending only on β.[6] Since b and Ω affect the utility of each customer through this transformation, the posteriors for b and Ω conditional on the β's depend on the choices of the respondents and no longer have the convenient form that we utilize. Boatwright *et al.* (1999) provide MCMC methods for truncated normals. These methods, which can be generalized to essentially any bounded distribution, are considerably more difficult and slower computationally than those we utilize in this paper. The question for the researcher for a particular application is whether transformations of normals that do not depend on b and Ω can adequately represent the relevant distributions of partworths. If so, the simple methods in this paper can be exploited; if not, the methods of Boatwright *et al.* can be used.[7]

While the estimation procedure that we describe is Bayesian, the results can be interpreted from either a Bayesian or classical perspective. Bayesian interpretation is of course straightforward since the procedure itself is Bayesian. Classical interpretation is less well recognized. The Bernstein-von Mises theorem (see, e.g., Train, 2003, for a discussion with historical references) establishes that, under conditions that are maintained in our specification, the mean of the Bayesian posterior is a classical estimator that is asymptotically equivalent to the maximum likelihood estimator. The theorem also establishes that the covariance of the posterior is the asymptotic sampling covariance of this estimator. The results from the Bayesian procedures can therefore be interpreted by a classical researcher in the same way that the researcher would interpret estimates obtained by maximum likelihood. To facilitate this interpretation, we present our results in the format that is standard for classically estimated models, namely by reporting the parameter estimates (which are the posterior means) and their standard errors (the posterior standard deviations).

In section 2, we describe Bayesian estimation of a mixed logit with normally distributed partworths. We then show in section 3 how this procedure is changed to accommodate transformations of the normal. We apply the method in section 4 to data on customers' choice among vehicles.

2. Mixed logit with normally distributed partworths

The behavioral derivation of mixed logit with repeated choices is given by Revelt and Train (1998) and Train (1998) for general distributions of partworths. The Bayesian procedure for estimating the model with normally distributed partworths was developed by Allenby (1997) and implemented by

[6] E.g., a one-dimensional normal truncated below at zero is created as $c = \Phi^{-1}(m(1-z)+z)\cdot\omega + b$ where $z = \Phi((\beta - b)/\omega)$ and $m = \Phi(-b/\omega)$.

[7] Classical estimation procedures accommodate truncated normals as readily as normals; see, e.g., Revelt (1999.) However, as stated above, classical procedure have difficulty dealing with correlated partworths due to the proliferation of parameters.

Sawtooth Software (1999).[8] We give the derivation and Bayesian procedure under normally distributed partworths in this section. The generalization to transformation of normals is described in the following section.

Person n faces a choice among J alternatives in each of T time periods. J can be as small as 2, and T can be as small as 1. The person's utility from alternative j in period t is

$$U_{njt} = \beta_n' x_{njt} + \varepsilon_{njt},$$

where $\varepsilon_{njt} \sim$ iid extreme value and $\beta_n \sim N(b, \Omega)$. The vectors of variables x_{njt} and partworths β_n have length K. Person n chooses alternative i in period t if $U_{nit} > U_{njt}\ \forall j \neq i$. Denote the person's chosen alternative in period t as y_{nt}, the person's sequence of choices over the T time periods as $y_n = \langle y_{n1}, \ldots, y_{nT} \rangle$, and the set of $y_n \forall n$ as Y. Conditional on β_n, the probability of person n's sequence of choices is the product of standard logit formulas:

$$L(y_n \mid \beta_n) = \prod_t \frac{e^{\beta_n' x_{ny_{nt}t}}}{\sum_j e^{\beta_n' x_{njt}}}.$$

The unconditional probability is the integral of $L(y_n \mid \beta_n)$ over all values of β_n weighted by the density of β_n:

$$P_n(y_n \mid b, \Omega) = \int L(y_n \mid \beta_n) g(\beta_n \mid b, \Omega) d\beta_n. \tag{2.1}$$

where $g(\cdot)$ is the multivariate normal density. This unconditional probability is called the mixed logit choice probability, since it is a product of logits mixed over a density of partworths.

For Bayesian analysis, it is necessary to specify the priors on the model parameters b, Ω, and $\beta_n \forall n$. Since we have already specified β_n to be normal with mean b and variance Ω, the prior on each β_n is proportional to this density times the prior on b and Ω.[9] We specify the prior on b to be a diffuse normal, denoted $N(b \mid 0, \Theta)$, which has zero mean and variance Θ sufficiently large that the density is effectively flat from a numerical perspective. The advantage

[8]Related methods for probit models were developed by Albert and Chib (1993), McColluch and Rossi (1994), and Allenby and Rossi (1999). Bayesian procedures for non-mixed logits are discussed by Koop and Poirier (1993, 1996) and Poirier (1994, 1996).

[9]Several terms have been used for these parameters. (1) Often, b and Ω are called population parameters that describe the distribution of customer-level β_n's in the population. With this usage, the distribution $g(\beta_n \mid b, \Omega)$ is interpreted as the actual distribution of partworths in the population. (2) In Bayesian analyses especially, b and Ω are often called hyper-parameters, since the prior on each β_n depends on b and Ω which themselves have priors. Under this usage, $g(\beta_n \mid b, \Omega)$ is interpreted as an aspect of the researcher's prior information about β_n. (3) Sometimes, the β_n's are called nuisance parameters, to reflect the concept that they are incorporated into the analysis simply (under this usage) to facilitate estimation of b and Ω.

of a normal prior on b is that it provides a conditional posterior on b (*i.e.*, conditional on $\beta_n \forall n$ and Ω) that is normal and hence easy to draw from, while the large variance assures that the prior has minimal (effectively no) influence on the posterior. The standard diffuse prior on Ω is inverted Wishart with K degrees of freedom and parameter KI where I is the K-dimensional identity matrix. This density is denoted $IW(\Omega \mid K, KI)$. It provides a conditional posterior on Ω that is IW and hence easy to draw from. The joint posterior on $\beta_n \forall n$, b and Ω is

$$\Lambda(\beta_n \forall n, b, \Omega \mid Y) \propto \prod_n L(y_n \mid \beta_n) \cdot g(\beta_n \mid b, \Omega) \cdot N(b \mid 0, \Theta) \cdot IW(\Omega \mid K, KI).$$

Information about the posterior is obtained by taking draws from the posterior and calculating relevant statistics, such as moments, over these draws. Gibbs sampling is used to facilitate the taking of draws (see Casella and George, 1992, for a general explanation of Gibbs sampling.) In particular, draws are taken sequentially from the conditional posterior of each parameter given the previous draw of the other parameters. The sequence of draws from the conditional posteriors converges to draws from the joint posterior.

The conditional posterior distributions in this model are especially convenient. Given β and Ω, the posterior on b is $N(\bar{\beta}, \Omega/N)$ with $\bar{\beta} = (1/N)\sum \beta_n$.[10] This distribution is easy to draw from: A draw of b is created as $b = \bar{\beta} + L\eta$, where L is the lower-triangular Choleski factor of Ω/N and η is K-dimensional vector of independent draws from a standard normal density. A draw of the vector b requires only K draws from a random number generator, K means over N terms each, and a few arithmetic calculations. It takes a tiny fraction of a second.

Given b and β, the conditional posterior of Ω is $IW(\Omega \mid K+N, KI+N\bar{V})$, where $\bar{V} = (1/N)\sum(\beta_n - b)(\beta_n - b)'$. Draws from the inverted Wishart are easily obtained. Take $K + N$ draws of K-dimensional vectors of iid standard normal deviates. Calculate the Choleski factor, M, of $(KI + N\bar{V})^{-1}$. Create $S = \sum_r (M\eta_r)(M\eta_r)'$. Then $\tilde{\Omega} = S^{-1}$ is a draw. This calculation is also extremely fast.

The only computationally intensive part is drawing $\beta_n \forall n$. Given b and Ω, the conditional posterior for β_n is proportional to $L(y_n \mid \beta_n) g(\beta_n \mid b, \Omega)$. The Metropolis-Hasting (M-H) algorithm is used to take draws from this distribution. (See Chib and Greenberg, 1995, for a general explanation of the M-H algorithm.) The previous draw is labeled β_n^0 and the new one is β_n^1. The new draw is obtained as follows.

[10] More precisely, the posterior on b approaches $N(\bar{\beta}, \Omega/N)$ as the variance of the prior on b rises without bound. This variance is specified to be sufficiently high such that the posterior is numerically indistinguishable from $N(\bar{\beta}, \Omega/N)$.

1. Calculate $d = \sigma L\eta$, where η is a draw of a K-dimensional vector of iid standard normal deviates, L is the Choleski factor of Ω, and σ is a scalar that the researcher sets in a way to be described below.

2. Create a "trial" value of β_n^1 as $\tilde{\beta}_n^1 = \beta_n^0 + d$.

3. Evaluate the posterior at this trial value and compare it with the posterior at the previous draw. That is, calculate the ratio

$$R = \frac{L(y_n \mid \tilde{\beta}_n^1) g(\tilde{\beta}_n^1 \mid b, \Omega)}{L(y_n \mid \beta_n^0) g(\beta_n^0 \mid b, \Omega)}.$$

4. Take a draw from a standard uniform and label the draw μ.

5. If $\mu < R$, accept the trial draw. Otherwise, reject the trial draw and use the previous draw as the current draw. That is, set $\beta_n^1 = \tilde{\beta}_n^1$ if $\mu < R$ and set $\beta_n^1 = \beta_n^0$ otherwise.

A sequence of draws taken by the M-H algorithm converges to draws from the target distribution, in this case the conditional posterior. One draw of β_n within the M-H algorithm for each person is taken in each iteration of the Gibbs sampling over b, Ω, and $\beta_n \forall n$. Movement to convergence in the M-H algorithm for each person and in the overall Gibbs sampling is thereby attained simultaneously. In our application we used 30,000 iterations for "burn-in" (i.e., movement toward convergence) followed by 20,000 iterations, of which the draws in every 10-th iteration were retained. (Run-times were only 1.5 hours, even with this large number of iterations.) The 2,000 retained draws are used to conduct inference. For example, the average of these draws is the simulated mean of the posterior, which, from a classical perspective, is the estimate of the parameters. The standard deviation of the draws is the simulated standard deviation of the posterior and the classicists' standard error of the estimate.

The value of σ in step (1) affects the acceptance rate in the M-H algorithm. For smaller values of σ, the acceptance rate is generally higher but the jumps between draws is smaller so that more draws are needed for the algorithm to reach convergence and, once at convergence, to traverse the conditional posterior. Gelman *et al.* (1995) found that the optimal acceptance rate is .4 for $K = 1$ and decreases to .23 for higher dimensions. They recommend an adaptive acceptance rate to achieve optimality. This adaptation is implemented by changing σ in each iteration of the Gibbs sampling based on the acceptance rate among the N trial draws of $\beta_n \forall n$ in the previous iteration. Following Sawtooth Software (1999) and in accordance with the optimal rate found by Gelman *et al.*, we lower σ if the acceptance rate is below .3 and raise it if the rate is above .3.

3. Transformation of normals

Denote the partworths of person n as c_n, which is a vector with the same length as β_n. The partworths are defined by $c_n = T(\beta_n)$, where T is a

transformation that depends only on β_n and is weakly monotonic (such that $\partial c_n^k / \partial \beta_n^k \geq 0$ for each element k of c_n and β_n). The distribution of c_n is determined by the transformation.

Little is changed in the estimation procedure by this transformation. Normally distributed β_n's are drawn as before but then transformed to c_n's when they enter utility. Utility is specified as

$$U_{njt} = T(\beta_n)' x_{njt} + \varepsilon_{njt}.$$

The probability of the person's choice sequence given β_n is

$$L(y_n \mid \beta_n) = \prod_t \frac{e^{T(\beta_n)' x_{ny_{nt}t}}}{\sum_j e^{T(\beta_n)' x_{njt}}}.$$

This probability is used in step 3 of the M-H algorithm instead of the probability based on untransformed β_n. The rest of the procedure is same.

In this set-up, β_n can be considered a latent value that determines the person's partworths. This latent value is normally distributed, with mean b and covariance Ω. The conditional posteriors for b and Ω are the same as before, and the conditional posterior of β_n changes only by the transformation that occurs in utility in the logit formula. The advantages of normal distributions within a Bayesian procedure are maintained while allowing the partworths to take other distributions. For any given value of β_n, the partworths c_n are calculated, and the distribution of β_n induces a distribution of c_n.

4. Application

We present an analysis of customers' choice among gas, electric, and hybrid gas-electric vehicles. We apply the methods described above to investigate the use of various bounded distributions of partworths, all of which involve transformations of normals.

Stated choice experiments were designed to elicit customers' choice among gas, electric, and hybrid vehicles under various prices, operating costs, and other attributes. The experiments were conducted as part of a survey of vehicle owners in California. The state of California is particularly relevant for electric and hybrid vehicles because the state's Air Resources Board has implemented, and is continually revising, regulations that promote these vehicles. Survey respondents were contacted through random-digit dialing throughout the state. Respondents intending to purchase a new vehicle within the next three years were asked to participate in the study. Those who were willing to participate in the study were sent a packet of materials, including information sheets that described the new vehicles and the choice experiments. The respondents were later called to go over the information, obtain their choices in the experiments, and ask demographic and other questions. A total of 500 respondents were obtained.

4.1 Choice experiments

Each choice experiment consisted of three vehicles. For each vehicle, the following information was listed:

- Body type, such as midsize car
- Engine type: gasoline, electric, or hybrid
- Purchase price
- Operating cost in dollars per month
- Performance
- Range: miles between refueling/recharging.

The respondent was asked to consider the attributes of all three vehicles and state which one he/she would buy if making the purchase today. Each respondent was presented with 15 choice experiments, with each experiment containing different vehicles with different attributes.

The choice experiments were designed to provide as wide variation in each attribute, and as little covariance among attributes, as possible while maintaining plausibility. Ten body types were considered in the experiments: mini car, small car, midsize car, large car, small SUV, midsize SUV, large SUV, compact pick-up, large pick-up, mini-van. Respondents were given examples of vehicles with that body type.

Each vehicle in each experiment was listed as being gasoline, electric, or hybrid. In any one experiment, the respondent might face a choice among two electric and one hybrid vehicle, among three gasoline vehicles, or any other combination.

The purchase price and operating cost of each vehicle were chosen randomly from a range of possible prices and operating costs.

The performance of each vehicle was described in terms of top speed and the number of seconds required to go from zero to 60 mph. These two performance measures were not varied independently, since respondents know that they are linked. Rather, three performance levels were specified, and each vehicle was randomly assigned one of the three performance levels. The three levels were: (1) Top speed of 120 mph, and 8 seconds to reach 60 mph. This level is called "high" performance in the discussions below; however, the respondent did not see the word "high". The respondent saw the numbers for top speed and seconds to 60. (2) Top speed of 100 mph, and 12 seconds to reach 60 mph. This level is called "mid" in the discussions below. (3) Top speed of 80 mph, and 16 seconds to reach 60 mph. This level is called "low." The performance for gas and hybrid vehicles was randomly chosen from all three

levels. The performance for electric vehicles was randomly chosen from the mid and low levels.

For the miles between refueling/recharging, a range of miles was given for each vehicle. The miles between refueling was given as "300-500" miles for gas vehicles and "400-700" miles for hybrid vehicles in all the experiments. A constant level was used for these vehicles because the study did not intend to estimate the value of increasing the range of vehicles that are refueled conventionally. The goal was to estimate the value to customers of increasing the range of electric vehicles. The range for gas and hybrid vehicles was stated so that the experiments would not place undue emphasis on the electric vehicle range in the eyes of the respondent. (If the range of electric vehicles was stated in the experiment but not the range of gas or hybrid vehicles, then respondents might be induced to place more importance on this aspect of electric vehicles than they otherwise would.) For electric vehicles, the possible ranges included every 10 mile increment starting with 60-70 and going up to 190-200. The range for each electric vehicles in the choice experiments was chosen randomly from these levels.

4.2 Models

Price, operating cost, and range are linearized, such that their partworths represent the value of a one-unit increment. The negative of price and operating cost are entered, such that their partworths are expected to be positive (so that log-normal distributions, which have positive support, can be used.) For performance, the low level is taken as the base and the medium and high levels are represented in increments. That is, two variables are entered for performance: a dummy indicating that the vehicle has either medium or high performance, and a dummy indicating that the vehicle has high performance. For engine types, gas is taken as the base, such that the partworths of the electric and hybrid vehicles are the value of these engine types relative to that of a gas engine. Similarly, the large car is taken as the base body type, with the partworths for the others representing value relative to the large car.

We start with a model in which all the partworths are distributed jointly normal $N(b, \Omega)$. As stated above, 2000 draws of b, Ω and β_n $\forall n$ are obtained from their posterior distribution. The means of the 2000 draws of b and of the diagonal elements of Ω are given in Table 7.1. (The partworths for body types are omitted from this and subsequent tables to save space and because they contain relatively less interpretable content.) From a classical perspective, these figures represent the estimated mean and variance of the β_n's in the population. And since the β_n's are untransformed, the figures also represent the mean and variance of partworths in the population. The standard deviations of the draws of b and the diagonal elements of Ω are given in parentheses. From

a classical perspective, these are the standard errors of the estimated mean and variance of β_n's in the population.

Table 7.1. Model of vehicle choice with all normal distributions.

β_n*'s and partworths for:*	*Mean*	*Variance*	*Share>0*
Price (negative):	.1900	.0632	.78
	(.0127)	(.0048)	
Operating cost (negative):	.0716	.0467	.63
	(.0127)	(.0032)	
Range:	1.213	4.050	.73
	(.2442)	(.7190)	
Electric vehicle:	–3.554	16.95	.19
	(.4120)	(3.096)	
Hybrid vehicle:	1.498	6.483	.72
	(.1584)	(.9729)	
High performance:	.3092	1.425	.60
	(.1004)	(.2545)	
Mid and high performance:	.8056	1.298	.76
	(.1030)	(.2384)	
Log-likehood	–6,835.5		

For example, for our first model, the partworth associated with range is normally distributed in the population with an estimated mean of 1.213 and estimated variance of 4.050. These estimates imply that 73 percent of the population have positive partworth for range while the other 27 percent have a negative partworth. (These negative partworths for range are of course implausible and the basis for our exploration of other distributions below.) The standard error on the estimated mean is 0.2442, which gives a t-statistic of 4.97, implying that the mean is significantly different from zero. Similarly, the standard error on the estimated variance is 0.7190, for a t-statistic of 5.63, implying that the variance is also significant (that is, the hypothesis of no variance can be rejected.) The classical log-likelihood of the model is given at the bottom of the Table. This value is the log-likelihood of the observed choices calculated at the estimated values of b and Ω (that is, at the mean of the draws of b and Ω.) We give this statistic, which is not used in Bayesian inference but is standard in classical analysis, to emphasize the ability of the Bayesian procedures to provide results that are interpretable classically.

The mean partworth associated with moving from low to medium or high performance (0.8056) is greater than that for moving from medium to high performance (0.3092), which is consistent with decreasing marginal utility of performance. The estimates for the other partworths are self-explanatory.

The model is implausible in several regards. The estimates imply that 22 percent of the population prefer higher prices. The existence of price coefficients with the wrong sign renders the model un-useable for calculation of willingness to pay and other welfare measures. The estimates also imply that 37 percent of people prefer higher operating costs, 27 percent prefer electric vehicles with shorter ranges over those that can be driven further between recharging, 24 percent prefer low performance over medium or high performance, and 40 percent prefer medium performance over high performance. Also, for any arbitrarily large value in either direction, the model implies that some people have a partworth in excess of this value. The model therefore implies that some people would buy a vehicle that is worse in all features than any other currently-offered vehicle. These implications are the result of using normal distributions for partworths when actual partworths have known signs and limited magnitudes.

Table 7.2 gives the correlation among the partworths implied by the estimate of Ω. The largest correlation (in magnitude) is between the partworths for range and electric vehicle: the correlation of –0.64 implies that people who are concerned about the range of an electric vehicle tend not to like electric vehicles at any range. This result is questionable. It is probably true that people who are concerned about range tend not to like electric vehicles, since electric vehicles generally have short ranges. However, the range of the electric vehicle is explicitly included in the experiments and the model. The negative correlation in partworths therefore implies that people who care about range tend to not like electric vehicles for reasons beyond range. It is not clear what these reasons might be or why they relate to concern about range.

Table 7.2. Correlations among partworths with all normal distributions.

Price	1.00	0.11	–0.10	0.05	–0.18	–0.07	–0.01
Operating cost		1.00	–0.05	0.15	0.01	0.01	–0.01
Range			1.00	–0.64	0.36	–0.01	0.15
Electric vehicle				1.00	0.12	0.02	–0.19
Hybrid vehicle					1.00	0.19	0.06
High performance						1.00	0.17
Med and high performance							1.00

We estimate two models under other distributional assumptions that are easy to accommodate within our estimation procedure. For the first of these two models, the two performance variables are given normal distributions that are censored from below at zero. With this distribution, a share of the population is completely unconcerned about performance (i.e., have a partworth of zero)

Table 7.3. Model of vehicle choice with transformations of normals.

	β_n		*Partworths*	
	Mean	*Variance*	*Mean*	*Variance*
Price (negative):	–2.531	0.9012	0.1204	0.0170
	(.0614)	(.1045)		
Operating cost (negative):	–3.572	1.015	0.0455	0.0031
	(.1100)	(.1600)		
Range:	–1.222	1.370	0.5658	0.8965
	(.2761)	(.3368)		
Electric vehicle:	–1.940	2.651	–1.9006	2.6735
	(.1916)	(.4965)		
Hybrid vehicle:	0.9994	2.870	1.0003	2.8803
	(.1267)	(.4174)		
High performance:	–.7400	2.358	0.3111	0.3877
	(.2953)	(.7324)		
Mid and high performance:	–.0263	1.859	0.5089	0.5849
	(.1538)	(.3781)		
Log-likehood	–6,171.5			

while the other share of the population places a positive value of higher performance with this value varying over people. The coefficients of price, operating cost, and range are given log-normal distributions (with the negative of price and operating cost entering the model.)

The mean and standard deviation of b and the diagonal elements of Ω are given in Table 7.3. Note that the log-likelihood of this model is considerably higher than that for the model with all normals: –6,171.5 compared to –6,835.5. As stated above, b and Ω are the mean and variance of the β_n in the population, which are transformed to obtain the partworths. The distribution of partworths is obtained through simulation on the estimated values of b and Ω.

In particular, draws of β_n are taken from a normal distribution with mean equal to the estimated value of b and variance equal to the estimated value of Ω. Each draw of β_n is then transformed to obtain a draw of partworths.[11] The mean and variance of these partworths are given in the latter columns of Table 7.3. The specification of the distributions assures that no one in the population dislikes (i.e., has a strictly negative partworth for) price reductions, oper-

[11] An alternative procedure, which is more consistent with Bayesian concepts and less consistent with classical concepts, is to retain the draw of the partworths (i.e., the transformation of the draw of β_n) for each person in each iteration of the MCMC procedure after convergence and calculate statistics such as means and variances over these draws.

ating cost reductions, range improvements, or either of the two performance improvements.[12] The mean partworths are smaller in this model than in the model with all normal distribution.[13] This difference is evidence of how the use of normal distribution can distort the estimated mean partworths. In particular: For a desirable attribute, the normal distribution gives an implausible negative sign for some share of customers; in estimation, the distribution is moved "up" to avoid the poor fit that these implausible values imply. With distributions that do not contain implausible values, the estimation procedure is not distorted to avoid implausible values.

The estimates imply that 51 percent of the population do not care about an improvement from low to mid-level performance, and 69 percent of the population do not care about an improvement from mid-level to high performance. These shares seem larger than expected (at least what we expected.) However, this result might simply indicate that the other attributes that were included in the choice experiments are more important to a large share of respondents, such that the partworth for performance appears to be zero for these respondents. If attributes that were considered less important than performance had been included in the experiments, with variation only over performance and these less important attributes, then a positive partworth for performance might have been evidenced.

Table 7.4. Correlations among partworths with transformations of normals.

Price	1.00	0.25	0.14	0.00	0.35	0.12	0.05
Operating cost		1.00	0.08	–0.10	0.17	0.02	–0.04
Range			1.00	–0.05	0.27	0.03	0.02
Electric vehicle				1.00	0.38	0.04	–0.11
Hybrid vehicle					1.00	0.22	0.09
High performance						1.00	0.14
Med and high performance							1.00

Table 7.4 gives the correlation among partworths implied by the estimated model. The implications are generally more reasonable that for the model with all normal distributions. People who are concerned about price are also concerned about operating cost. People who like electric vehicles also tend

[12]The mean β_n is negative for many of these attributes, even though the partworths themselves are positive. For log-normal distributions, β_n is exponentiated such that the partworth is positive even if β_n is negative. In this case, a negative mean for β_n implies that the median partworth is between zero and one. Similarly, if the partworth is a normal censored at zero, a negative mean for β_n implies than more than half of the population does not care about the attribute.

[13]Even though the means drop, the ratios of means move in both directions relative to those with all normal distributions.

to like hybrid vehicles. This result suggests that customers have a willingness or unwillingness to consider new technologies that transcends the particular technology. The questionable correlation between the partworths of range and electric vehicles that arose in the model with all normal distributions is not evidenced in the model with transformations of normals.

As specified, the partworths for price, operating cost and range have log-normal distributions, which allow unboundedly large partworths and have fairly thick tails. It might be more appropriate to give these partworths a S_B distribution. To investigate this question, and to illustrate how various distributions can be tested, we estimated a model that is the same as the one just described except that the partworths for these three variables are specified as S_B with a lower bound of zero and a high upper bound for each. The upper bounds are 1.0 for price and operating cost, and 2.0 for range, which are high enough to accommodate nearly all of the cumulative distribution under the respective log-normals and yet allow a different shape of the distribution within the relevant range. The log-likelihood for this model is higher than for the previous one: –6,159.7 compared to –6,171.5. For price and operating cost, the mean and variance of the partworths are about the same with the S_B distribution as the log-normal; however, the shape differed, with the tail of the S_B distribution being considerably smaller even within the support of the S_B. Figure 7.1 illustrates the difference for the price coefficient, with the solid line representing the S_B distribution and the dashed line representing the log-normal.

For the partworths associated with range, using the S_B distribution instead of the log-normal had a substantial effect. Figure 7.2 shows the estimated densities under the S_B distribution (solid line) and log-normal (dashed line). The S_B distribution provides a plateau shape that cannot be accommodated with a log-normal. The question arises of whether this shape is the result of placing a maximum of 2.0 on the S_B distribution when the estimated log-normal distribution gives a non-negligible share above 2. We re-estimated the model with the maximum of the range coefficient set at 5.0 instead of 2.0. The estimated S_B density with this maximum takes a shape that is similar to the log-normal. However, the log-likelihood dropped considerably, from –6,159.7 with the maximum set at 2.0 to –6,163.1 with the maximum set at 5.0. Apparently the improvement in fit that arises from using S_B distribution instead of the log-normal is due to the plateau shape that the S_B distribution takes when its maximum is set at 2.0 for this partworth.

The bounds of the S_B distribution can be estimated as parameters, rather than specified by the researcher. Doing so requires an extra layer of Gibbs sampling, with the bounds drawn from their posterior distribution conditional on $\beta_n \ \forall n$. The conditional posterior is proportional to the logit likelihood for the entire sample, $\prod_n L(y_n \mid \beta_n)$, times the prior on the bounds, where the

Figure 7.1. Price coefficients under the S_B distribution (solid line) and log-normal (dashed line).

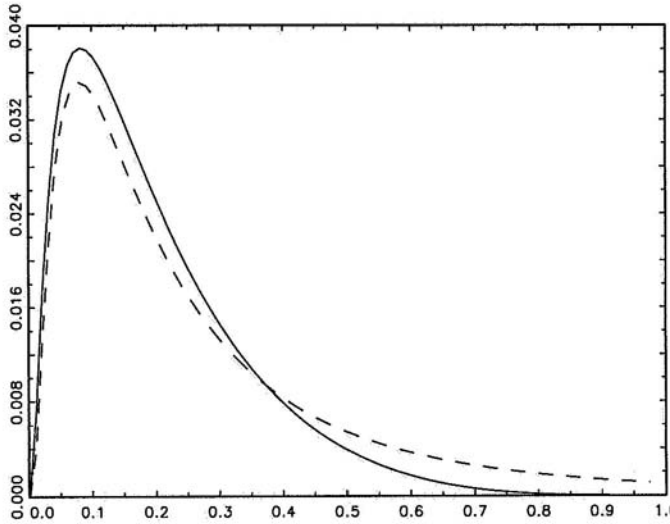
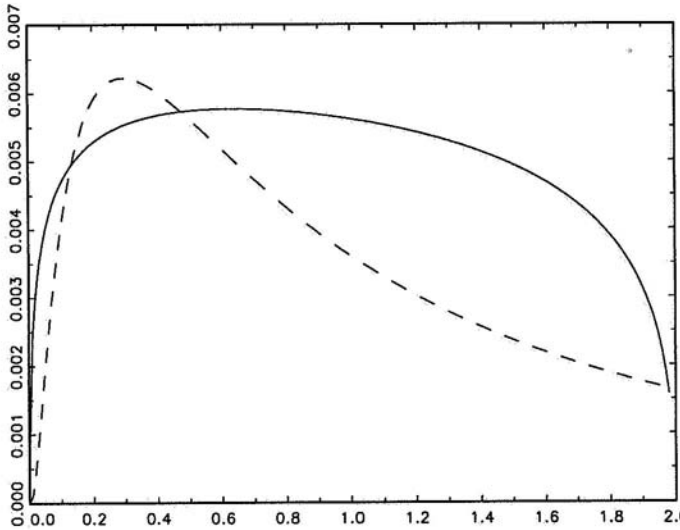


Figure 7.2. Range coefficients under the S_B distribution (solid line) and log-normal (dashed line).



utility that is used to calculate the logit formulas in $L(\cdot)$ depends on the bounds of the S_B distributions. A M-H algorithm is used to take draws from this conditional posterior, similar to that used by Train (2001) for fixed coefficients.

We estimated a model with the upper bound of the S_B distribution for the range coefficient treated as a parameter. Using a flat prior, the estimated value was 2.86 with a standard error of 0.42. The log-likelihood of the model dropped slightly from –6,159.7 with the upper bound set at 2.0 to –6,160.56 with the estimated bound. Run time approximately doubled, since the M-H algorithm for the bounds of the S_B distribution requires about the same amount of calculation as the M-H algorithm for $\beta_n \ \forall n$. As noted above, run times are fairly short with the procedure such that doubling them is not a burden. However, identification becomes an issue when the bounds are treated as parameters, since the difference between the upper and lower bounds, $u - \ell$, is closely related to the variance ω of the latent normal term. An important area for further work is whether the S_B distributions can be re-parameterized in a way that improves identification of each parameter when the researcher does not specify the bounds.

Chapter 8

KUHN-TUCKER DEMAND SYSTEM APPROACHES TO NON-MARKET VALUATION

Roger H. von Haefen
Department of Agricultural and Resource Economics,
University of Arizona, AZ, U.S.A.
and
Department of Economics and Center for Environmental Science and Policy,
Stanford University, CA, U.S.A.
rogervh@ag.arizona.edu

Daniel J. Phaneuf
Department of Agricultural and Resource Economics,
North Carolina State University, NC, U.S.A.
dan_phaneuf@ncsu.edu

Abstract In this chapter we summarize recent advances with Kuhn-Tucker demand system approaches to non-market valuation. Over the past five years, simulation-based estimation and welfare calculation strategies have been developed that enable the Kuhn-Tucker framework to address policy-relevant valuation questions in applications with many quality-differentiated goods. We review these modeling innovations in the context of three generic Kuhn-Tucker specifications that differ in terms of their ability to account for unobserved preference heterogeneity. For illustration, we apply the alternative specifications to Canadian moose hunting data and present parameter and welfare estimates. We conclude the chapter by suggesting important areas for future research within the Kuhn-Tucker framework.

Keywords: Kuhn-Tucker models, applied demand analysis, simulation-based estimation, welfare measurement.

1. Introduction

Policy recommendations arising from economic research often depend critically on the public's willingness to pay or willingness to accept for price and quality changes. To construct these welfare measures, analysts must often estimate coherent demand systems for the affected commodities. In microeconometric applications with numerous quality-differentiated goods, both demand system estimation and welfare calculation can be daunting in practice.

Environmental economists routinely confront these challenges when using outdoor recreation data to infer the non-market value of changes in site attributes and availability. In this chapter we provide an overview and synthesis of existing continuous demand system approaches to addressing these issues. Our discussion is couched in the context of recreation demand analysis and focused on so-called "Kuhn-Tucker" approaches, or continuous demand system models estimated within the primal framework.[1] Significant progress has been made with these models in recent years, and we emphasize throughout the critical role that simulation has played. We illustrate these advances with recreation data on Canadian moose hunting.

Frequently in recreation demand applications, the analyst has seasonal trip data for a sample of individuals and a relatively large number of recreational sites (10 or more), socio-demographic information, and site attributes. For a given individual, many sites are unvisited but some are visited more than once. To consistently derive welfare measures for price and attribute changes with such data, structural econometric models that behaviorally and statistically account for the mixture of corner solutions (unvisited sites) as well as interior solutions (sites with one or more trips) are required. Over the past three decades, recreation demand modelers have developed an impressive array of econometric models to fit data with these characteristics (See Phaneuf and Smith (forthcoming) for a recent review), but since their introduction by Hanemann (1978), variations of the discrete choice random utility maximization (RUM) model have dominated. In discrete choice RUM models, the recreation season is decomposed into separable choice occasions with independent discrete trip choices made on each. These models are tractable in estimation, can account for a potentially large number of sites, and provide a consistent theoretical framework from which choice occasion welfare measures can be calculated. However, by focusing on the decision process at the choice occasion, discrete choice models are ill suited for estimating seasonal demands

[1]Continuous demand system models estimated within the dual framework have also been proposed by Lee and Pitt (1986). Although Phaneuf (1999) and Wang (2003) have recently used the framework to estimate preferences for 5 and 15 sites in the recreation demand context, respectively, considerably less progress has been made with the dual approach relative to the primal framework in terms of estimation and welfare calculation. Our discussion in this chapter therefore focuses on the primal framework.

and welfare measures. Although numerous approaches for linking decisions made on individual choice occasions to decisions made over the season have been proposed (see Parsons *et al.* (1999) for an overview and comparison), no consensus on how this can be consistently done has emerged.

At the time he proposed the discrete choice framework for modeling seasonal recreation choices, Hanemann also developed a continuous demand system approach that was based on the primal representation of the consumer's constrained optimization problem and exploited the Kuhn-Tucker conditions in estimation. These "Kuhn-Tucker" models (hereafter KT models), as Wales and Woodland (1983) called them in separate but concomitant work, consistently account for both the extensive (which sites to visit) and intensive (how many trips to take) margins of choice and can be used to recover a coherent representation of an individual's seasonal preferences. As such, the KT framework has a significant conceptual advantage over discrete choice approaches for modeling seasonal recreation demand.

Despite its conceptual appeal, no recreation demand application exploited the KT framework prior to 2000. Moreover, only a handful of empirical studies have employed the KT framework outside of environmental economics, and all of these studies used data sets with less than 10 goods and did not report welfare measures. This historical avoidance of the KT framework in applied demand analysis can be explained by the absence of computational algorithms for flexibly estimating the structural parameters of and constructing Hicksian welfare measures from KT models in realistic, policy-relevant applications. Over the past 5 years, however, advances in simulation have facilitated the development of the necessary algorithms and in turn the wider application of the KT framework.

We begin the chapter by describing the generic KT framework as well as the estimation and welfare calculation challenges that have historically limited its application. We then consider three KT specifications that fit within a common additively separable preference structure but vary in terms of their ability to flexibly account for unobserved heterogeneity. We use these specifications as vehicles for discussing recent computational advances in KT estimation and welfare calculation. Next, we illustrate these advances with an application to moose hunting in Canada. We close the chapter with an appraisal of current KT approaches to non-market valuation and recommendations for future research.

2. Modeling framework and challenges

The point of departure for the KT framework is the consumer's direct utility function $U(\mathbf{x}, \mathbf{Q}, z, \beta, \mathbf{E})$ where $\mathbf{x}$ is an M-dimensional vector of visits to the available recreation sites, $\mathbf{Q} = [\mathbf{q}_1, \cdots, \mathbf{q}_M]$ denotes an $L \times M$ matrix of site-specific quality attributes for the recreation sites, z is a strictly positive

numeraire representing spending on other goods, β is a vector of structural parameters entering preferences, and $\mathbf{E}$ is a vector or matrix of unobserved heterogeneity. Because $\mathbf{E}$ is interpreted as the components of the utility function that are known to the individual but unobserved by the analyst, the setup is consistent with McFadden's random utility maximization hypothesis, and the elements of $\mathbf{E}$ have a structural interpretation. The consumer chooses $\mathbf{x}$ and z to maximize utility subject to budget and non-negativity constraints:

$$\max_{\mathbf{x},z} U(\mathbf{x}, \mathbf{Q}, z, \beta, \mathbf{E}) \quad s.t. \quad y = \mathbf{p}^\top \mathbf{x} + z, \quad x_j \geq 0, j = 1, \cdots, M, \tag{2.1}$$

where $\mathbf{p}$ is an M-dimensional vector of travel costs for the recreation sites and y is income. Assuming trips are infinitely divisible (i.e., ignoring the non-negative integer nature of trip demand),[2] (2.1) represents a continuous nonlinear programming problem that can be solved using the approach of Kuhn and Tucker (1951). Assuming U satisfies the standard regularity conditions, the Kuhn-Tucker conditions that implicitly define the optimal consumption bundle $(\mathbf{x}^*, z^*)$ are

$$\frac{\partial U/\partial x_j}{\partial U/\partial z} \leq p_j, j = 1, \ldots, M, \tag{2.2}$$

$$x_j \times \left[\frac{\partial U/\partial x_j}{\partial U/\partial z} - p_j \right] = 0, j = 1, \ldots, M. \tag{2.3}$$

The left hand side of (2.2) can be interpreted as the Marshallian "virtual" price of x_j, the price that would make the individual freely choose x_j^* if the non-negativity constraints were somehow relaxed. Each virtual price is an endogenous function of $(\mathbf{x}^*, z^*)$, and as suggested by equation (2.2), its relationship relative to its corresponding travel cost provides the conceptual link between observable demands and the structure of preferences. For a visited site (i.e., an interior solution), the site's virtual and travel cost prices are equal, and trip demand is chosen such that the marginal rate of substitution between trips and the numeraire equal the travel cost. For an unvisited site (a corner solution), the site's virtual price is bounded from above by its travel cost and can be interpreted as the individual's reservation price, or the price below which trip demand will be strictly greater than zero.

Combining the M weak inequalities in (2.2) with distributional assumptions for $\mathbf{E}$ implies likelihoods for alternative values of $(\mathbf{x}^*)$ conditional on β. As Hanemann (1978) and Wales and Woodland (1983) argued, estimation of β can proceed by maximum likelihood, but to do so in general requires evaluation of complex multivariate integrals whose dimensions equal the number

[2] See von Haefen and Phaneuf (2003) for an empirical investigation of this assumption.

of unvisited sites. Hanemann showed how strategic use of convenient distributional assumptions can imply closed form likelihood functions, but exploiting such assumptions comes at the high cost of behavioral implausibility. Thus prior to the development of modern simulation techniques, the computational difficulties associated with evaluating high dimensional integrals effectively limited the application of KT models to data sets with four or fewer goods.

Welfare measurement in the KT framework raises additional challenges. The Hicksian compensating surplus CS^H arising from a change in prices and/or qualities from ($\mathbf{p}^0$, $\mathbf{Q}^0$) to ($\mathbf{p}^1$, $\mathbf{Q}^1$) can be defined implicitly using indirect utility functions:

$$\nu(\mathbf{p}^0, \mathbf{Q}^0, y, \beta, \mathbf{E}) = \nu(\mathbf{p}^1, \mathbf{Q}^1, y - CS^H, \beta, \mathbf{E}), \tag{2.4}$$

or explicitly using expenditure functions:

$$CS^H = y - e(\mathbf{p}^1, \mathbf{Q}^1, U^0, \beta, \mathbf{E}), \tag{2.5}$$

where $U^0 = \nu(\mathbf{p}^0, \mathbf{Q}^0, y, \beta, \mathbf{E})$. Whether the analyst uses equation (2.4) or (2.5) to solve for CS^H, two computational issues must be addressed. First, both $\nu(\cdot)$ and $e(\cdot)$ can be interpreted as endogenous regime switching functions where regimes correspond to the 2^M combinations of interior and corner solutions for the M sites. The functions are endogenous because the rational individual chooses the regime that maximizes her utility or minimizes her expenditures. Particularly when M is large, solving for $\nu(\cdot)$ and $e(\cdot)$ (and thus CS^H) can be computationally difficult. Second, the unobserved heterogeneity entering preferences is by definition unknown to the analyst, and thus CS^H is a random variable from the analyst's perspective and cannot be ascertained precisely. The analyst can at best compute a measure of its central tendency such as its expectation, $E(CS^H)$, using Monte Carlo simulation techniques. Although conceptually straightforward, the use of Monte Carlo techniques to estimate $E(CS^H)$ can be computationally intensive when the costs of solving for CS^H conditional on simulated $\mathbf{E}$ values are high. Furthermore, this computational burden is more onerous if confidence intervals for $E(CS^H)$ are desired because the entire routine for estimating $E(CS^H)$ must be replicated for alternative parameter values.

In sum, the computational challenges associated with estimation and welfare calculation in the KT framework are significant. The lack of tractable computational procedures to address these challenges explains why no empirical KT application in environmental economics existed prior to 2000. Moreover, in the relatively few KT applications outside of environmental economics, the data used consisted of 10 or fewer goods and no attempt was made to construct consistent welfare measures. In recent years, however, modeling innovations by Phaneuf, Kling and Herriges (2000), von Haefen, Phaneuf, and Parsons

(2004), and von Haefen (2004a) have enabled analysts to estimate and construct welfare measures from KT models in applications with as many as 89 goods (von Haefen, 2004b). The next two sections summarize these innovations in the context of three representative KT specifications that we apply to moose hunting data in our empirical illustration.

3. Empirical specifications and implementation strategies

All three specifications we consider are variations of the following direct utility function:

$$\begin{array}{ll} u(\cdot) & = \sum_{j=1}^{M} \Psi_j \ln(\phi_j x_j + \theta) + \frac{1}{\rho} z^{\rho} \\ \ln \Psi_j & = \delta^{\top} \mathbf{s} + \mu \varepsilon_j \\ \ln \phi_j & = \gamma^{\top} \mathbf{q}_j \\ \rho & = 1 - \exp(\rho^*) \\ \ln \theta & = \theta^* \\ \ln \mu & = \mu^*, \end{array} \tag{3.1}$$

where $\mathbf{s}$ is a vector of individual characteristics, $(\delta, \gamma, \theta^*, \rho^*, \mu^*)$ are structural parameters that may vary randomly across individuals in the population to capture unobserved heterogeneity, and $\mathbf{E} = (\varepsilon_1, \ldots, \varepsilon_M)$ represents additional unobserved heterogeneity that varies randomly across individuals and sites. Preferences in (3.1) embody additive separability, and as such, rule out *a priori* inferior goods and any Hicksian complementarities among goods. For a cross section of outdoor recreators, additive separability implies that wealthier individuals will take more trips to more sites on average. Although plausible in some applications, these implications may be troubling in others. As we describe below and in the next section, the appeal of additive separability is largely computational, particularly when constructing welfare measures. In addition, preferences in (3.1) are consistent with weak complementarity ($\partial u / \partial \mathbf{q}_j = 0$ if $x_j = 0, \forall j$; see Herriges, Kling, and Phaneuf (2004) and von Haefen (2004a) for recent discussions) because $\mathbf{q}_j$ enters preferences only through the j^{th} good's pure repackaging parameter, ϕ_j. Finally, for (3.1) to be a coherent preference specification, the ρ parameter cannot be greater than one. For numerical simplicity, we strengthen this restriction to $\rho < 1$.

Maximizing (3.1) with respect to the linear budget and non-negativity constraints implies a set of first order conditions that, with some manipulation, can be written as

$$\varepsilon_j \leq \frac{1}{\mu} \left(-\delta^{\top} \mathbf{s} + \ln \frac{p_j}{\phi_j} + \ln(\phi_j x_j + \theta) + (\rho - 1) \ln(y - \mathbf{p}^{\top} \mathbf{x}) \right), \forall j. \tag{3.2}$$

We assume each element of $\mathbf{E}$ is an independent draw from the normalized type I extreme value distribution. Hanemann (1978) recognized that this convenient but restrictive distributional assumption implies a closed form solution for the likelihood of observing $\mathbf{x}$ conditional on $(\delta, \gamma, \theta^*, \rho^*, \mu^*)$:

$$l(\mathbf{x}|\delta, \gamma, \theta^*, \rho^*, \mu^*) = |\mathbf{J}| \prod_j \left[\exp\left(-g_j(\cdot)\right)/\mu)\right]^{1(x_j>0)} \times \exp\left[-\exp(-g_j(\cdot))\right], \tag{3.3}$$

where $g_j(\cdot)$ is the right hand side of (3.2), $|\mathbf{J}|$ is the determinant of the Jacobian of transformation, and $1(x_j > 0)$ is an indicator function equal to one if x_j is strictly positive and zero otherwise.

Our alternative KT specifications differ in their treatment of $(\delta, \gamma, \theta^*, \rho^*, \mu^*)$ as fixed or random across the population. For all three specifications, estimation of the model's structural parameters in principle can proceed within either classical or Bayesian frameworks (see Train (2003) for a recent discussion of the subtle but important differences between the two). However, classical estimation approaches are easier to implement for the two more restrictive specifications but more difficult for the more general third. Although classical and Bayesian paradigms imply very different interpretations for the estimated parameters, the Bernstein-von Mises theorem suggests that Bayesian posterior mean estimates, interpreted within the classical framework, are asymptotically equivalent to their classical maximum likelihood counterparts when the analyst has correctly specified the data generating process. Following Train (2003), we interpret this result as suggesting that both approaches should generate qualitatively similar inference, and thus the analyst's choice of which to use in practice can be driven by computational convenience. We therefore discuss estimation of the more restrictive specifications in the classical framework and estimation of the more general specification in the Bayesian framework.

3.1 Fixed and random parameter classical estimation approaches

Our first empirical KT specification follows Phaneuf, Kling, and Herriges (2000) and assumes that all elements of $(\delta, \gamma, \theta^*, \rho^*, \mu^*)$ are equal across the population. This fixed parameter KT specification can be criticized as implausibly restrictive because only the independent random variables capture unobserved heterogeneity in the model.[3] However, its main appeal is that simulation is not required in estimation because the likelihood simplifies to $l(\mathbf{x}) = l(\mathbf{x}|\delta, \gamma, \theta^*, \rho^*, \mu^*)$, where the right hand side is as in equation (3.3).

[3]Phaneuf, Kling, and Herriges also consider a fixed parameter GEV variant of this model that allows for some correlations in the unobserved determinants of choice.

This closed form analytical structure implies that recovering maximum likelihood parameter estimates using numerical gradient-based search methods is straightforward even when M is large.[4]

Our second specification generalizes the fixed parameter specification by allowing a subset of parameters in $(\delta, \gamma, \theta^*, \rho^*, \mu^*)$ to vary randomly across the population. As von Haefen, Phaneuf, and Parsons (2004) discuss, estimation of this random parameter specification in the classical framework requires simulation because the likelihood function no longer has a closed form and numerical integration techniques can only be used when the dimension of the random parameter space is sufficiently small. The computational burden in estimation is consequently far more substantial. However, by strategically holding elements of $(\delta, \gamma, \theta^*, \rho^*, \mu^*)$ fixed across the target population and exploiting the same simulation-based estimation approaches that have been widely used in random parameter logit applications, estimation of this random parameter specification is tractable in the classical framework for a relatively large number of goods.

Our choice of which elements in $(\delta, \gamma, \theta^*, \rho^*, \mu^*)$ are fixed and random is significantly influenced by computational considerations. To understand why, consider the general case where all elements of $(\delta, \gamma, \theta^*, \rho^*, \mu^*)$ are random. The likelihood function takes the general form:

$$\begin{aligned} l(\mathbf{x}) &= \int l(\mathbf{x}|\delta, \gamma, \theta^*, \rho^*, \mu^*) d\delta d\gamma d\theta^* d\rho^* d\mu^* \\ &= \int |\mathbf{J}| \prod_j \left[\exp\left(-g_j(\cdot)/\mu\right)/\mu\right]^{1(x_j>0)} d\delta d\gamma d\theta^* d\rho^* d\mu^*. \end{aligned} \tag{3.4}$$

Simulating $l(\mathbf{x})$ involves: 1) simulating values of $(\delta, \gamma, \theta^*, \rho^*, \mu^*)$; 2) calculating $l(\mathbf{x}) = l(\mathbf{x}|\delta, \gamma, \theta^*, \rho^*, \mu^*)$, conditional on the simulated values; and 3) iterating R times and averaging the simulated likelihoods. With all parameters random across the population, the computational difficulty of this task is substantial because the determinant of the Jacobian of transformation, $|\mathbf{J}|$, is a function of $(\gamma, \theta^*, \rho^*)$ and must be recalculated on each simulation for each observation. However, if the analyst assumes $(\gamma, \theta^*, \rho^*)$ are fixed parameters, (3.4) simplifies to

$$\begin{aligned} l(\mathbf{x}) &= \int |\mathbf{J}| \prod_j \left\{ \left[\exp\left(-g_j(\cdot)\right)/\mu\right]^{1(x_j>0)} \exp\left[-\exp(-g_j(\cdot))\right] \right\} d\delta d\mu^*, \\ &= |\mathbf{J}| \int \prod_j \left\{ \left[\exp\left(-g_j(\cdot)\right)/\mu\right]^{1(x_j>0)} \exp\left[-\exp(-g_j(\cdot))\right] \right\} d\delta d\mu^*. \end{aligned} \tag{3.5}$$

Because $|\mathbf{J}|$ can be pulled outside the integral in (3.5), it must only be calculated once per observation when simulating $l(\mathbf{x})$. This significant compu-

[4]Using analytical gradient-based search routines is more difficult because the Jacobian of transformation in general has a complex structure that cannot be easily differentiated. Although they are generally slower and less precise than analytical gradient-based search routines, our experience has been that numerical gradient-based search routines are sufficiently fast and precise for estimating KT models.

tational savings explains why we follow von Haefen, Phaneuf, and Parsons (2004) and assume $(\gamma, \theta^*, \rho^*)$ are fixed parameters.

To complete our random parameter specification, we make the following assumptions. Although not essential for computational reasons, we assume μ^* is fixed. We also assume the elements of δ are draws from the multivariate normal distribution with mean $\bar{\delta}$ and variance-covariance matrix Σ_δ. For simplicity, we restrict the off-diagonal elements of Σ_δ to zero, although von Haefen (2004b) has estimated similar KT models with all elements of Σ_δ unconstrained. Relative to our fixed parameter specification, this random parameter specification is in many ways more general because it allows for both heteroskedasticity and positive correlations in the unobserved determinants of choice across the KT equations.

Estimation proceeds by maximum simulated likelihood (Gourieroux and Monfort, 1996). For our model, the simulated likelihood function is

$$\tilde{l}(\mathbf{x}) = R^{-1} \sum_{r=1}^{R} l(\mathbf{x}|\delta^r), \tag{3.6}$$

where δ^r is the r^{th} $(r = 1, \ldots, R)$ random draw of δ. In our maximum simulated likelihood routine $\tilde{l}(\mathbf{x})$ replaces $l(\mathbf{x})$ and numerical gradient-based search methods are again used to find the parameter values that maximize the sample's log-likelihood. As summarized in Train (2003, p. 259), these values are consistent, efficient, asymptotically normal, and equivalent to maximum likelihood values if R rises faster than that square root of the sample size.

In other applications of maximum simulated likelihood, a common empirical finding has been that several random draws are necessary before parameter estimates are unaffected by the addition of more random draws. This finding has caused many researchers to use a large number of random draws when calculating $\tilde{l}(\mathbf{x})$. As a result, the computational time necessary for model convergence increases substantially. In the mixed logit literature, an increasingly popular strategy for circumventing this difficulty is to use quasi-random draws such as Halton sequences (Train, 2003, p. 224) in place of random draws when simulating the likelihood function. The main advantage of quasi-random draws is that they can more systematically and completely fill the parameters' support and in the process increase simulation precision. As a result, the likelihood can be calculated to an arbitrary degree precision with fewer simulations and less run time. Several studies have documented dramatic gains from the use of Halton draws relative to random draws. Bhat (2001), for example, found that 100 Halton draws provided more precision that 1000 random draws in the context of mixed logit models. In light of these findings, we use Halton draws to calculate (3.6) in our application.

3.2 Random parameter Bayesian estimation approaches

In our third and most general specification, we assume $\beta = (\delta, \gamma, \theta^*, \rho^*, \mu^*)$ are distributed multivariate normal with mean $\bar{\beta}$ and unrestricted variance-covariance matrix Σ_β. Estimating this random parameter specification in the classical framework is computationally intensive not only because the Jacobian of transformation must be recalculated on each simulation, but also because the dimension of the parameter space is large (if β is a K-dimensional vector, there are $K + K(K+1)/2$ unique parameters in $(\bar{\beta}, \Sigma_\beta)$). However, recent developments in Monte Carlo Markov Chain (MCMC) methods have made the model tractable in the Bayesian framework.

The Bayesian framework assumes that the analyst has initial beliefs about the unknown parameters $(\bar{\beta}, \Sigma_\beta)$ that can be summarized by a prior probability distribution function, $f(\bar{\beta}, \Sigma_\beta)$. When the analyst observes $\mathbf{x}$, she combines this choice information with the assumed data generating process to form the likelihood of $\mathbf{x}$ conditional on alternative values of $(\bar{\beta}, \Sigma_\beta)$, $l(\mathbf{x}|\bar{\beta}, \Sigma_\beta)$. The analyst then updates her prior beliefs about the distribution of $(\bar{\beta}, \Sigma_\beta)$ to form a posterior distribution for $(\bar{\beta}, \Sigma_\beta)$ conditional on the data, $f(\bar{\beta}, \Sigma_\beta|\mathbf{x})$. By Bayes' rule, $f(\bar{\beta}, \Sigma_\beta|\mathbf{x})$ is proportional to the product of the prior distribution and likelihood, i.e., $f(\bar{\beta}, \Sigma_\beta|\mathbf{x}) \propto f(\bar{\beta}, \Sigma_\beta) l(\mathbf{x}|\bar{\beta}, \Sigma_\beta)/C$ where C is a constant. In general, $f(\bar{\beta}, \Sigma_\beta|\mathbf{x})$ will not have an analytical solution, and thus deriving inference about the moments and other relevant properties of $(\bar{\beta}, \Sigma_\beta)$ conditional on the data is difficult. However, Bayesian econometricians have developed a number of procedures to simulate random samples from $f(\bar{\beta}, \Sigma_\beta|\mathbf{x})$ and in the process draw inference about the posterior distribution of $(\bar{\beta}, \Sigma_\beta)$.

We follow Kim, Allenby, and Rossi (2002) and von Haefen (2004a) by specifying diffuse priors and using a Gibbs sampler with an adaptive Metropolis-Hastings component to simulate from $f(\bar{\beta}, \Sigma_\beta|\mathbf{x})$. By decomposing the parameter space into disjoint sets and iteratively simulating each parameter set conditionally on the others, the Gibbs sampler generates simulations from the unconditional posterior distribution after a sufficiently long burn-in. The algorithm is virtually identical to the one developed by Allenby and Lenk (1994) and described in Train (2003) for random parameter logit estimation except for the KT likelihood in (3.3) replacing the logit probability throughout.

We assume the following diffuse priors for β and Σ_β :

$$\beta \sim N(\beta^{FP}, \eta I_k) \text{ and } \Sigma_\beta \sim IW(k, I_k), \tag{3.7}$$

where $N(\cdot)$ and $IW(\cdot)$ denote the multivariate normal and inverse Wishart distributions, respectively, β^{FP} are the fixed parameter maximum likelihood estimates, η is a scalar chosen such that $1/\eta$ approaches zero, k is the dimension of β, and I_k is a k-dimensional identity matrix. These priors, in combination with our assumed data generating process, imply the following conditional posterior

distributions for $\bar{\beta}$ and Σ_β as well as the individual specific β_t (where we have added subscript t to index the T individuals in the sample):

$$\begin{array}{l} f(\bar{\beta}|\beta_1,\cdots,\beta_T,\Sigma_\beta,\mathbf{x}_1,\cdots,\mathbf{x}_T) \propto N(\bar{B},\Sigma_\beta/T) \\ f(\Sigma_\beta|\beta_1,\cdots,\beta_T,\bar{\beta},\mathbf{x}_1,\cdots,\mathbf{x}_T) \propto IW\left[k+T,(kI_k+T\bar{S})/(k+T)\right] \\ f(\beta_t|\bar{\beta},\Sigma_\beta,\mathbf{x}_t) \propto l(\mathbf{x}_t|\beta_t)\times n(\beta_t|\bar{\beta},\Sigma_\beta) \quad \forall t, \end{array} \tag{3.8}$$

where $l(\mathbf{x}_t|\beta_t)$ is the conditional likelihood function from equation (3.3) for individual t, $n(\cdot)$ is the multivariate normal density function, T is the sample size, and

$$\bar{\beta} = T^{-1}\sum_t \beta_t$$
$$\bar{S} = T^{-1}\sum_t (\beta_t-\bar{\beta})^\top(\beta_t-\bar{\beta}).$$

Aside from the form of the conditional likelihood function, these distributions are identical to those described in Train (2003) for estimating the mixed logit model using Bayesian techniques.

The Gibbs sampling algorithm proceeds by iteratively drawing from the conditional distributions in (3.8), with each draw being made conditional on the remaining parameters' most recent draws. As Train (2003) describes, simulating from the multivariate normal and inverse Wishart distributions is relatively straightforward. However, simulating from β_t's posterior distribution is more complex and requires an adaptive Metropolis-Hastings algorithm (Chib and Greenberg, 1995). Thus iteration i of the Gibbs sampler involves the following steps:

- **Step 1:** Simulate $\bar{\beta}^i$ from $N(\bar{B}^{i-1},\Sigma_\beta^{i-1}/T)$. To initialize the algorithm set $\Sigma_\beta^0 = kI_k$ and $\beta_t^0 = \bar{B}^0 = \beta^{FP}, \forall t$.

- **Step 2:** Simulate Σ_β^i from $IW\left[k+T,(kI_k+T\bar{S}^i)/(k+T)\right]$, where $\bar{S}^i = T^{-1}\sum_t(\beta_t^{i-1}-\bar{\beta}^i)^\top(\beta_t^{i-1}-\bar{\beta}^i)$.

- **Step 3:** Simulate β_t^i for each observation using one iteration from the following Metropolis-Hastings algorithm:
  1) For each observation, simulate a vector $\tilde{\beta}_t^i$ from $N(\beta_t^{i-1}, r^{i-1}\Sigma_\beta^i)$, where r^{i-1} is a constant. To initialize the sequence, set $r^0 = 0.1$.
  2) For each observation, construct the following statistic:

$$\chi_t^i = \frac{l(\mathbf{x}_t|\tilde{\beta}_t^i)n(\tilde{\beta}_t^i|\bar{\beta}^i,\Sigma_\beta^i)}{l(\mathbf{x}_t|\beta_t^{i-1})n(\beta_t^{i-1}|\bar{\beta}^i,\Sigma_\beta^i)}.$$

If $\chi_t^i \geq U_t^i$ where U_t^i is a uniform random draw, accept the candidate random parameters, i.e., $\beta_t^i = \tilde{\beta}_t^i$. Otherwise, set $\beta_t^i = \beta_t^{i-1}$.
3) Gelman *et al.* (1995) argue that the Metropolis-Hastings algorithm for the normal distribution is most efficient if the acceptance rate of candidate parameters averages between 0.23 and 0.44. Therefore, set $r^i = (1.01)r^{i-1}$ if the sample's proportion of accepted candidate parameter values is less than 0.3. Otherwise, set $r^i = (0.99)r^{i-1}$.

- **Step 4:** Iterate.

After a sufficiently long burn-in, this algorithm generates random draws from the posterior distributions of β_t, $\bar{\beta}$, and Σ. In practice, the burn-in length necessary to achieve convergence (i.e., random draws from the posterior distributions) is difficult to ascertain. However, our experience has been that the Gibbs sampler algorithm is very fast even in large choice set applications, and thus the analyst can cheaply add burn-in iterations if convergence is in doubt. Finally, because the Gibbs sampler induces serial correlation in β_t, $\bar{\beta}$, and Σ_β , we only use each 10^{th} simulation after the burn-in to construct distributional summary statistics and, as we discuss below, welfare measures.

3.3 Welfare analysis

Our simulation-based approach to welfare measurement follows von Haefen, Phaneuf, and Parsons (2004) by incorporating the implications of observed choice. In particular, we simulate the unobserved heterogeneity such that our model perfectly predicts observed behavior at baseline conditions and use the model's structure to predict how individuals respond to price, quality, and income changes. As von Haefen (2003) discusses, this conditional approach to welfare measurement differs from the more traditional unconditional approach where the structural model is used to predict both behavior at baseline conditions and responses to price, quality, and income changes. The law of iterated expectations implies that sample means of conditional and unconditional welfare estimates should converge in expectation assuming the analyst has correctly specified the data generating process. Similar to the analyst's choice of classical or Bayesian estimation approaches, the choice of conditional or unconditional welfare measures can thus be governed by convenience. Because our past experience with the conditional approach to welfare measurement in the KT framework suggests that it is up to three times faster than the unconditional approach in achieving an arbitrary level of precision in the welfare estimates, we employ the approach with our empirical specifications here.[5]

[5]An additional rationale for using the conditional approach to welfare measures arises when one finds poor in-sample trip predictions (von Haefen and Phaneuf (2003)). Incorporating observed choice can mask these prediction problems and in our view result in more reliable policy inference.

There are two key steps to constructing the Hicksian consumer surplus (CS^H) on each iteration of our simulation-based algorithm. The first involves simulating the unobserved heterogeneity such that it is consistent with observed choice at baseline conditions, and the second involves solving for CS^H conditional on the simulated unobserved heterogeneity values. In the next two subsections, we discuss our strategies for resolving each of these issues in the context of our three empirical specifications.

3.4 Simulating the unobserved heterogeneity

To simulate the unobserved heterogeneity entering all three of our specifications, we need to draw from $f(\beta_t, \mathbf{E}_t|\mathbf{x}_t)$, the joint distribution for individual t's structural parameters $\beta_t = (\delta_t, \gamma_t, \theta_t^*, \rho_t^*, \mu_t^*)$ and independent type I extreme value draws $\mathbf{E}_t$ conditional on her observed trip demands. The following decomposition suggests a convenient strategy for accomplishing this task:

$$f(\beta_t, \mathbf{E}_t|\mathbf{x}_t) = f(\beta_t|\mathbf{x}_t) f(\mathbf{E}_t|\beta_t, \mathbf{x}_t). \tag{3.9}$$

Equation (3.9) suggests that simulating from $f(\beta_t, \mathbf{E}_t|\mathbf{x}_t)$ can be accomplished by first simulating from $f(\beta_t|\mathbf{x}_t)$ and then conditionally from $f(\mathbf{E}_t|\beta_t, \mathbf{x}_t)$. For the fixed parameter classical specification, no simulation from $f(\beta_t|\mathbf{x}_t)$ is required. Moreover for the random parameter Bayesian specification, the random draws from $f(\beta_t|\mathbf{x}_t)$ generated as part of the Gibbs sampler can be used in welfare analysis. For the random coefficient classical specification, however, simulating from $f(\beta_t|\mathbf{x}_t)$ requires an additional subroutine. Following von Haefen, Phaneuf, and Parsons (2004), we use an adaptive Metropolis-Hastings algorithm to simulate from $f(\beta_t|\mathbf{x}_t)$ for this specification. The algorithm is qualitatively similar to the one used in the Gibbs sampler except that only the elements in δ_t vary randomly across the population. The key steps of the algorithm are summarized below:

1 At iteration i, simulate for all T observations a candidate vector of unobserved heterogeneity $\tilde{\delta}_t^i$ from the normal distribution with location parameters δ_t^{i-1} and scale parameters $(r^{i-1}\Sigma_\delta)$ where r^{i-1} is a constant. To initialize the algorithm, set $\delta_t^0 = \bar{\delta}$ and $r^0 = 0.1$.

2 For all T observations, construct the statistic

$$\chi_t^i = \frac{l(\mathbf{x}_t|\tilde{\delta}_t^i) n(\tilde{\delta}_t^i|\bar{\delta}, \Sigma_\delta)}{l(\mathbf{x}_t|\delta_t^{i-1}) n(\delta_t^{i-1}|\bar{\delta}, \Sigma_\delta)} \tag{3.10}$$

where is defined in equation (3.3). If $\chi_t^i \geq U_t^i$ where U_t^i is a uniform random draw, accept the candidate random parameter (i.e., $\delta_t^{i-1} = \tilde{\delta}_t^i$). Otherwise set $\delta_t^i = \delta_t^{i-1}$.

3 If the sample's proportion of accepted candidate parameters is less than 0.3, set $r^i = (1.01)r^{i-1}$. Otherwise, set $r^i = (0.99)r^{i-1}$.

4 Iterate.

After a relatively short burn-in period, this algorithm generates random draws from $f(\beta_t|\mathbf{x}_t)$. To reduce the serial correlation in the random draws, we only use every 10^{th} simulation after the burn-in to construct welfare measures.

Once the parameters of the model have been simulated, drawing from $f(\mathbf{E}_t|\beta_t, \mathbf{x}_t)$ is relatively straightforward. For each element of $\mathbf{E}_t$, simulation depends critically on whether a site is visited. If, for example, site j is visited, the structure of the model, the simulated random parameters, and the observed trip demands imply that $\varepsilon_{tj} = g_{tj}(\cdot)$ where $g_{tj}(\cdot)$ equals the right hand side of equation (3.2). By contrast if site j is unvisited, ε_{tj} is strictly less than $g_{tj}(\cdot)$ and can be simulated from the truncated type I extreme value distributional assumption using the probability integral transformation, i.e.,

$$\varepsilon_{tj} = -\ln\left[-\ln(\exp(-\exp(-g_{tj}(\cdot)))U_{tj})\right] \tag{3.11}$$

where U_{tj} is a uniform random draw.

3.5 Calculating CS^H

Since Phaneuf, Kling, and Herriges (2000) proposed the first strategy to solve for CS^H conditional on the simulated unobserved heterogeneity, numerous refinements to their approach have been developed that reduce the computational burden of the task significantly. To place these refinements in context, we first describe the Phaneuf, Kling and Herriges approach. Their strategy exploits the implicit, indirect utility-based definition of CS^H in equation (2.4) and uses a numerical bisection routine to iteratively solve for the income compensation that equates utility before and after the price and quality changes. At each step of the numerical bisection routine the analyst must analytically solve the consumer's constrained optimization problem conditional on $(\mathbf{p}^1, \mathbf{Q}^1, \beta_t, \mathbf{E}_t)$ and an arbitrary income level to determine the individual's utility. Phaneuf, Kling and Herriges proposed doing this by calculating each of the 2^M regime-specific conditional indirect utility functions and determining which one generated the highest utility.

Although their approach generates consistent estimates of CS^H, several factors limit its practical appeal. Perhaps the most significant is that analytically calculating 2^M regime-specific conditional indirect utility functions for several observations and simulated values of $(\beta_t, \mathbf{E}_t)$ is not practical when M is large. Phaneuf and Herriges (1999) have shown that the approach is feasible for recreation applications with as many as 15 sites, but with the addition of only a few more sites, the approach becomes intractable. Moreover, the approach is limiting in the sense that the analyst must choose preference specifications with

closed form conditional indirect utility functions. Since the number of preference specifications that satisfy this condition is relatively small, this constraint is notable. In addition, the approach is inefficient because it is based on the implicit, indirect utility-based function definition of CS^H in (2.4) instead of its explicit, expenditure-based function definition in (2.5). As a result, the analyst must use a numerical bisection algorithm to solve for the income compensation that equates utility.

von Haefen, Phaneuf, and Parsons (2004) significantly refined Phaneuf, Kling, and Herriges' approach by introducing an iterative algorithm that numerically solves the consumer's constrained optimization problem at each iteration of the numerical bisection routine. By using a numerical approach based on the Kuhn-Tucker conditions to solve for the consumer's optimal consumption bundle and utility, von Haefen, Phaneuf, and Parsons' approach can be applied to preference specifications without closed form indirect utility functions (such as our three specifications in this chapter) as well as data with many goods. Like Phaneuf, Kling, and Herriges' approach, however, their approach exploits the implicit definition of CS^H in (2.4) and therefore is not fully efficient. Nonetheless, experience has shown it to be surprisingly fast in practice regardless of the number of goods.

von Haefen, Phaneuf, and Parsons' approach can be best appreciated by inspecting the generic Kuhn-Tucker conditions when preferences are additively separable (i.e., $u(\cdot) = \sum_k u_k(x_k) + u_z(z)$):

$$\frac{\partial u_j(x_j)}{\partial x_j} \leq \frac{\partial u_z(z)}{\partial z} p_j, \quad x_j \geq 0, \quad \forall j \tag{3.12}$$

and

$$z = y - \sum_{k=1}^{M} p_k x_k. \tag{3.13}$$

Notably in (3.12) only x_j and z enter the j^{th} Kuhn-Tucker equation. This simple structure suggests that if the analyst knew the optimal value for z, she could use equation (3.12) to solve for each x_j. Therefore under additive separability, solving the consumer's problem reduces to solving for the optimal value of z. Based on this insight the following numerical bisection algorithm can be used to solve the consumer's problem conditional on values for the exogenous variables and the simulated unobserved heterogeneity:

1 At iteration i, set $z_a^i = (z_l^{i-1} + z_u^{i-1})/2$. To initialize the algorithm, set $z_l^0 = 0$ and $z_u^0 = y$.

2 Conditional on z_a^i, solve for $\mathbf{x}_i$ using (3.12). Use (3.13) and $\mathbf{x}_i$ to construct $\tilde{z}^i$.

3 If $\tilde{z}^i > z_a^i$, set $z_l^i = z_a^i$ and $z_u^i = z_u^{i-1}$. Otherwise, set $z_l^i = z_l^{i-1}$ and $z_u^i = z_a^i$.

4 Iterate until $|(z_l^i - z_u^i)| \leq c$ where c is arbitrarily small.

von Haefen, Phaneuf, and Parsons show that the strict concavity of the utility function implies the algorithm will find the unique solution to the consumer's problem. Plugging the optimal solutions into (3.1) allows the analyst to evaluate the consumer's utility conditional on $(\mathbf{p}^1, \mathbf{Q}^1, \beta_t, \mathbf{E}_t)$ and income. Nesting this algorithm within the numerical bisection routine that iteratively solves for the income compensation that equates utility before and after the price and quality change allows the analyst to construct the individual's Hicksian compensating surplus conditional on $(\beta_t, \mathbf{E}_t)$.

Building on von Haefen, Phaneuf, and Parsons' numerical approach, von Haefen (2004a) has recently developed a more efficient numerical algorithm that relies on expenditure functions and the explicit definition of CS^H in equation (2.5). The computational savings arising from his approach are substantial; whereas von Haefen, Phaneuf, and Parsons' numerical approach requires the analyst to solve roughly two dozen constrained maximization problems, von Haefen's approach requires that the analyst solve only one constrained minimization problem.

Under additive separability the generic Kuhn-Tucker conditions for the individual's expenditure minimization problem are equation (3.12) and

$$\bar{u} = \sum_{k=1}^{M} u_k(x_k) + u_z(z). \tag{3.14}$$

A similar iterative algorithm can be used to solve the individual's constrained minimization problem. Assuming utility is strictly concave, the following algorithm finds the values of x and z that minimize expenditure:

1 At iteration i, set $z_a^i = (z_l^{i-1} + z_u^{i-1})/2$. To initialize the algorithm, set $z_l^0 = 0$ and $z_u^0 = u_z^{-1}(\bar{U} - \sum u_j(0))$.

2 Conditional on z_a^i, solve for $\mathbf{x}_i$ using (3.12). Solve for $\tilde{u}^i = U(\mathbf{x}^i, z_a^i)$ using (3.14).

3 If $\tilde{U}^i < \bar{U}$ set $z_l^i = z_a^i$ and $z_u^i = z_u^{i-1}$. Otherwise set $z_l^i = z_l^{i-1}$ and $z_u^i = z_a^i$.

4 Iterate until $|(z_l^i - z_u^i)| \leq c$ where c is arbitrarily small.

With the optimal solutions for $\mathbf{x}$ and z in hand, the expenditure needed to reach baseline utility under the new prices and quality can be calculated and used to efficiently compute the compensating surplus via (2.5).

3.6 Our welfare calculation approach: a restatement

For clarity, we now restate the key steps of our generic approach to constructing Hicksian welfare estimates:

1 On iteration i, simulate β_t initially from $f(\beta_t|\mathbf{x}_t)$ and $\mathbf{E}_t$ conditionally from $f(\mathbf{E}_t|\beta_t, \mathbf{x}_t)$. No simulation from $f(\beta_t|\mathbf{x}_t)$ is necessary for the fixed parameter classical model, and simulations from $f(\beta_t|\mathbf{x}_t)$ are generated at each step of Gibbs sampler in the random parameter Bayesian model. For the random parameter classical model, the adaptive Metropolis-Hastings algorithm described in section 3.4 can be used to generate simulations of β_t. For the classical and Bayesian random parameter specifications, we use every 10^{th} simulation after the burn-in to construct welfare estimates to reduce the serial correlation induced by the Monte Carlo Markov Chain. For each element of $\mathbf{E}_t$, set $\varepsilon_{tj} = g_{tj}(\cdot)$ if site j is visited and simulate ε_{tj} from the truncated type I extreme value distribution using equation (3.11) otherwise.

2 Conditional on the simulated $(\beta_t, \mathbf{E}_t)$, compute CS^H arising from the price and quality change using either von Haefen, Phaneuf, and Parsons' indirect utility function-based approach or von Haefen's expenditure function-based approach.

   (a) With the indirect utility function approach, we use numerical bisection to find the income differential that equates baseline and new utility levels. At each iteration of this search routine, use von Haefen, Phaneuf, and Parsons' algorithm to solve the consumer's constrained optimization problem and utility.

   (b) With the expenditure function approach, use von Haefen's algorithm to compute the minimum expenditure necessary to achieve baseline utility at the new price and quality values. Because this approach is significantly faster than von Haefen, Phaneuf, and Parsons', we use it exclusively in our subsequent application.

3 Average each of the simulated compensating surplus values to construct an estimate of $E(CS)$, the individual's expected Hicksian surplus measure.

4. Empirical illustration

To illustrate our estimation and welfare calculation approaches, we use 1992 moose hunting recreation data from Alberta, Canada.[6] The data was previously

[6]We thank Vic Adamowicz and Peter Boxall for generously providing this data.

analyzed by Adamowicz *et al.* (1997) and is further documented in McCloud *et al.* (1993). Here we present only a few of its salient features. In 1992, a survey of Alberta moose hunting license holders was conducted to ascertain the role of alternative forest management policies on moose hunting behavior. In-person interviews at five centrally located towns in the province were used to elicit revealed preference, stated preference, and demographic information. A total of 199 of the original 412 hunters who were initially contacted supplied complete revealed preference data and are the focus of our analysis.

The revealed preference component of the survey collected information about trip-taking behavior over the past year to fourteen Wildlife Management Units (WMUs). These WMUs represent forestry-intensive regions of west-central Alberta that range in size from roughly 1,000 to 13,000 hectares. Round trip travel distances were manually calculated with maps and rotary planimeters from each individual's residence to a point near the center of each WMU that could be reached by road or truck trail. Travel costs were constructed as the round trip travel distance valued at $0.27/km (1992 Canadian dollars) plus travel time valued at the wage rate for those who could have worked while they were hunting and zero for those who could not. Summary statistics for the trip and travel cost data for the fourteen sites, as well as additional economic and demographic variables used in our analysis, are given in Table 8.1.

Table 8.1. Summary Statistics (Sample Size = 199, Sites = 14)

Variable	*Description*	*Mean*	*Std. Err.*
Trips	Trips to all 14 WMUs	4.925	6.12
Tcost	Travel cost to each WMUs[a]	220	151
Income	Income [a]	51,722	22,809
Edmonton	Edmonton resident	0.412	0.493
Age	Age/100	0.396	0.107
HS_diploma	High school diploma	0.91	0.288
Some_post	Some post-secondary education	0.412	0.493
Col_grad	College graduate	0.06	0.239
Gen_exp	General hunting experience[b]	2.019	1.024
Moose_exp	Moose hunting experience[b]	1.688	0.987
Moose_den	Moose density[c]	0.568	0.341
FMA	Percentage of WMU in Forest Management Area	0.483	0.327

[a] In 1992 Canadian dollars.
[b] Years/10.
[c] Estimated moose hunting population per square hectare.

5. Results

5.1 Parameter estimates

Table 8.2 reports parameter estimates for three alternative Kuhn-Tucker specifications.[7] The first column contains fixed coefficient classical estimates, column two contains the restricted random coefficient classical estimates, and column three contains unrestricted random coefficient Bayesian estimates. All three sets of estimates are generated from models that consistently account for the truncation of trips (i.e., the fact that all 199 sample respondents take a positive number of trips) induced by the sampling design. For the Bayesian estimates, 78 additional parameters (the off-diagonal elements of the random parameter variance-covariance matrix) were also estimated but are not reported in Table 8.2. Thus, the Bayesian random parameter specification with 104 total parameters estimated can account for a much richer structure of unobserved heterogeneity relative to the classical random parameter specification with 20 parameters estimated. Moreover, the run time required to produce the Bayesian estimates was less than one-third the run time needed for the classical random parameter specification. Combined, these points suggest the significant appeal of the Bayesian framework when estimating Kuhn-Tucker demand models.

Across all three specifications, the parameter estimates reported in Table 8.2 suggest an increase in moose population per hectare (the *moose_den* variable) makes a WMU more attractive for hunters, although the Bayesian specification indicates there is variation in the strength of this effect in the population. Likewise, a percentage increase in protected Forest Management Areas (FMA) tends to make a WMU more attractive for moose hunters, although again the Bayesian estimates suggest diversity of preferences across the population. The presence of income effects in the data can be ascertained by the sign and magnitude of ρ^*, where a large negative estimate implies zero income effects. While the classically estimated models imply income effects are present, the Bayesian estimates suggest otherwise. This empirical finding is probably due to the fact that once one allows for a richer pattern for the unobserved heterogeneity through correlated random parameters, income's role in explaining behavior is substantially reduced.

In general, it is difficult to statistically discriminate among the alternative specifications. A naïve likelihood ratio test that the standard errors jointly equal zero in the random coefficient classical model suggests that random coefficient classical model fits the data better (p-value = 0.013).[8] However, using

[7] GAUSS code for all estimation and welfare calculations discussed in this chapter is available from either author upon request.

[8] Chen and Cosslett (1998) have pointed out that traditional hypothesis testing procedures should not be used in this context because the null hypothesis essentially restricts the standard error parameters to their

Table 8.2. Parameter estimates[a]

	Fixed Coeff. Classical	*Random Coeff. Classical*[b]		*Random Coeff. Bayesian*[c]	
Log-Likelihood	–1,590.71	–1,581.85		–	
CAIC	3,269.71	3,289.56		–	
Ψ Parameters	Fixed	Mean	St. Dev.	Mean	St. Dev.
Constant	0.0885	–0.8089	–	1.9468	1.5046
	(1.095)	(1.265)		(0.539)	(0.322)
Edmonton	0.6303	0.6276	0.0046	0.0241	1.1868
	(0.127)	(0.122)	(0.007)	(0.386)	(0.216)
Age	–3.0644	–2.935	0.0043	–0.7103	1.6494
	(1.161)	(0.984)	(0.500)	(1.036)	(0.375)
HS_diploma	0.1697	0.1892	0.2778	1.448	1.3513
	(0.305)	(0.257)	(0.098)	(0.575)	(0.297)
Some_post	–0.2033	–0.1965	0.2477	–0.4981	1.2161
	(0.137)	(0.147)	(0.288)	(0.423)	(0.223)
Col_grad	–0.2561	–0.2597	0.0038	–0.5564	1.4801
	(0.307)	(0.322)	(0.077)	(1.241)	(0.357)
Gen_exp	0.1604	0.1348	0.0174	–0.0804	1.1057
	(0.141)	(0.137)	(0.026)	(0.36)	(0.256)
Moose_exp	0.0916	0.0484	0.1616	–0.2091	1.1461
	(0.135)	(0.138)	(0.053)	(0.422)	(0.256)
ϕ Parameters					
Moose_den	0.9485	0.9097	–	1.0689	0.8817
	(0.113)	(0.110)		(0.183)	(0.132)
FMA	1.1185	1.1111	–	1.2773	1.0527
	(0.159)	(0.153)		(0.216)	(0.181)
Additional Parameters					
ρ^*	–0.8351	–0.6389	–	–6.6153	1.4453
	(0.254)	(0.228)		(1.28)	(0.354)
θ^*	2.1208	2.1487	–	2.2868	1.1136
	(0.176)	(0.169)		(0.229)	(0.201)
μ^*	0.6704	0.624	–	0.2983	0.559
	(0.027)	(0.028)		(0.076)	(0.047)

[a] For the classical models robust standard errors are reported in parentheses. For the Bayesian model posterior standard errors are reported in parentheses.
[b] Parameter estimates generated with 500 Halton draws.
[c] Posterior mean and standard error estimates generated with 50,000 Gibbs sampler iterations. Simulations from the first 25,000 iterations were discarded as burn-in, and simulations from every tenth iteration thereafter were used to construct the reported estimates.

lower bound values (i.e., zero). As a result, naïve hypothesis tests as we use here to discriminate fixed and random coefficient classical models will tend to reject the null hypothesis too infrequently. Because we reject the null hypothesis using the more stringent naïve critical values, we can conclude that we would have also rejected the hypothesis if we had used the correct critical values.

the more stringent Consistent Akaike Information Criteria (*CAIC* in Table 8.2) to discriminate between the models, we would choose the fixed coefficient specification. Although the Bernstein-von Mises theorem implies the Bayesian estimates have a classical interpretation, formally comparing the overall statistical fit of the Bayesian to the two classical models is difficult because of the fundamentally different paradigms used to derive the estimates.

Table 8.3. Welfare estimates

Policy Scenario	*Fixed Coeff. Classical*[a]	*Random Coeff. Classical*[b]	*Random Coeff. Bayesian*[c]
Increase in moose density to minimum	$22.14	$21.28	$25.38
threshold (0.5 per hectare) at all WMUs	(3.239)	(3.507)	(6.95)
$25 increase in entrance	–$93.15	–$92.55	–$91.81
fees at all WMUs	(0.777)	(0.884)	(1.233)
Closure of WMU-344	–$14.23	–$13.88	–$18.37
to moose hunting	(0.752)	(0.769)	(4.619)

For the classical estimates parametric bootstrap standard errors estimates based on 200 iterations are reported in parentheses. For Bayesian estimates posterior standard error estimates are reported in parentheses.

[a] For the classical estimates parametric bootstrap standard errors based on 200 iterations are reported in parentheses. For Bayesian estimates posterior standard errors are reported in parentheses.

[b] Welfare estimates generated with 2,500 draws. The first 500 were discarded as burn-in, and every tenth iteration thereafter was used to construct the reported estimates.

[c] Posterior means and standard errors generated with 50,000 Gibbs sampler iterations. Simulations from the first 25,000 iterations were discarded as burn-in, and every tenth iteration thereafter was used to construct the reported estimates.

5.2 Welfare estimates

Table 8.3 reports Hicksian welfare estimates for three separate policy scenarios:

- an increase in moose density to a minimum threshold of 0.5 moose per hectare in each WMU;
- a $25 increase in per trip entrance fees at all WMUs; and
- closure of WMU-344 to moose hunting.

The first policy scenario affects six of the most logged WMUs in the region and is designed to inform policy makers of the potential benefits of a more restrictive forestry policy in these regions. The second policy affects all fourteen sites equally and could assist policy makers in determining the welfare implications of a per trip increase in moose hunting fees. The third policy affects

only WMU-344, where past and present logging practices have substantially diminished moose populations.

The results in Table 8.3 suggest the three specifications imply qualitatively similar policy inference for all three scenarios. The welfare estimates for the fixed and random coefficient classical specifications are virtually identical across all three policy scenarios, and the random coefficient Bayesian estimates are slightly larger in absolute value for the first and third scenarios, but very similar for the second. Although these results might suggest that little is gained from moving from simpler to more complex models, we caution against drawing such an inference. In separate work, von Haefen, Phaneuf, and Parsons (2004) found relatively large welfare differences between fixed and random coefficient classical models, and although we are not aware of any previous studies that compare classical and Bayesian welfare measures, our sense is that the greater flexibility afforded by the Bayesian models might very well translate into meaningful differences in welfare estimates in future applications.

6. Discussion

Each chapter of this volume has highlighted how recent innovations in simulation-based methods have advanced the practice of environmental economics. In our view, the gains from simulation in KT approaches to nonmarket valuation have been transformational. Over twenty years ago Hanemann and Wales and Woodland proposed behavioral micro-econometric models that consistently account for interior and corner solutions, but only recently have simulation techniques been developed that allow these models to answer policy relevant questions. As a result, the last five years have seen KT models used to measure the benefits of a wide variety of environmental goods including beach nourishment programs, nonpoint source pollution abatement, and increases in angler catch rates. These advances have also permitted a reconsideration of fundamental issues arising in nonmarket valuation such as the role of weak complementarity (Herriges, Kling, and Phaneuf, 2004; von Haefen, 2004a) and site definition (Phaneuf and Herriges, 1999). Our objective in this chapter has been to summarize and illustrate the recent simulation-based innovations that have made these applications and methodological investigations possible. While five years ago the practical usefulness of KT models for nonmarket valuation was in doubt, it is now the case that parameter estimates and welfare measures for large demand systems that account for rich patterns of unobserved heterogeneity can be recovered in single overnight runs.[9]

[9]Indeed, all of the point and standard error estimates reported in this paper (admittedly for a relatively small data set) were generated in runs that took less than an afternoon of computer time, and the Bayesian estimation and welfare results in particular were generated in less than an hour.

Despite this progress, we should emphasize in closing that the KT approach to nonmarket valuation remains in its infancy. For the approach to mature and realize its full potential, considerable progress along several dimensions must be achieved. Perhaps the most significant is the development of tractable estimation and welfare calculation techniques for more flexible preference structures than the additively separable specifications considered in this chapter. Although we have focused exclusively on demand system approaches estimated within the primal framework here, our sense is that the dual (e.g., Phaneuf, 1999) as well as the primal approaches may be fruitful frameworks for developing these more general models. Wang's (2003) dissertation, in particular, suggests how the dual framework and Bayesian estimation techniques can be used to recover non-additive preference parameters for moderately large demand systems (15 goods) that flexibly allow for unobserved variation in tastes.

Two other lines of research may also produce significant improvements in existing KT modeling. The first involves a reconsideration in the KT framework of the economic and statistical insights that have been generated elsewhere in the recreation literature. In some sense, this process has already begun in that a number of the simulation-based estimation and welfare calculation techniques discussed in this chapter were first developed and used in the discrete choice context. Other possibilities include the use of spike and zero-inflated models to more fully address nonparticipation (von Haefen and Phaneuf, 2003), latent consideration set models to address choice set definition issues (von Haefen, 2004), and exploiting estimation procedures developed in the count data literature to address the problem of endogenous stratification arising with on-site sampling. In addition, KT models will likely benefit from a careful consideration of what insights from the aggregate data consumer demand system literature might transfer. Our sense is that the wealth of estimation strategies, demand specifications, and economic insights derived in the more traditional demand systems literature has the potential to significantly advance current KT practice.

Chapter 9

HIERARCHICAL ANALYSIS OF PRODUCTION EFFICIENCY IN A COASTAL TRAWL FISHERY*

Garth Holloway
Department of Agricultural and Food Economics
University of Reading
Reading, POBox 237, RG6 6AR, U.K.
garth.holloway@reading.ac.uk

David Tomberlin
National Marine Fisheries Service
110 Shaffer Road, Santa Cruz, CA 95060, U.S.A.
David.Tomberlin@noaa.gov

Xavier Irz
Department of Agricultural and Food Economics
University of Reading
Reading, PO Box 237, RG6 6AR, U.K.
x.t.irz@reading.ac.uk

Abstract We present, pedagogically, the Bayesian approach to composed error models under alternative, hierarchical characterizations; demonstrate, briefly, the Bayesian approach to model comparison using recent advances in Markov Chain Monte Carlo (MCMC) methods; and illustrate, empirically, the value of these techniques to natural resource economics and coastal fisheries management, in particular. The Bayesian approach to fisheries efficiency analysis is interesting for

*Copies of computer code used to process the data in this study are available from David Tomberlin at: David.Tomberlin@noaa.gov. We are grateful to several colleagues with whom we enjoyed discussions. Siddhartha Chib, Gary Koop and Bill Greene, in particular, deserve special thanks. We are also highly indebted to William Daspit and Brad Stenberg of the Pacific States Marine Fisheries Commission who were extremely helpful with data access and interpretation. Responsibility for error remains ours.

at least three reasons. First, it is a robust and highly flexible alternative to commonly applied, frequentist procedures, which dominate the literature. Second, the Bayesian approach is extremely simple to implement, requiring only a modest addition to most natural-resource economist tool-kits. Third, despite its attractions, applications of Bayesian methodology in coastal fisheries management are few.

Keywords: Hierarchical production efficiency, coastal trawl fishery

1. Introduction

Since its inception (Lindley and Smith, 1972)[1] the hierarchical, normal, linear model has been an important investigative structure for applied Bayesian practitioners. Arguably, its impact has been greatest in situations in which unobserved heterogeneity confounds traditional approaches to estimation, inference and model comparison. One important context — the topic of this paper — is the allocation of efficiency measures across subunits within a sample. In coastal fisheries analysis it is desirable to have available robust methods for comparing efficiency across firms (incumbent vessels), across time dimensions (years or seasons in which the vessels have operated), and across other dimensions important to the sampling process (including species and spatial aspects of the fishing activity). One feature of the data-generating environment is the likelihood that firm or seasonal effects are linked and that this linkage, in turn, leads to an underlying but random relationship in the data. The hierarchical approach allows one to draw these relationships together in a probabilistically consistent, but random, fashion. It allows one to make statements about the performance of subunits or units or the sample as a whole, recognizing the inter-linkages and the intra-linkages across time or across sample units. It is also advantageous from an inferential perspective. As Robert (2001) notes (p. 458):

> "A general characteristic of the hierarchical modelling is that it improves the robustness of the resulting Bayes estimators: while still incorporating prior information, the estimators are also well performing from a frequentist point of view (minimaxity and admissibility), although these two requirements are difficult to reconcile."

One additional motivation for hierarchical analysis is that it often opens additional avenues for empirical enquiry. Such avenues have often the ability

[1]For discussions of hierarchical models see Berger (1985); Gelman *et al.* (1995); and Robert (2001). More detailed representations appear in Deely and Lindley (1981); Dumouchel and Harris (1981); George (1986); and Draper (2000). Although Lindley and Smith (1972) are often credited with its first appearance, the hierarchical modelling concept is actually embedded in ideas originating from Good (see the references in Good (1980) and Good (1983) for origins).

to reap deeper insights from the target population and, usually, although not always, the policy environment. As Robert, again, notes (p. 458):

> "Additional justifications of hierarchical Bayes modelling stem from real life, since there are settings in medicine, biology, animal breeding, economics, and so on, where the population of interest can be perceived as a sub-population of a meta-population, or even as a subpopulation of a subpopulation of a global population."

To this list of applications we hope, within this chapter, to add the subjects 'natural resource economics' and 'fisheries economics' in particular. In fisheries economics interest centres on the efficiency of sampling units (usually the vessels that operate within the fishery) and post-estimation processing of the efficiency measures. Almost exclusively, it seems, these studies employ frequentist techniques, raising scope for alternative methodology (Fernàndez *et al.*, 2002) is the notable exception).

A typical investigation applies the maximum-likelihood procedure to a sample – most often a cross-section, although occasionally a panel – of boats within the fishery. The investigator then derives estimates of the relationship between the covariates affecting random catch and the catch level (usually referred to as the 'production frontier') and derives estimates of the distance by which each sub-unit differs from the frontier. The inefficiency measures so derived are then used to make statements that are relevant for policy and for the management of the fishery. At the heart of the investigation is the composed-error model of Aigner *et al.* (1977), which evolved from the deterministic framework of Aigner and Chu (1968). The composed error model has spawned a vast and seemingly endless literature and is the model of choice presently. In our context, the presence of panel data affords additional opportunity for analysis. Are the efficiencies of the boats in question constant across time periods, or have they changed in a directional manner? Are the efficiencies of the time periods constant, or have they also altered during the sample period? And, why? Through a carefully chosen subdivision of the panel units, some deeper questions arise. Are the efficiencies of the firms that exit lower than those that remain? Do recent entrants operate relative to the same frontier? And, so on.

At this point, it seems natural to outline, briefly, the way Bayesian investigation proceeds. Generally speaking, the objective is to derive a conditional distribution, say $\pi(\mathbf{A}|\mathbf{B})$ (see Mood *et al.,* 1974) for an excellent introduction to conditional probability, which summarizes all of the information available to the researcher. Bayesians set $\mathbf{B}$ equal to data (usually denoted $\mathbf{y}$) and set $\mathbf{A}$ equal to some unknown quantity of interest, usually a set of model parameters (and usually denoted $\boldsymbol{\theta}$). Sometimes deriving probability statements about the model itself is the objective. In this case we set $\mathbf{A}$ equal to the data and condition on the model of interest by setting $\mathbf{B}$ equal to the characteristics embedded in the model, say $\mathbf{M}$. Estimation of parameters, for

example, proceeds by extracting the form of the function $\pi(\boldsymbol{\theta}|\mathbf{y})$ and, where $\boldsymbol{\theta}$ is multidimensional, the marginal distributions of its component parts, say $\pi(\boldsymbol{\theta}_1|\mathbf{y}), \pi(\boldsymbol{\theta}_2|\mathbf{y}), \ldots, \pi(\boldsymbol{\theta}_K|\mathbf{y})$. In the case of model comparison, the objective is to derive what is termed the predictive density for the data, $\pi(\mathbf{y}|\mathbf{M})$. This is, essentially, the nature of the exercise. The procedure, therefore, is conceptually straight-forward, but the mechanics of characterizing $\pi(\boldsymbol{\theta}|\mathbf{y})$ and $\pi(\mathbf{y}|\mathbf{M})$ often prove troublesome. It is noteworthy that several of the other chapters within this important volume devote effort to explaining these mechanics. Our primary interest is the development of the hierarchical model. In the hierarchical approach we allow subsets of the data to evolve from distinctly different parameter values but require those values to be related, in turn, in a consistent stochastic manner to the common values in $\boldsymbol{\theta}$. The hierarchical approach, we argue, affords additional layers of insight that a purely conventional approach does not and, while we postpone more detailed discussion, one point seems particularly noteworthy at this stage. In the context of panel analysis with multiple dimensions (such as the example that follows), a natural question at the heart of policy is the linkage between the subunits that comprise the whole. Here we refer to the intrinsic, latent level of efficiency that characterizes the fishery as 'the whole' and note its importance. Regulators comparing the performances of alternative fisheries may seek a single, summary measure by which to rank and select; and analysts may seek the same measure, because, although ancillary to specific objectives, it enables robust analysis. In short, the presence of an over-arching or under-pinning summary measure is useful because it allows the model to better replicate data generation.

When the data are multi-dimensional questions often arise about the interlinkages of the sub-units within the hierarchical chain. In the fishery that we analyse, the data consist of a panel of 13 years (1987-1999) of observations across a cross-section of 13 boats which fished a single location, namely, offshore Washington State. How the inefficiency measures evolve within this fishery is the topic of attention. Inevitably this question leads to a set of comparisons across models which, for various reasons, we deem competitive. An important contribution of this chapter is to lay out, as cogently as possible, how this model comparison proceeds; how it relates to Bayesian estimation of the composed error model using MCMC (Markov Chain Monte Carlo) techniques; and how the added information that it offers in the context of fisheries efficiency analysis is useful. The model-comparison technique is developed in a seminal paper by Chib (1995) and, although some time has elapsed since its introduction, it remains hitherto under-exploited in the natural-resource and fisheries-economics literatures.[2] Comparing alternative specifications of efficiency in the Washington

[2]Notably, the model-comparisons technique is employed in at least one of the other chapters in this volume (Layton and Levine, this volume chapter 10).

fishery, we showcase the Chib (1995) calculation in an important empirical setting. Designed for the novice, we lay out pedagogically the details of hierarchical analysis; demonstrate that Bayesian composed error modelling sheds light on important questions in the coastal fishery; and, because it offers a flexible and tractable alternative to current methodologies, offers enormous scope for innovation in fisheries efficiency analysis. Section two reviews, briefly, recent literature on fisheries efficiency. Section three presents key aspects of the Bayesian hierarchical approach to composed error modelling. Section four presents the background to the empirical model and section five presents the data. Empirical results are presented in section six and conclusions are offered in section seven. Extensions are also discussed.

2. Coastal fisheries efficiency measurement

Until the late 1990s, examples of measurement of technical efficiency in fisheries were few (Bjorndal *et al.*, 2002). Recent contributions, however, are numerous (Andersen, 2002) and we now turn, briefly, to review the literature on efficiency measurement in fisheries; present the predominant methods used; discuss the practical and conceptual difficulties involved; and summarize lessons emerging from this work.

2.1 Why measure technical efficiency in fisheries?

Technical-efficiency-in-fisheries studies improve our understanding of the relationship between the inputs employed and the catch obtained in fishing operations. The results often culminate in indices of economic performance and often characterize relative performances among vessels. The common objective is to suggest policies that may enhance the overall productivity of the fishery. More recently, the results have been used to evaluate the performance of various regulatory and conservation policies. Examples include Kompas *et al.* (2003); Viswanathan and Omar (2001); Alvarez (2001); and Pascoe *et al.* (2001).

2.2 Modelling Technical Efficiency

Focusing on technical efficiency excludes analyses of allocative and economic efficiency, which abound in the general economics literature. However, allocative- and economic-efficiency studies require price data and require that the principal objective of boat captains consists of maximising profits or minimising costs. Here we restrict attention to the purely technical relationships between inputs and outputs. One branch of early empirical work comparing catch per unit of effort of the operating boats has been criticized (Wilen, 1979) because it takes into account neither the multi-dimensional nature of effort nor the randomness of the fisheries production process. Hence, recent studies rely

on multi-dimensional frameworks and accommodate random shocks in the production process. These studies are attractive for reasons other than the fact that they can be implemented with only input and output quantities, but this remains one compelling reason for their prevalence in applied econometric work.

Methods of efficiency measurement (reviewed in Coelli *et al.*, 1998 and in Fried *et al.*, 1993) are divisible into two categories, namely parametric and non-parametric approaches. Data envelopment analysis (DEA) (Charnes *et al.*, 1978) is the predominant non-parametric approach. The method has been used extensively in agriculture and in banking but has been used only rarely in fisheries (Walden *et al.* (2003) is one recent exception). The relative unpopularity of non-parametric approaches stems from their inability to accommodate randomness, which, because of the biological processes impacting marine resources (weather, resource availability and environmental influences), is fundamental to fisheries-efficiency analysis (Bjorndal *et al.*, 2002; Kirkley *et al.*, 1995). In addition, non-parametric methods generate little information about the nature of the harvesting technology, such as its returns to scale or the degree of substitutability among inputs and outputs, both of which have important implications for fisheries management.

First generation, parametric methods (Aigner and Chu, 1968; Afriat, 1972) rely on deterministic frontiers and do not accommodate randomness. Although now largely obsolete, these methods are noteworthy because they include the first recorded fisheries efficiency study (Hannesson, 1983). Nowadays, parametric studies are derived from stochastic (composed-error) production frameworks (first proposed by Aigner *et al.*, 1977). Unlike its precursor, composed-error models allow technical inefficiency to be distinguished from random shocks. A generic formulation has structure

$$y_i = f(\mathbf{x}_i; \boldsymbol{\beta}) + u_i - z_i, \tag{2.1}$$

where y_i denotes output of boat *i* and $f(\mathbf{x}_i; \boldsymbol{\beta})$ is the production function that depends on a vector of parameters $\boldsymbol{\beta}$ and a vector of inputs $\mathbf{x}_i$. The stochastic component of the model includes a random shock u_i, which is usually assumed to be identically, independently and symmetrically distributed with variance σ^2; whereas the inefficiency term, z_i, follows a one-sided, non-negative distribution. The likelihood function can be expressed algebraically and maximized numerically to produce estimates of the unknown parameters. Typically, the inefficiency components take half-normal, truncated-normal or gamma distributed forms and predicted inefficiencies can be computed for each boat (Jondrow *et al.*, 1982). One major limitation of this approach is that the efficiency scores cannot be estimated consistently (Khumbhakar and Lovell, 2000). Fisheries applications are numerous (see, for example, Kirkley *et al.*, 1995; Grafton *et al.*, 2000).

A somewhat deeper issue absorbing considerable econometric attention is locating the determinants of inter-vessel differences in efficiency; but the task, it seems, is beset with difficulty. Because, during estimation, it is usually necessary to invoke the assumption that the z_i are iid across firms, it is inconsistent, in a second step, to explain efficiency levels by firm-specific characteristics (see Coelli *et al.*, 1998; for discussion). A one-step procedure is developed by Battese and Coelli (1995) whereby the distribution of the $z_i \geq 0$ is truncated normal with mean μ_i determined according to

$$u_i = \mathbf{w}_i'\delta, \tag{2.2}$$

where $\mathbf{w}_i \equiv (w_{1i}, w_{2i}, \ldots, w_{Ki})'$ denotes the variables conjectured to influence efficiency and $\delta \equiv (\delta_1, \delta_2, \ldots, \delta_K)'$ denotes their impact on the conditional mean, $\boldsymbol{\mu}_i$. A statistically significant coefficient associated with any w_{ij} indicates that the variable in question affects boat efficiency. This model has been applied extensively in fisheries research and remains popular currently. Recent examples include Pascoe and Coglan (2002); Squires *et al.* (1999); Andersen (2002); and Pascoe *et al. (*2003). The model's popularity is almost surely due to its ease of estimation as facilitated by *FRONTIER*, the freely-available software developed by Coelli.

Panel data remain mostly unexploited in fisheries efficiency analyses. In fact, with the exception of Squires *et al.* (2003) each of the aforementioned studies have panel structures with repeated (time series) observations on each boat. Yet, the potential richness of the panel structure remains untapped (Alvarez, 2001). In an effort to address this issue two recent approaches are noteworthy. One approach consists of modifying the previous framework to allow inefficiencies to vary parametrically over time. The most popular model here is the one advocated by Battese and Coelli (1992). They suggest that the inefficiency terms $z_{i1}, z_{i2}, \ldots, z_{iT}$ evolve according to

$$z_{it} = z_i \exp\left[-\eta(t - T)\right], \tag{2.3}$$

where the z_i are *iid* normal with mean μ, truncated at zero; *t* denotes the present time period; *T* denotes the terminal period; and η denotes a parameter to be estimated. In this model, technical inefficiencies are a monotonic function of time, which is increasing for $\eta < 0$ and decreasing for $\eta > 0$ and, thus, the estimation of η is paramount. The underlying idea, they claim, is that managers should learn from previous experience and technical inefficiency should evolve in some consistent pattern (Coelli *et al.*, 1998). Applications include Herrero and Pascoe (2003), Pascoe *et al.* (2003) and Bjorndal *et al.* (2002).

A second approach to time-varying efficiency adjusts standard panel estimation techniques. The advantages are that panel-data models do not require the strong distributional and independence assumptions made in maximum-likelihood estimation and that panel models permit the consistent estimation

of efficiency scores (Schmidt and Sickles, 1984). One common reformulation of equation (2.1) assumes the existence of a firm-specific, non-stochastic effect. Under the assumption that z_i is itself dependent on a vector of covariates the model is estimated using ordinary least squares (OLS) and produces consistent estimates of the efficiency scores (Kumbakhar and Lovell, 2000). Unfortunately, the approach has additional limitations. Most notably, the fixed effects provide reliable estimates of firm-level technical efficiency if (and only if) each of the frontier covariates is time-variant. Time-invariant covariates create major problems, which is a point worth emphasizing in the current context. In fisheries investigations boat characteristics remain largely invariant over the sample period and many of the frontier covariates will be time-invariant (see Alvarez, 2001 for a detailed discussion). Consequently, the fixed-effects model has rarely been applied to fisheries efficiency investigations (although Alvarez *et al.* (2003) lists Squires and Kirkley (1999) and Kirkley *et al.* (1995) as exceptions).

The introduction of random-effects into fisheries efficiency studies is important for two reasons. First, their introduction enables the time-invariance issue to be circumvented. Second, the (classical) random-effects methodology provides an important link to the hierarchical methodology that we exploit. As noted by McCulloch and Rossi (1994), in the Bayesian context there is no distinction between fixed and random-effects (because the parameters themselves are random), there is only a distinction between hierarchical and non-hierarchical analyses. However, Koop *et al.* (1997) suggest that it is useful, for pedagogic purposes, to retain this (frequentist) terminology. Applications of the random-effects methodology in fisheries include Squires and Kirkley (1999) and Alvarez and Perez (2000).

2.3 What have we learned?

We now attempt to draw general lessons from the technical-efficiency-in-fisheries literature. One natural question is whether fishing operations are technically efficient. Generally speaking, authors using maximum-likelihood test the null hypothesis of absence of inefficiencies (Viswanathan and Omar, 2001; Pascoe and Coglan, 2002; Squires *et al.*, 2003; Kirkley, *et al.*, 1995; Bjorndal *et al.*, 2002; Kompas *et al.*, 2003). In virtually all cases, the stochastic frontier specification is accepted over the alternative hypothesis consisting of a single, symmetric error. Hence, non-negligible inefficiency appears to be the rule rather than the exception. This point is important because inefficiency affects the equity and efficiency impacts of regulation. Second, refutation of full efficiency guides additional empirical inquiry. Because the parameters of a production function estimated conventionally are biased when inefficiencies are present, the statistical significance of inefficiency implies that important

characteristics (such as returns to scale or input-output jointness) are best investigated using a stochastic frontier.

The distribution of technical performance in fishing fleets is largely context-specific. Estimates of average efficiencies vary from less than 40 percent for the otter trawlers of the North Sea demersal fishery in the early 1990s (Bjorndal *et al.*, 2002) to 88 percent for the Malaysian gill net artisanal fishery (Squires *et al.*, 2003). Higher moments of the distribution also vary widely across studies. For example, Viswanathan and Omar (2001) find a large spread of efficiency scores for the Malaysian trawl fishery, which contrasts with the relatively concentrated distribution reported by Pascoe and Coglan (2002) for the English Channel fishery. The literature, it seems, does not lead to meta-conclusions concerning the overall distribution of efficiency scores.

Because of its policy implications and its link to the theoretical literature, the analysis of how regulations and management influence fisheries efficiency is a matter of much interest. Here, again, the literature presents a wide range of results. However, distinct themes do emerge. First, measures aimed at solving the open-access nature of fisheries tend to increase the general level of technical efficiency in the fishery. Second, this positive influence is greater the more flexible is the policy intervention. For example, Pascoe *et al.* (2003) find that area restrictions increase efficiency due to reductions in congestion and 'crowding out.' Felthoven (2002) concludes that the move towards well-defined property rights leads to increased technical efficiency and reduces excess capacity. And Grafton *et al.* (2000) find that the introduction of individual vessel quotas results in increased efficiency, although it seems that gains were slow to materialize and remained somewhat limited by restrictive regulations on property rights, such as bundling or limiting transferability, duration and divisibility.

The effect of the introduction of alternative regulations is the focus of some attention. For example, Kompas *et al.* (2003) find that increasing restrictions on vessel size and engine capacity has a negative effect on technical efficiency and Pascoe et al. (2003) conclude that gear restrictions and restrictive TACs have negative impacts on efficiency. Finally, Andersen (2003) establishes that a move from monthly limits towards more flexible yearly catch limits enhances technical efficiency.

A second strand of the literature aims to test the so-called 'good captain hypothesis.' This hypothesis posits that differences in catches among vessels is largely explained by the distribution of skills among skippers. The composed-error model appears particularly suited to test this hypothesis because it distinguishes the effect on production of luck (random events captured by the symmetric error term u_i in equation (2.1) as distinct from the effect of management decisions (embodied in the non-negative term z_i). Initial studies on this topic (see Alvarez *et al.* 2003 for a review) lend support to the hypothesis.

For example, Viswanathan and Omar (2001) find that skipper skills are key determinants of production, but that the concept of skill is difficult to relate to the socio-demographic attributes of the skipper. More recently, Alvarez *et al.* (2003) demonstrate that luck *per se* is more important than skill in explaining catch. Their estimates suggest that skills accounted for up to 9% of the total variation in catches between boats compared to approximately 11% for luck. Pascoe and Coglan (2002) reach similar conclusions. Generally speaking, both skipper skills and luck matter when explaining variations in boat catches, but their relative importance cannot be disentangled *a priori*.

Finally, several studies explain technical performance by factors other than regulations or skills. In particular, vessel characteristics appear to be particularly important explanatory factors. For example, Pascoe and Coglan (2002) and Pascoe *et al.* (2003) find that vessel age has a negative impact on technical efficiency. This finding is important because it suggests that boat vintages should be considered when measuring fleet capacity. There are also indications that larger vessels tend to achieve higher levels of technical efficiency (Andersen 2003, Basch *et al.*, 2002; Kompas *et al.*, 2003; Pascoe *et al.*, 2003), although examples to the converse (most notably, Eggert, 2001) exist. Altogether, it is fair to conclude that the factors explaining technical efficiency in fisheries differ considerably by locality, time period, and the full heterogeneity of the extant conditions (see Squires *et al.* 2003 for further discussion).

In summary, studies of technical efficiency in fisheries are innovative, highly heterogeneous and mostly policy-driven; the potential richness of the panel-data environment remains mostly untapped; and the literature to date is dominated by the maximum-likelihood procedures developed by Battese and Coelli (1992, 1995). Notwithstanding the importance of previous work, it seems that scope exists for analysis with alternative methodology.

3. Bayesian composed error model development

By way of notation, let $\boldsymbol{\theta}$ denote a vector of parameters of interest, $\pi(\boldsymbol{\theta})$ the prior probability density function (pdf) of $\boldsymbol{\theta}$ and $\pi(\boldsymbol{\theta}|\mathbf{y})$ the posterior pdf, where $\mathbf{y} \equiv (y_1, y_2, \ldots, y_N)'$ denotes data. Frequently we reference the data-generating model $f(y|\boldsymbol{\theta})$ which, when viewed as a function of $\boldsymbol{\theta}$, is the likelihood function. Sometimes we make use of variants of the $f(\cdot|\cdot)$ notation in order to reference particular probability density functions. The pdfs that we refer to are:

- the m-dimensional multivariate normal pdf:
  $f^N(\mathbf{x}|\boldsymbol{\mu}, \Sigma) \equiv (2\pi)^{-m/2}|\Sigma|^{-1/2} \exp\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\}$;
- the univariate truncated-normal pdf:
  $f^{TN}(x|\mu, \sigma) \equiv (2\pi)^{-1/2}\sigma^{-1/2} \exp\{-\frac{1}{2}\sigma^{-2}(x-\mu)'(x-\mu)\}/(1-\Phi)$;

- the gamma distribution:
  $f^G(x|\rho,\lambda) \equiv \Gamma(\rho)^{-1}\lambda^\rho x^{\rho-1}\exp\{-\lambda x\}$
- and two variants, namely the inverted-gamma pdf:
  $f^{IG}(x|v,s) \equiv (2/\Gamma(v/2))(vs^2/2)^{v/2}(1/x^{v+1})\exp\{-vs^2/2x^2\}$
  and the exponential pdf:
  $f^E(x|\lambda) \equiv \lambda\exp\{-\lambda x\}$.

Often, we reference just the variable part of the density (in other words, the part with the integrating constant excluded) by noting its proportionality to a true pdf and by using the symbol '$\propto$'. Thus, the multivariate normal pdf, for example, is proportional to the scale factor '$\exp\{-1/2(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\}$' and we write $f(\mathbf{x}|\boldsymbol{\mu},\Sigma) \propto \exp\{-1/2(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\}$. Finally, we use $i = 1,2,\ldots,N$ to denote the upper-level units within the panel; $j = 1,2,\ldots,N_i$ to denote sub units; $k = 1,2,\ldots,N_{ij}$ to denote lower-level units, and so on. In this way, a sample observation with three dimensions relevant to its generation is denoted by subscripts 'ijk'.

3.1 Composed Error Frontier Model With Exponential Inefficiency

Because a part of our remit is pedagogic, we first introduce a basic cross-section model to which we manufacture subsequent extensions. The basic formulation is the composed-error model with a normal distribution for the sampling error and an exponential distribution for the inefficiency. This construction has many limitations, but serves to illustrate the types of calculations that we encounter subsequently.

The observational equation is:

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + u_i - z_i; \tag{3.1}$$

where y_i denotes the level of output produced by sample unit i; $\mathbf{x}_i$ denotes a K-vector of covariates relevant to the production frontier; the vector $\boldsymbol{\beta}$ denotes the relationship between the frontier and the K covariates; u_i denotes random error; and z_i denotes displacement between the frontier and the output of sample unit i. We assume that sampling error is normally distributed with mean zero and variance σ^2 and assume that the displacement value, z_i, is positive. Thus, z_i, denotes the level of 'inefficiency' corresponding to sample unit i and, because larger values of z_i correspond to output realizations farther below the frontier, sample units (firms) with larger z_i are 'less efficient'. For a similar development in the context of cost-function frontiers see Koop *et al.* (1997).

Much of our illustration centres around alternative specifications of the inefficiency term, z_i, and many of the early developments in the literature involve

generalizations of their structure. However, a convenient starting point is to assume that the z_i are derived from a common exponential distribution. Formally, assume that $\mathbf{z} \sim f^{\mathbf{E}}(z_i|\lambda)$. While this assumption may be inappropriate for many empirical settings, it serves as a benchmark from which to incorporate additional institutional detail. Its immediate advantage is simplicity.

With a normal distribution assumed for the sampling error and an exponential distribution assumed for the inefficiency, the data density for a single observation is

$$f(y_i|\boldsymbol{\theta}) = f^N(y_i|\mathbf{x}_i'\boldsymbol{\beta} - z_i, \sigma) \times f^E(z_i|\lambda), \tag{3.2}$$

which is the product of a (conditional) normal distribution for the observed data and a (marginal) exponential distribution for the unobserved inefficiency. Throughout the paper we follow the convention that (lower-case) Roman numerals denote data and (lower-case) Greek symbols denote parameters. It is worth emphasizing that the observable entities in the sampling model in (3.2) are the output level, y_i, and the vector of covariates $\mathbf{x}_i$. The unobservable elements are the scale parameter, σ; the vector of location parameters, $\boldsymbol{\beta}$; the parameter λ underlying the inefficiency distribution, $f^E(\cdot|\cdot)$; and the level of inefficiency itself, z_i. The interpretation of the z_i terms deserves discussion. Although they are unknown, estimated from the data, and justify interpretation as 'parameters' of the model, we prefer to view them as 'latent' or 'missing data.' This distinction is important for two reasons. The first concerns whether a prior is placed directly on the z_i's or, instead, on the upper level parameter, λ, that constitutes the hierarchy. In our analysis we consider priors only on the upper-level components. Second, the specific interpretation of the z_i's becomes important when we estimate the model's marginal likelihood, which is the key input to model comparison. When the z_i's are latent they are not considered part of the model parameters and, so, a prior and a posterior density value, which appear for each model parameter, do not enter calculations. Consequently we use $\boldsymbol{\theta} \equiv (\sigma, \boldsymbol{\beta}', \lambda)'$ to denote the model parameters (which are unknown) and use $\mathbf{z}$ to denote the vector of N inefficiencies (also unknown) across the sample. Hence, assuming independence across observations, the density of the data, $\mathbf{y}$, given the parameters, $\boldsymbol{\theta}$, is[3]

$$f(y|\boldsymbol{\theta}) = \prod_{i=1}^{N} f(y_i|\boldsymbol{\theta}). \tag{3.3}$$

[3]Another issue concerns notation. With $\mathbf{z} \equiv (z_1, z_2, \ldots, z_N)$ latent, it is sometimes important to distinguish between the *observed-data* likelihood, generically, $f(\mathbf{y}|\boldsymbol{\theta})$ and the so-termed *complete-data* likelihood $f(\mathbf{y}|\boldsymbol{\theta}, \mathbf{z})$ which is conditioned by the latent data. The likelihood in (3.3) is actually the latter, in which the latent $\mathbf{z}$ is suppressed.

The econometrician observes y but not $\boldsymbol{\theta}$; forms a prior pdf for the unknowns, $\pi(\boldsymbol{\theta})$; and conducts posterior inference with respect to the resulting posterior,

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}), \tag{3.4}$$

which is proportional to the product of the data density and the prior pdf. Assuming that information *a priori* is diffuse, we employ the improper measure (the so-called Jeffrey's prior) $\pi(\boldsymbol{\theta}) \propto \sigma^{-1}$. Consequently, the posterior pdf assumes the form

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto \sigma^{-1} \prod_{i=1}^{N} f^N(y_i|x_i'\boldsymbol{\beta} - z_i, \sigma) f^E(z_i|\lambda) \tag{3.5}$$

and the key ingredients of our Bayesian approach are now in place. The posterior, however, is difficult to work with. The objective, which is common to all Bayesian exercises, is to derive the marginal distributions that summarize knowledge about the parameters, $\boldsymbol{\theta}$, conditional on the data, $\mathbf{y}$. Here, the phrase 'difficult to work with' means that at least one of the required posterior marginal distributions – respectively, $\pi(\sigma|y)$, $\pi(\boldsymbol{\beta}|y)$ and $\pi(\lambda|y)$ – is not available in closed form. The density $\pi(\sigma|y)$ is the inverted-Gamma form and the density $\pi(\boldsymbol{\beta}|y)$ is the multivariate-T distribution, but the marginal distribution $\pi(\lambda|y)$ is not easily derived. The reason is that the posterior distribution for λ depends on the z_i's and, although the conditional distribution of λ, namely $\pi(\lambda|\mathbf{y}, \mathbf{z})$, has a simple form (this conditional distribution is the Gamma distribution), the marginal distribution is slightly more difficult to obtain.

Although, in this simplified setting, this integration can be achieved (the integration involves the well-known Gamma integral, see, for example, Mood *et al.*, 1974; p. 534), in other, more general settings such integration is usually not possible, leaving us with a major impediment to analysis. However, a solution exists. The solution is to use Gibbs sampling.

3.2 Gibbs sampling the composed error model

The necessary conditions for Gibbs sampling the composed error model, or any model, for that matter, are two. First, the fully conditional distributions comprising the joint posterior must be available in closed form. Second, these forms must be tractable in the sense that it is easy to draw samples from them. In the present setting each of the full conditional distributions has a convenient form. Specifically, the conditional distribution of the regression standard error, $\pi(\sigma|\boldsymbol{\beta}, \lambda, \mathbf{z}, \mathbf{y})$, is the inverted-Gamma pdf; the distribution of the location vector, $\pi(\boldsymbol{\beta}|\lambda, \sigma, \mathbf{z}, \mathbf{y})$, is the multivariate-Normal pdf; the distribution of the mean inefficiency, $\pi(\lambda|\sigma, \boldsymbol{\beta}, \mathbf{z}, \mathbf{y})$, is the Gamma pdf; and the fully conditional distribution of the vector of latent inefficiencies, $\pi(\mathbf{z}|\sigma, \boldsymbol{\beta}, \lambda, \mathbf{y})$, is the truncated-normal distribution. In addition, we note that $\pi(\sigma|\boldsymbol{\beta}, \lambda, \mathbf{z}, \mathbf{y})$ and

$\pi(\boldsymbol{\beta}|\lambda, \sigma, \mathbf{z}, \mathbf{y})$ are independent of λ ; that $\pi(\lambda|\sigma, \boldsymbol{\beta}, \mathbf{z}, \mathbf{y})$ is independent of σ, $\boldsymbol{\beta}$ and $\mathbf{y}$; and that $\pi(\mathbf{z}|\sigma, \boldsymbol{\beta}, \lambda, \mathbf{y})$ is dependent on all of the conditioning variables. Hence, we focus attention on $\pi(\sigma|\boldsymbol{\beta}, \mathbf{z}, \mathbf{y})$, $\pi(\boldsymbol{\beta}|\sigma, \mathbf{z}, \mathbf{y})$, $\pi(\lambda|\mathbf{z})$ and $\pi(\mathbf{z}|\sigma, \boldsymbol{\beta}, \lambda, \mathbf{y})$ and turn to two issues that, although now de-emphasized in the theoretical and applied literatures, are of much importance to pedagogic development.

The Gibbs sampler has origins that stem from a seminal work in the 1950's (Metropolis *et al.*, 1953), its exploration in the 1980's (Geman and Geman, 1984) and its subsequent re-interpretation under Bayesian assumptions in the early 1990's (Gelfand and Smith, 1990; Gelfand *et al.*, 1990). Essentially, iterative sampling from the component conditional distributions comprising the joint posterior forms a Markov chain. Under weak regularity conditions, this chain converges in distribution to the true marginal quantities (the marginal posterior pdfs) that we seek (Tierney, 1994). Since the early 1990s, this realization has spawned a veritable explosion in Bayesian applications in the pure and social sciences. An excellent introduction to the Gibbs sampler is Casella and George (1992) and an excellent introduction to the closely related Metropolis-Hastings algorithm is Chib and Greenberg (1995). Other useful introductions to some of the techniques employed here include Gelman *et al.*, 1990; Robert and Casella, 1999; Robert, 2001; and Liu, 2001. The other issue that present extensions of the literature de-emphasize, is the process by which we confirm the forms of the fully conditional distributions. Rather than relegate derivations to an appendix, it seems pertinent and informative to motivate them in the context of the simple, exponential formulation. The procedure is very simple and relies only on knowledge of the forms of the component distributions, some basic algebra and the notion that only the shapes of the distributions in question are of interest. Specifically, we view the posterior from four different vantage points, conditioning on alternative sets of quantities each time. This practice is made simpler by reformulating the posterior (equation (3.5)) in terms of the vector $\mathbf{z}$; writing

$$\pi(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z}) \propto \sigma^{-1} f^N(\mathbf{y}|\mathbf{x}\boldsymbol{\beta} - \mathbf{z}, \sigma^2 \mathbf{I_N}) f^E(\mathbf{z}|\lambda), \tag{3.6}$$

where $x \equiv (x_1', x_2', \ldots, x_N')'$ and $\mathbf{I_N}$ denotes the N-dimensional identity matrix; and working, sequentially, with respect to σ, $\boldsymbol{\beta}$, λ and $\mathbf{z}$.

First, when the posterior is viewed as a function solely of the regression standard error, σ, the pdf representing the exponential inefficiencies drops out (because σ does not appear as a function of the exponential distribution). In addition, the remaining terms in the multivariate normal distribution that do not contain σ are ignored and we are left with the expression

$$\pi(\sigma|\boldsymbol{\beta}, \mathbf{z}, \mathbf{y}) \propto \sigma^{-(N+1)} \exp\{-.5\sigma^{-2}(\mathbf{y} - \mathbf{x}\boldsymbol{\beta} + \mathbf{z})'(\mathbf{y} - \mathbf{x}\boldsymbol{\beta} + \mathbf{z})\}, \tag{3.7}$$

as the form of the conditional posterior for σ. Now, with reference to a standard source (for example, Zellner, 1996 equation (A.37b), p. 371), the form of the function on the right-side of (3.7) and the form of the inverted Gamma distribution are seen to be the same when all inconsequential constants are ignored. Hence, we redefine $\nu \equiv N$ and $\nu s^2 \equiv (\mathbf{y} - \mathbf{x}\boldsymbol{\beta} + \mathbf{z})'(\mathbf{y} - \mathbf{x}\boldsymbol{\beta} + \mathbf{z})$. Consequently, $\pi(\sigma|\boldsymbol{\beta}, \mathbf{z}, \mathbf{y})$ has the inverse-Gamma form; we write $\pi(\sigma|\boldsymbol{\beta}, \mathbf{z}, \mathbf{y}) \equiv f^{IG}(\sigma|\nu, s^2)$; and we note that one component of the Gibbs-sampling algorithm will be a draw from the inverted-Gamma distribution. There are several algorithms available to simulate a draw from an inverse-Gamma distribution, one of which, used currently, involves drawing from the so-called scaled inverse chi-squared distribution (see Gelman *et al.,* 1995).

Second, we view the posterior solely as a function of the location vector $\boldsymbol{\beta}$. Once again, the exponential distribution drops out, constants preceding the kernel are ignored and we are left with

$$\pi(\boldsymbol{\beta}|\sigma, \mathbf{z}, \mathbf{y}) \propto \exp\{-.5\sigma^{-2}(\mathbf{y} - \mathbf{x}\boldsymbol{\beta} + \mathbf{z})'(\mathbf{y} - \mathbf{x}\boldsymbol{\beta} + \mathbf{z})\}. \tag{3.8}$$

This form can be brought into the multivariate normal form by the process of 'completing the square' in $\boldsymbol{\beta}$ (see Zellner, pp. 380-381, for an example). Redefining $\mathbf{a} \equiv \mathbf{y} + \mathbf{z}$, the right side of equation (3.8) is:

$$\begin{aligned}
\exp\{-.5\sigma^{-2}(\mathbf{a} - \mathbf{x}\boldsymbol{\beta})'(\mathbf{a} - \mathbf{x}\boldsymbol{\beta})\} \propto \\
\exp\{-.5\sigma^{-2}\{-\mathbf{a}'\mathbf{x}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{x}'\mathbf{a} + \boldsymbol{\beta}'\mathbf{x}'\mathbf{x}\boldsymbol{\beta}\}\} \propto \\
\exp\{-.5\{\boldsymbol{\beta}'(\sigma^{-2}\mathbf{x}'\,\mathbf{x})\boldsymbol{\beta} - \boldsymbol{\beta}'(\sigma^{-2}\mathbf{x}'\mathbf{a}) - (\sigma^{-2}\mathbf{a}'\mathbf{x})\boldsymbol{\beta}\} \propto \\
\exp\{-.5(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\mathbf{C}_{\hat{\boldsymbol{\beta}}}^{-1}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\}
\end{aligned}$$

where $\hat{\boldsymbol{\beta}} \equiv (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{a}$ and $\mathbf{C}_{\hat{\boldsymbol{\beta}}} \equiv \sigma^2(\mathbf{x}'\mathbf{x})^{-1}$. This form is the same as that for the multivariate normal distribution presented in Zellner (equation (B.1), p. 379) with the redefinitions $\mathbf{x} \equiv \boldsymbol{\beta}$, $\boldsymbol{\theta} \equiv \hat{\boldsymbol{\beta}}$ and $\Sigma \equiv \mathbf{C}_{\hat{\boldsymbol{\beta}}}$. Consequently, $\pi(\boldsymbol{\beta}|\sigma, \mathbf{z}, \mathbf{y}) \equiv f^{MN}(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}, \mathbf{C}_{\hat{\boldsymbol{\beta}}})$ and the second component of the algorithm will be a draw from the multivariate-normal distribution.

Third, we seek the full conditional distribution for the latent $\mathbf{z}$. In contrast to the previous two situations, the elements of $\mathbf{z}$ span both distributions and the derivation is slightly more complex, but only slightly. Ignoring irrelevant constants, we have

$$\begin{aligned}
\pi(\mathbf{z}|\sigma, \boldsymbol{\beta}, \lambda, \mathbf{y}) \quad &\propto \quad \exp\{-.5\sigma^{-2}(y - \mathbf{x}\boldsymbol{\beta} + \mathbf{z})'(y - \mathbf{x}\boldsymbol{\beta} + \mathbf{z})\} \\
&\times \quad \exp\{-\lambda\mathbf{z}\}.
\end{aligned} \tag{3.9}$$

Upon completing the square, we derive that $\pi(\mathbf{z}|\sigma, \boldsymbol{\beta}, \lambda, \mathbf{y}) \equiv f^{TN}(\mathbf{z}|\hat{\mathbf{z}}, \mathbf{C}_{\hat{\mathbf{z}}})$, which is the truncated normal distribution with parameters $\hat{\mathbf{z}} \equiv -\mathbf{y} + \mathbf{x}\boldsymbol{\beta} - \lambda\sigma^2$ and $\mathbf{C}_{\hat{\mathbf{z}}} \equiv \sigma^2\mathbf{I}_{\mathbf{N}}$ and it is useful to note, in passing, that this derivation is the

same as one in Koop *et al.* (1997) under the nomenclature 'marginal efficiency distribution.' To simulate a draw from the truncated normal distribution, two avenues are available. One is to draw from the normal distribution, accepting the draw if it satisfies the constraint requirements. The second method (advocated, for example, by Geweke (1992), and by Chib (1992), and employed currently) is to construct efficient one-for-one draws by modifying a draw from the standard uniform distribution through the probability integral transform (see, for example, Mood *et al.*, 1974, p. 202). In the interests of completeness, a draw from the standard normal distribution, say v (a scalar), truncated below by $\underline{a}$ and from above by $\underline{b}$ is simulated by transforming a draw from the standard uniform distribution $x \sim f^U(x|0,1)$ as $v = \Phi^{-1}(x((\Phi(b) - \Phi(a)) + \Phi(a))$, where $\Phi(\cdot)$ denotes the cumulative distribution function (cdf) corresponding to the standard normal distribution and $\Phi^{-1}(\cdot)$ denotes its inverse.

Finally, we outline the draw for the location parameter underlying the inefficiency distribution. Focusing in (3.5) on λ , the normal density drops out and we are left with

$$\pi(\lambda|\mathbf{z}) \propto \lambda^N \exp\{-\lambda \boldsymbol{\iota}_N' \mathbf{z}\}. \tag{3.10}$$

This density is in the form of a Gamma distribution and is specified, for example, in Zellner (equation (A.30), p. 369) with parameters $\alpha \equiv N + 1$ and $\gamma \equiv \boldsymbol{\iota}_N' \mathbf{z}$ with $\boldsymbol{\iota}_N$ the N-dimensional unit vector. Consequently, $\pi(\lambda|\mathbf{z})$ is the Gamma distribution and we write $\pi(\lambda|\mathbf{z}) \equiv f^G(\lambda|\alpha,\gamma)$. It follows that one additional step is to draw a pseudo-random number from the Gamma distribution, for which a number of alternative algorithms exist. In short, robust estimates of the exponential inefficiencies, composed error model are obtained by sampling iteratively in the algorithm:

- Step 1: Simulate a draw $\sigma^{(s)}$ from (3.7).
- Step 2: Simulate a draw $\boldsymbol{\beta}^{(s)}$ from (3.8).
- Step 3: Simulate a draw $\mathbf{z}^{(s)}$ from (3.9).
- Step 4: Simulate a draw $\lambda^{(s)}$ from (3.10).

Once again, although it is not emphasized in the notation, the conditioning in each draw is on the previous values of the model parameters $\boldsymbol{\theta}^{(s)} \equiv (\sigma^{(s)}, \boldsymbol{\beta}^{(s)\prime}, \lambda^{(s)})'$ and the latent data $\mathbf{z}^{(s)}$. Sampling iteratively from Steps 1-4 until the draws are independent of the starting values, say $\boldsymbol{\beta}^{(0)}$, $\mathbf{z}^{(0)}$ and $\lambda^{(0)}$; we continue to sample $g = 1, 2, \ldots, G$; collect outputs $\{\boldsymbol{\theta}^{(g)}\}_{g=1}^G$ and $\{\mathbf{z}^{(g)}\}_{g=1}^G$; and conduct posterior inference. This procedure is simple; relies only on a basic understanding of random-number generation; requires modest amounts of computer memory in order to collect samples of reasonable size (say, G = 5, 000); and leads to robust estimates of parameters of the composed-error model.

3.3 Composed-error models comparison

Bayesian model comparison relies on the estimation of Bayes factors (see Berger (1985) or Berger and Perrichi (1996) for discussion). The essential inputs in Bayes factors calculations are the marginal likelihoods of competing models. The marginal likelihood, denoted $m(\mathbf{y})$, is nothing other than the integrating constant that allows us to write equation (3.5) as a strict equality instead of two quantities that are proportional to each other. Until recently, the computation of marginal likelihoods has proved extremely troublesome for all but the simplest of models. The innovation that paves the way for general application is Chib (1995). Briefly, the marginal likelihood

$$m(\mathbf{y}) = \int f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}, \tag{3.11}$$

can be written, alternatively, as the product of the sampling density and the prior for $\boldsymbol{\theta}$ divided by the posterior for $\boldsymbol{\theta}$,

$$m(\mathbf{y}) = \frac{f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta}|\mathbf{y})} \tag{3.12}$$

By exploiting this basic marginal likelihood identity we are able to derive robust estimates of $m(\mathbf{y})$. Once converted to the (computationally convenient) log scale, an estimate of the log of the marginal likelihood, say, $\ln \hat{m}(\mathbf{y})$ is available as

$$\ln \hat{m}(\mathbf{y}) = \ln f(\mathbf{y}|\boldsymbol{\theta}^*) + \ln \pi(\boldsymbol{\theta}^*) - \ln \hat{\pi}(\boldsymbol{\theta}^*|\mathbf{y}), \tag{3.13}$$

where $f(\mathbf{y}|\boldsymbol{\theta}^*)$, $\pi(\boldsymbol{\theta}^*)$ and $\pi(\boldsymbol{\theta}^*|\mathbf{y})$ denote estimates of, respectively, the sampling density, the prior ordinate and the posterior ordinate at the point $\boldsymbol{\theta} = \boldsymbol{\theta}^*$. In the context of our formulation, $\boldsymbol{\theta}^* \equiv (\sigma^*, \boldsymbol{\beta}^{*\prime}, \lambda^*)$ and $\boldsymbol{\beta}^* \equiv (\beta_1^*, \ldots, \beta_K^*)'$. It is important to note the simplicity of this relationship. Because equation (3.13) holds at any point in the parameter space, all that is required in order to evaluate it are three quantities. The sampling density for $\mathbf{y}$ is not available directly from equation (3.3) because we have, in effect, used $f(\mathbf{y}|\boldsymbol{\theta}^*, \mathbf{z})$ when we require $f(\mathbf{y}|\boldsymbol{\theta}^*) = \int f(\mathbf{y}|\boldsymbol{\theta}^*, \mathbf{z})d\mathbf{z}$. Nevertheless, an estimate is available from exploiting the Rao-Blackwell method (Gelfand and Smith, 1990), drawing estimates of the sampling density conditional for the draws for the latent data; in other words,

$$\hat{f(\mathbf{y}|\boldsymbol{\theta}^*)} \equiv G^{-1} \sum_{g=1}^{G} f(\mathbf{y}|\boldsymbol{\theta}^*, \mathbf{z}^{(g)}). \tag{3.14}$$

The prior probability densities are available directly once specified and the ones that we employ are, respectively, $\pi(\sigma) \equiv f^{IG}(\sigma|\nu_0, s_0^2)$, $\nu_0 = 1$,

$s_0^2 = 0.001$; $\pi(\boldsymbol{\beta}) \equiv f^N(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}_0, \mathbf{C}_{\hat{\boldsymbol{\beta}}_0})$; where $\hat{\boldsymbol{\beta}}_0 = \mathbf{0_K}$ and $\mathbf{C}_{\hat{\boldsymbol{\beta}}_0} = \mathbf{I_K} \times 1000$; and $\pi(\lambda) \equiv f^E(\lambda|\delta_0), \delta_0 = 0.001$. These are the natural conjugate prior distributions, respectively, for the inverted-Gamma, Normal and Exponential distributions and they reflect weak prior information about the model parameters. In terms of the steps in equations (3.7) through (3.10), they imply minor changes to the parameters defining the component conditional distributions. Specifically, in $f^{IG}(\sigma|\nu, s^2), \nu \equiv N+\nu_0$ and $\nu s^2 \equiv (y-\mathbf{x}\boldsymbol{\beta}+\mathbf{z})'(y-\mathbf{x}\boldsymbol{\beta}+\mathbf{z})+\nu_0 s_0^2$; in $f^{MN}(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}, \mathbf{C}_{\hat{\boldsymbol{\beta}}})$, $\hat{\boldsymbol{\beta}} \equiv \mathbf{C}_{\hat{\boldsymbol{\beta}}}(\sigma^{-2}\mathbf{x}'\mathbf{a} + C^{-1}_{\hat{\boldsymbol{\beta}}_0}\hat{\boldsymbol{\beta}}_0)$ and $\mathbf{C}_{\hat{\boldsymbol{\beta}}} \equiv (\sigma^{-2}\mathbf{x}'\mathbf{x} + \mathbf{C}^{-1}_{\hat{\boldsymbol{\beta}}_0})$; and in $f^G(\lambda|\alpha, \gamma), \alpha \equiv N+1$ and $\gamma \equiv \iota_N'\mathbf{z} + \delta_0$.

An estimate of the posterior ordinate is available from decomposing the posterior as $\pi(\boldsymbol{\theta}^*|\mathbf{y}) = \pi(\sigma^*|\boldsymbol{\beta}^*, \lambda^*, \mathbf{y}) \times \pi(\boldsymbol{\beta}^*|\lambda^*, \mathbf{y}) \times \pi(\lambda^*|\mathbf{y})$ and the estimation of the posterior quantity involves three steps. Recalling that the fully conditional distribution for λ is independent of σ, $\boldsymbol{\beta}$ and also $\mathbf{y}$, but is dependent on $\mathbf{z}$; the first step involves averaging over the draws in the original Gibbs run at the point $\lambda = \lambda^*$ to obtain the estimate $\hat{\pi}(\lambda^*) = G^{-1}\sum_{g=1}^G \pi(\lambda^*|\mathbf{z}^{(g)})$. Second, recalling that $\pi(\boldsymbol{\beta}|\sigma, \mathbf{z}, \mathbf{y})$, is independent of λ, we compute $\hat{\pi}(\boldsymbol{\beta}^*|\mathbf{y}) = G^{-1}\sum_{g=1}^G \pi(\boldsymbol{\beta}^*|\sigma^{(g)}, \mathbf{z}^{(g)}, \mathbf{y})$ at the point $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ from the original Gibbs run. Third, recalling that $\pi(\sigma|\boldsymbol{\beta}, \mathbf{y})$ is also independent of λ, we set $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ and run the algorithm for one additional run obtaining the estimate $\hat{\pi}(\sigma^*|\boldsymbol{\beta}^*, \mathbf{y}) = G^{-1}\sum_{g=1}^G \pi(\sigma^*|\boldsymbol{\beta}^*, \mathbf{z}^{(g)}, \mathbf{y})$. At the end of this one additional run each of the components of the marginal likelihood is available and posterior inference about model comparison beckons.

Several points are noteworthy. First, as emphasized in Chib (1995), this computation is attractive because all that is required are some modest amendments to the basic Gibbs algorithm. Second, although any point $\boldsymbol{\theta}^*$ will suffice for computation, efficiency arguments motivate choosing $\boldsymbol{\theta}^*$ at a high-density point (for example, the maximum-likelihood point). Third, a measure of the variability of the estimate in repeated implementation (with the same data) is available from exploiting a result in Newey and West (1987) and the delta method for deriving estimates of the variance of non-linear functions. Specifically, defining

$$\mathbf{h}^{(g)} \equiv \begin{pmatrix} h_1(\mathbf{z}) \\ h_2(\mathbf{z}) \\ h_3(\sigma,\ \mathbf{z}) \\ h_4(\ \mathbf{z}) \end{pmatrix} = \begin{pmatrix} \pi(\mathbf{y}|\boldsymbol{\theta}^*, \mathbf{z}^{(g)}) \\ \pi(\lambda^*|\mathbf{z}^{(g)}) \\ \pi(\boldsymbol{\beta}^*|\sigma^{(g)}, \mathbf{z}^{(g)}, \mathbf{y}) \\ \pi(\sigma^*|\boldsymbol{\beta}^*, \mathbf{z}^{(g)}, \mathbf{y}) \end{pmatrix} \tag{3.15}$$

where $\pi(\mathbf{y}|\boldsymbol{\theta}^*, \mathbf{z}) \equiv f(\mathbf{y}|\boldsymbol{\theta}^*, \mathbf{z})$ denotes the complete-data density, $\hat{\mathbf{h}} \equiv G^{-1}\sum_{g=1}^G \mathbf{h}^{(g)}$ and $\Omega_s \equiv G^{-1}\sum_{g=s+1}^G (\mathbf{h}^{(g)} - \hat{\mathbf{h}})(\mathbf{h}^{(g-s)} - \hat{\mathbf{h}})'$; the variance

of the vector $\hat{\mathbf{h}}$ is

$$var(\hat{\mathbf{h}}) = G^{-1} \sum_{s=0}^{q} \left(1 - \frac{s}{q+1}\right) (\Omega_s + \Omega_s'), \tag{3.16}$$

where q denotes the value at which the autocorrelation function of the Gibbs sample tapers off (Chib, 1995, p. 316). Using the delta method, an estimate of the numerical standard error of the marginal likelihood on the logarithmic scale is

$$se \ln(\hat{\mathbf{h}}) = \sqrt{\mathbf{h}^{-1\prime} var(\hat{\mathbf{h}}) \mathbf{h}^{-1}}, \tag{3.17}$$

where $\mathbf{h}^{-1} \equiv (h_1^{-1}, h_2^{-1}, h_3^{-1}, h_4^{-1})'$. In short, a complete set of composed-error model estimates is available from combining ideas in Gelfand and Smith (1990), Koop *et al.* (2001) and Chib (1995).

3.4 Truncated-normal, panel data estimation

The cross-sectional exponential formulation has served its purpose. Two major limitations preclude using it in empirical work. First, and most importantly, the model is actually under-identified. Second, because of its simplicity, the exponential distribution places undue restriction on the pattern of inefficiency, motivating search for more flexible formulations. A simple counting procedure motivates the intuition about why the cross-sectional model is under-identified. There are, in total, $N+3$ parameters being estimated, namely the N inefficiency measures $z_1, z_2 \ldots, z_N$; the two parameters characterizing the sampling distribution, namely σ and $\boldsymbol{\beta}$; and the parameter λ characterizing the inefficiency distribution. Yet, only $i = 1, 2 \ldots, N$ sample points avail themselves for estimating these $N + 3$ unknowns. This point is taken up in Fernàndez *et al.* (1997), where a simple remedy is also suggested. As long as we have panel data and the inefficiency terms are held constant across subsets of the data (usually its time dimension), the complete set of model parameters is fully identified. The second problem concerns the actual shape of the inefficiency distribution across the firms. The exponential distribution is a single-parameter distribution, offering considerable tractability, and, while highly useful for expositional purposes, it severely restricts the underlying inefficiencies. The form on which we focus attention in the remainder of the paper is the truncated-normal distribution. In the panel with $j = 1, \ldots, N_i$ subunits (say, periods) corresponding to each of the $i = 1, \ldots, N$ sample units (say, firms), we consider (firm) inefficiencies $z_1, z_2 \ldots, z_N \sim f^{TN}(z_i|\lambda, \gamma)$, constant across time periods; and sampling errors $u_{ij} \sim f^{N}(u_{ij}|0, \sigma)$ for all $j = 1, 2 \ldots, N_i$, $i = 1, 2 \ldots, N$ and the availability of panel data has two implications. It circumvents the under-identification issue and it affords great tractability permitting multiple linkages across sub-units and across time peri-

ods. In the empirical application such flexibility, we will show, is extremely important. In short, panel data raises scope for nuanced empirical inquiry.

With a diffuse prior on the parameters $\boldsymbol{\theta} \equiv (\sigma, \boldsymbol{\beta}', \lambda, \gamma)'$ the posterior has the form

$$\pi(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z}) \propto \sigma^{-1}\gamma^{-1} \prod_{i=1}^{N} f^{N}(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta} - \iota_i z_i, \sigma^2 \mathbf{I}_{Ni}) f^{TN}(z_i|\lambda, \gamma) \qquad (3.18)$$

which should be compared to (3.5). The differences are important. First, y_i is an $N_i \times 1$ series of observations on firm i's production; $\mathbf{x}_i$ denotes an $N_i \times K$ matrix of covariates defining the frontier; $\boldsymbol{\beta}$ remains as before; ι_i is an N_i vector of ones; and I_{Ni} denotes the N_i-dimensional identity matrix. The total number of observations in the sample is $S = \Sigma_i N_i$ and the counting problem is usually overcome because the $S = \Sigma_i N_i$ data points are now estimating $N + K + 3$ quantities, namely $\mathbf{z} \equiv (z_1, \ldots, z_N)'$, $\boldsymbol{\beta} \equiv (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K)', \sigma, \lambda$ and γ.

Despite the added flexibility of the truncated-normal framework, the Gibbs algorithm follows closely the steps that we derived for the exponential model and each of the fully conditional distributions is one that we have already encountered. The regression standard error has an inverse-Gamma distribution $\pi(\sigma|\boldsymbol{\beta}, \mathbf{z}, \mathbf{y}) \equiv f^{iG}(\sigma|v, s^2), v \equiv N, vs^2 \equiv (\mathbf{y} - \mathbf{x}\boldsymbol{\beta} + \mathbf{wz})'(\mathbf{y} - \mathbf{x}\boldsymbol{\beta} + \mathbf{wz}), \mathbf{y} \equiv (y_1', y_2', \ldots, y_N)', \mathbf{x} \equiv (\mathbf{x}_1', \mathbf{x}_2', \ldots, \mathbf{x}_N')'$ and $\mathbf{w} \equiv diag(\iota_1, \iota_2, \ldots, \iota_N)$, of dimension $S \times N$, denotes an appropriate arrangement of the unit vectors corresponding to each element of $\mathbf{z}$ in the regression $\mathbf{y} = \mathbf{x}\boldsymbol{\beta} - \mathbf{wz} + \mathbf{u}$ where $\mathbf{u} \equiv (u_1', u_2', \ldots, u_{N'})'$, The frontier parameters have the multivariate normal distribution $\pi(\boldsymbol{\beta}|\sigma, \mathbf{z}, \mathbf{y}) \equiv f^{N}(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}, \mathbf{C}_{\hat{\boldsymbol{\beta}}}), \hat{\boldsymbol{\beta}} \equiv \mathbf{C}_{\hat{\boldsymbol{\beta}}}^{-1}\sigma^{-2}\mathbf{x}'(\mathbf{y} + \mathbf{wz})$, $\mathbf{C}_{\hat{\boldsymbol{\beta}}} \equiv \sigma^2(\mathbf{x}'\mathbf{x})^{-1}$. The vector of inefficiency terms is multivariate-Normal with truncation at zero, $\pi(\mathbf{z}|\sigma, \boldsymbol{\beta}, \gamma, \mathbf{z}, \mathbf{y}) \equiv f^{TN}(\mathbf{z}|\hat{\mathbf{z}}, \mathbf{C}_{\hat{\mathbf{z}}}), \hat{\mathbf{z}} \equiv \mathbf{C}_{\hat{\mathbf{z}}}^{-1}(\mathbf{ch}\lambda - \mathbf{w}'\mathbf{ba})$, $\mathbf{C}_{\hat{\mathbf{z}}}(\mathbf{w}'\mathbf{bw} + \mathbf{c}), \mathbf{a} \equiv \mathbf{y} - \mathbf{x}\boldsymbol{\beta}, \mathbf{b} \equiv \sigma^{-2}I_s, \mathbf{c} \equiv \gamma^{-2}I_N$ and $\mathbf{h}$ is the N-dimensional unit vector. The inefficiency standard deviation has an inverse-Gamma distribution $\pi(\gamma|\lambda, \mathbf{z}) \equiv f^{IG}(\gamma|\rho, t^2), \rho \equiv N, pt^2 \equiv (\mathbf{z} - \mathbf{h}\lambda)'(\mathbf{z} - \mathbf{h}\lambda)$ and the mean inefficiency has a truncated-normal distribution

$$\pi(\lambda|\mathbf{z}) \equiv f^{TN}(\lambda|\hat{\lambda}, v_{\hat{\lambda}}), \hat{\lambda} \equiv (\mathbf{h}'\mathbf{h})^{-1}\mathbf{h}'\mathbf{z}, v_{\hat{\lambda}} \equiv \gamma^2(\mathbf{h}'\mathbf{h})^{-1}.$$

A single pass through these five distributions simulates a draw from the joint posterior; is easily implemented using available software; and is readily extended to accommodate marginal likelihood calculations. In fact, the truncated-normal, composed-error model is almost fully operational once the simpler, exponential specification is in place.

3.5 Estimation in a two-layer hierarchy

One advantage of the truncated-normal set-up – the subject of this section – is its ability to interlay additional levels of plausible empirical distributions. When data arrive in multidimensional forms (such as those in the example that follows) it is natural to consider implanting additional layers in the hierarchy. These notions, of course, have more than academic appeal because, as we have already argued, they enhance the robustness of model estimation. Turning to a two-layer extension of the single-level set-up, almost all of the developments are identical to the ones above, and we can be terse with the presentation.

Suppose, in the panel, with $k = 1, 2, \ldots, N_{ij}$ production 'cycles' in each of $j = 1, 2, \ldots, N_i$ time periods across $i = 1, 2, \ldots, N$ firms; inefficiencies z_{ij}, $j = 1, 2, \ldots, N_i$, $i = 1, 2, \ldots, N$ evolve according to a two-layer chain $z_{ij} \sim f^{TN}(z_{ij}|\alpha_i, \omega)$, $\alpha_i \sim f^{TN}(\alpha_i|\lambda, \gamma)$; and suppose, in addition, that sampling error follows $u_{ijk} \sim f^{N}(u_{ijk}|0, \sigma)$. In this model inefficiencies are permitted to vary across time periods and across firms but are constrained to be constant over the intra-firm, intra-period production cycles. The sampling errors are, once again, assumed to be *iid* Normal. A notion that has much importance and one to which we return is conceptualising the i, the j and the k.

In the fisheries empirical application a natural interpretation of the sequence i-j-k is that i denotes a boat, j denotes a season and k denotes a voyage. In other words, a boat-season-voyage interpretation of i-j-k seems natural and observed quantity y_{ijk} thus denotes the quantity of catch by boat i in period j in which voyage k was undertaken. But a season-boat-voyage interpretation may be equally plausible. Moreover, such an interpretation may have quite significant implications for the way in which we process the posterior information and, in particular, the pattern of inefficiencies that are estimated across the sample. This issue—choosing a preferred specification of the hierarchical chain—is important and can be reconciled empirically as part of a model selection exercise. Placing a diffuse prior on the model parameters $\boldsymbol{\theta} \equiv (\sigma, \boldsymbol{\beta}', \omega, \gamma, \lambda)'$ the posterior distribution has the form

$$\pi(\boldsymbol{\theta}|\mathbf{y}, \alpha, \mathbf{z}) \propto \frac{1}{(\sigma\omega\gamma)^{-1}} \prod_{i=1}^{N}\prod_{j=1}^{N} f^{N}(\mathbf{y}_{ij}|\mathbf{x}_{ij}\boldsymbol{\beta} - \boldsymbol{\iota}_{ij}z_{ij}, \sigma^2\mathbf{I}_{Nij}) \times f^{TN}(z_{ij}|\alpha_i, \omega) f_{TN}(\alpha_i|\lambda, \gamma), \tag{3.19}$$

which should be compared to (3.18), above. Presently, $\mathbf{y}_{ij} \equiv (y_{ij1}, \ldots, y_{ijN_{ij}})'$ is an $N_{ij} \times 1$ series of observations on firm i's output in production 'cycles' occurring in period j; $\mathbf{x}_{ij} \equiv (\mathbf{x}'_{ij1}, \ldots, \mathbf{x}'_{ijNij})'$ denotes an $N_{ij} \times K$ matrix of covariates affecting the frontier; $\boldsymbol{\beta} \equiv (\beta_1, \ldots, \beta_K)'$ as before; ι_{ij} denotes an N_{ij} unit vector; and $\mathbf{I}_{Nij}$ denotes the N_{ij} -dimensional identity matrix. There are now $S = \Sigma_i\Sigma_j N_{ij}$ total observations in the sam-

ple. There are $P = \Sigma_i N_i$ latent quantities in the vector $\mathbf{z} \equiv (\mathbf{z}_1', \mathbf{z}_2', \ldots, \mathbf{z}_N')'$, $\mathbf{z}_1 \equiv (z_{11}, \ldots, z_{1N1})$, $\mathbf{z}_2 \equiv (z_{21}, \ldots, z_{2N2})', \ldots,$ $\mathbf{z}_N \equiv (z_{N1}, \ldots, z_{NNN})'$; there are N latent quantities in the vector $\boldsymbol{\alpha} \equiv (\alpha_1, \ldots, \alpha_N)'$; and there are now $K + 4$ model parameters in $\boldsymbol{\theta} \equiv (\sigma, \boldsymbol{\beta}', \omega, \lambda, \gamma)'$. Our objective is to extract samples $\{\boldsymbol{\theta}^{(g)}\}_{g=1}^G$, $\{\alpha^{(g)}\}_{g=1}^G$ and $\{\mathbf{z}^{(g)}\}_{g=1}^G$ and make inferences about posterior quantities of interest.

Once again, the conditional distributions are easy to sample. The regression standard-error has an inverse-Gamma distribution $\pi(\sigma|\boldsymbol{\beta}, \mathbf{z}, \mathbf{y}) \equiv f^{IG}(\sigma|\nu, s^2), \nu \equiv S, \nu s^2 \equiv (\mathbf{y} - \mathbf{x}\boldsymbol{\beta} + \mathbf{wz})'(\mathbf{y} - \mathbf{x}\boldsymbol{\beta} + \mathbf{wz}), \mathbf{y} \equiv (\mathbf{y}_1', \ldots, \mathbf{y}_N')'$, $\mathbf{y}_1 \equiv (\mathbf{y}_{11}', \ldots, \mathbf{y}_{1N1}')', \mathbf{y}_2 \equiv (\mathbf{y}_{21}', \ldots, \mathbf{y}_{2N2}')', \ldots, \mathbf{y}_N \equiv (\mathbf{y}_{N1}', \ldots, \mathbf{y}_{NNN}')'$; $\mathbf{x} \equiv (\mathbf{x}_1', \ldots, \mathbf{x}_N')'$, $\mathbf{x}_1 \equiv (\mathbf{x}_{11}', \ldots, \mathbf{x}_{1N1}')'$, $\mathbf{x}_2 \equiv (\mathbf{x}_{21}', \ldots, \mathbf{x}_{2N2}')'$, ..., $\mathbf{x}_N \equiv (\mathbf{x}_{N1}', \ldots, \mathbf{x}_{NNN}')'$; $\mathbf{w} \equiv diag(\mathbf{v}_1, \ldots, \mathbf{v}_N), \mathbf{v}_i \equiv diag(\boldsymbol{\iota}_{i1}, \boldsymbol{\iota}_{i2}..., \boldsymbol{\iota}_{iN})$, $\boldsymbol{\iota}_{ij}$ denotes anN_{ij}-dimensional unit vector and, thus, $\mathbf{w}$ denotes an appropriate arrangement of the unit vectors corresponding to the z_{ij}'s within the regression $\mathbf{y} = \mathbf{x}\boldsymbol{\beta} - \mathbf{wz} + \mathbf{u}$. The frontier parameters have the multivariate normal distribution $\pi(\boldsymbol{\beta}|\sigma, \mathbf{z}, \mathbf{y}) \equiv f^N(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}, \mathbf{C}_{\hat{\boldsymbol{\beta}}})$, $\hat{\boldsymbol{\beta}} \equiv (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'(\mathbf{y} + \mathbf{wz})$, $\mathbf{C}_{\hat{\boldsymbol{\beta}}} \equiv \sigma^2(\mathbf{x}'\mathbf{x})^{-1}$. The vector of inefficiency terms is multivariate-Normal with truncation at zero, $\pi(\mathbf{z}|\sigma, \boldsymbol{\beta}, \omega, \alpha, \mathbf{z}, \mathbf{y}) \equiv f^{TN}(\mathbf{z}|\hat{\mathbf{z}}, \mathbf{C}_{\hat{\mathbf{z}}})$, $\hat{\mathbf{z}} \equiv \mathbf{C}_{\hat{\mathbf{z}}}(\mathbf{cvz} - \mathbf{w}'\mathbf{ba})$, $\mathbf{C}_{\hat{\mathbf{z}}} \equiv (\mathbf{w}'\mathbf{bw} + \mathbf{c})^{-1}$, $\mathbf{a} \equiv \mathbf{y} - \mathbf{x}\boldsymbol{\beta}$, $\mathbf{b} \equiv \sigma^{-2}\mathbf{I_s}$, $\mathbf{c} \equiv \omega^{-2}\mathbf{I_p}$ and $\mathbf{v} \equiv diag(v_1, v_2, .., v_N)$ is the $P \times N$-dimensional binary matrix that is consistent with $\mathbf{z}$. The inefficiency standard deviation in the first level of the hierarchy has an inverse-Gamma distribution $\pi(\omega|\alpha, \mathbf{z}, \mathbf{y}) \equiv f^{IG}(\omega|\delta, q^2)$, $\delta \equiv P$, $\delta q^2 \equiv (\mathbf{z} - \mathbf{h}\boldsymbol{\alpha})'(\mathbf{z} - \mathbf{h}\boldsymbol{\alpha})$. In the upper level of the hierarchy the mean inefficiency has a truncated-Normal distribution $\pi(\boldsymbol{\alpha}|\omega, \gamma, \lambda, \mathbf{z}) \equiv f^{TN}(\boldsymbol{\alpha}|\hat{\boldsymbol{\alpha}}, \mathbf{C}_{\hat{\boldsymbol{\alpha}}})$, $\hat{\boldsymbol{\alpha}} \equiv \mathbf{C}_{\hat{\boldsymbol{\alpha}}}(\mathbf{v}'\mathbf{cz} + \mathbf{dh}\lambda)$, $\mathbf{C}_{\hat{\boldsymbol{\alpha}}} \equiv (\mathbf{vc}'\mathbf{v} + \mathbf{d})^{-1}$, $\mathbf{d} \equiv \gamma_N^{-2}$. The inefficiency standard deviation in the upper level of the hierarchy has an inverse-Gamma distribution $\pi(\gamma|\lambda, \boldsymbol{\alpha}) \equiv f^{IG}(\gamma|\rho, t^2)$, $\rho \equiv N$, $\rho t^2 \equiv (\boldsymbol{\alpha} - \mathbf{h}\lambda)'(\boldsymbol{\alpha} - \mathbf{h}\lambda)$. And the upper-level mean inefficiency has the truncated-Normal distribution $\pi(\lambda|\mathbf{z}) \equiv f^{TN}(\lambda|\hat{\lambda}, v_{\hat{\lambda}})$, $\hat{\lambda} \equiv (\mathbf{h}'\mathbf{h})^{-1}\mathbf{h}'\boldsymbol{\alpha}$, $v_{\hat{\lambda}} \equiv \gamma^2(\mathbf{h}'\mathbf{h})^{-1}$. One pass through these six distributions simulates a draw from the joint posterior and the algorithm can be applied to derive estimates of model parameters or marginal likelihoods. Thus, a tractable extension of the single-layer hierarchy avails itself to nuanced empirical inquiry.

4. Empirical evidence from the Pacific Coast Groundfish Fishery

In the 1960s and 1970s, the United States government undertook various measures to develop the domestic fishing fleet, including the exclusion of foreign vessels and investment incentives for US fishermen. Even as the number of boats in the domestic fleet was increasing, new technologies enabled boats to fish more efficiently. In the Pacific Coast groundfish fishery, by the

early 1980s concern was growing about the effect of the fleet on particular fish stocks. Since 2000, the fishery has operated in a climate of crisis. The fishery has been formally designated a disaster, most of the continental slope has been closed to trawling, and the federal government has administered a buyback program that retired approximately one third of the fishing fleet.

The groundfish fishery is managed by the Pacific Fisheries Management Council, composed of representatives of the states, tribes, industry groups, recreational fishermen, and the federal government. The Council's goals are to protect fish stocks while attempting to meet the needs of commercial, recreational, and tribal fishermen. The Council's groundfish regulatory measures initially consisted primarily of trip limits, but have since evolved into a complex combination of trip limits, monthly and bi-monthly limits, gear restrictions, area closures, and seasonal restrictions. Some of the more significant regulatory changes in recent years have been the adoption of small-footrope requirements, which discourage fishermen from fishing in areas where there are rocks on the ocean floor, and depth-based area restrictions. Each year the Council sets total allowable harvests for management complexes, i.e., groups of species in particular management zones. During the year, the Council adjusts regulations in an attempt to see that the total catch target is met but not exceeded.

Increasingly strict regulation has not averted persistent overfishing: nine species of groundfish are now designated overfished by the federal government. Coupled with overfishing is the problem of overcapacity, i.e., the fleet's ability to catch fish exceeds the total allowable catch. In a 2001 report, the Council's Scientific and Statistical Committee estimated that only 27-41% of the trawl vessels with groundfish permits would be needed to harvest the 2000 quotas, and asserted that overcapacity was the most pressing groundfish management problem. In 2000, the Secretary of Commerce designated the fishery a disaster, a prelude to the allocation of funds for job retraining, data collection, and buyouts of fishing licenses and boats. Then in 2003, the federal government oversaw a buyout of groundfish permits, with the dual goals of protecting fish stocks and restoring profitability to the fishery. A total of $46 million in financing from the industry and the federal government was used to purchase the fishing rights of boats in the limited-entry fishery. Those boats permanently relinquished their fishing permits and may never fish again anywhere in the world.

This brief description of the fishery suggests several ways in which the analysis of technical efficiency relates to important questions in fisheries management. In addition to giving managers an aggregate indication of the fleet's condition, efficiency measures can provide a comparison of efficiency among sub-groups (e.g., boat size or home port), which may cast light on equity issues such as the differential impacts of regulatory changes. Common fisheries

regulations, including both input and output restrictions, are likely to affect the technical efficiency (hence profitability) of fishing fleets. Because these changes also affect the likely future impact of the fleet on fish stocks, technical efficiency analysis can also inform fish stock projections. More generally, questions of the relative impact on fishing enterprises' viability of time-varying factors that affect all boats (ocean conditions and regulations being the most obvious) versus boat-specific factors (such as skipper skill) suggest that many policy issues might be illuminated by the application of hierarchical composed-error models as developed previously in this chapter.

5. Data

Groundfish trawlers with federal limited-entry permits are required to keep trip logs, including information on departure and return ports, location and duration of fishing, catch, crew, fuel consumption, and numbers of tows. These logbook data (maintained by the *Pacific States Marine Fisheries Commission*) are used to estimate truncated-normal stochastic production frontiers. We limit analysis to data from the state of Washington in order to keep the observational units fairly homogeneous; in order to avoid complications due to inter-state variation in regulations and data collection; and because Washington has the highest logbook reporting compliance rate of the three West Coast states. Further, because our primary focus is the way in which the hierarchical structures affect the results, we restrict attention to a panel that is, ostensibly, 'square, ' consisting of the thirteen most prolific vessels in operation in the thirteen year period 1987-1999. The resulting sample consists of 5,030 tows and the variance of this sample across the respective partitions (boats or years) is important in what follows. The mean number of years (boats) in which a boat (year) appears is 10.69 and the standard deviation of the tows partitioned across boats is 3.38 and partitioned across years is 1.38. The mean number of tows per boat (tows per year) is 386.92 and the standard deviation across boats is 104.7 and across years is 90.37. Thus, slight variations in deviations exist, but they are small in relation to the respective sample means.

In terms of our notation, the data are the catch (in pounds) for trips 'k' in years 'j' by vessels 'i' (namely, y_{ijk}) and the covariates (x_{ijk}) conjectured to condition catch levels. Among the covariates to which we have access, we adopt a fairly parsimonious specification consisting of just four, namely proxies for boats' capital services (horsepower), labor inputs (number of crew per trip), effort intensity (total duration of tows), and a measure designed to reflect possible differences in tow-duration technology between trips (the total number of tows undertaken). Prior to estimation, each of the variables is transformed to natural logarithms and, thus, technological possibilities are constrained to be Cobb-Douglas.

6. Results

Table 9.1 reports results of respective applications to the groundfish data. The first two columns of entries are the estimates of the truncated-normal model with single-layer hierarchies (Fishery-Boats (FB) and Fishery-Years (FY), respectively) and the third and fourth columns report estimates with two-layer hierarchies (Fishery-Boats-Years (FBY) and Fishery-Years-Boats (FYB), respectively). Generally speaking, the estimates across the four formulations are quite similar and one particularly consistent feature across specifications is the signs of the coefficient estimates. The output elasticity of a change in horsepower ranges from a low of 1.45 (FB design) to a high of 1.93 (FBY design) and is, perhaps, the most homogeneous of the estimates. The output elasticities of the remaining covariates (number of crew, duration of tows and number of tows) are positive but variable across specifications. Important outputs of

Table 9.1. Model Estimates

	Hierarchical Designs			
	Fishery Boats	*Fishery Years*	*Fishery Boats Years*	*Fishery Years Boats*
$\beta_{\text{horse-power}}$	1.45	1.22	1.93	1.66
	(1.40, 1.52)	(1.17, 1.30)	(1.73, 2.12)	(1.50, 1.91)
$\beta_{\text{number-of-crew}}$	0.03	0.49	0.14	0.25
	(–0.12, 0.18)	(0.36, 0.60)	(–0.10, 0.39)	(0.01, 0.47)
$\beta_{\text{total-duration-of-tows}}$	0.18	0.44	0.11	0.11
	(0.10, 0.26)	(0.37, 0.51)	(0.01, 0.18)	(0.02, 0.20)
$\beta_{\text{total-number-of-tows}}$	0.67	0.58	0.81	0.81
	(0.58, 0.76)	(0.51, 0.65)	(0.71, 0.91)	(0.71, 0.91)
σ	0.96	1.01	0.87	0.87
	(0.94, 0.98)	(1.03, 1.06)	(0.85, 0.89)	(0.85, 0.90)
γ	0.61	0.24	0.55	0.06
	(0.37, 1.10)	(0.14, 0.44)	(0.30, 1.03)	(0.01, 0.24)
λ	1	0.55	3.90	2.36
	(0.47, 1.57)	(0.26, 0.95)	(2.62, 5.04)	(1.47, 3.73)
ω			0.5	0.66
			(0.41, 0.60)	(0.55, 0.79)
Mean eff. score	0.90	0.94	0.69	0.79
Std err.	0.06	0.03	0.06	0.07
Log max. like.	–6971.53	–6945.45	–6763.88	–6763.89
Log marg. like.	–7007.59	–6981.25	–6813.19	–6812.74
nse.	0.0033	0.002	0.0174	0.0113

Note: Numbers in brackets are 99 percentile highest posterior density regions.

the estimation are the *percent efficiency scores*. These scores are derived from the well-known Farrell measures of efficiency and are computed as the ratio of the predicted value of the regression to the maximum predicted value as in Jondrow *et al.* (1982). Significantly, the mean efficiency scores depend *fundamentally* on the nature of the hierarchical structure. The estimated percent efficiency scores under the single layer specifications (0.90 under FB and 0.94 under FY, respectively) decline dramatically under the two-layer specifications (to 0.69 under FBY and 0.79 under FYB, respectively) and suggest that the mean efficiency level predicted for the fishery is highly sensitive to the form of the hierarchical structure. From a model-selection perspective, an interesting informal comparison is the way in which the estimated noise—evidenced by estimates of σ—varies across the respective formulations. It is significantly higher in the single-layer specifications compared to the two-layer models. To the extent that this reduction in noise is desirable, it appears, at least *prima facie*, that adding layers in the hierarchy significantly improves fit. Inevitably, it raises scope for a formal comparison of the various formulations. The maximized likelihood values reported at the base of the table, suggest that the two-layer formulations are, indeed, the dominant structure supported by the data. This conclusion is supported formally by comparing the marginal likelihood values of each of the models, where the accuracy of the respective estimates—indicated by the numerical standard error (nse) values in the last line of the table—is particularly noteworthy. Again, the two-level specifications are overwhelmingly dominant. Under the prior, the data marginally support the Fishery-Year-Boats formulation over the Fishery-Boats-Years formulation, with estimated odds of 1.57 in favour of the former. The odds favouring both of the two-layer specifications over either of the single-layer formulations are in the billions. This finding makes clear that, among the candidate models that we implement, the preferred specification is a two-layer hierarchy with inefficiency levels held constant either across boats (within years) or across years (within vessels). Specifications that fail to account for this important distinction are likely to grossly overstate mean levels of efficiency among trawling operations. Moreover, as evidenced by the proximity of the *maximized* and *marginalized* likelihood values reported in the base of the table, these conclusions are fairly insensitive to the type of prior information adopted. In short, the two-layer designs afford greater flexibility than is present in the single-layer designs; are almost as easy to implement; and affect significantly posterior inferences about production efficiency in the coastal Washington fishery.

7. Conclusions and Extensions

Focusing mostly on pedagogic developments, this chapter has presented a fairly comprehensive guide to implementing Bayesian, hierarchical,

composed-error models. With reference to two of its simplest structures (the exponential and the truncated-normal forms) we outlined the essential inputs for posterior inference and model comparison. The Bayesian hierarchical approach, we argue, is a robust and flexible method of estimation that is insightful and intuitive and has the potential to enhance the current stock of simulation techniques available to natural-resource and environmental economists. Recent innovations and the continued availability of high-speed computing raise additional scope for application in the natural-resource and environmental-economics literatures. Extensions of the current effort pertain to distributional form, prior elicitation and a host of features too numerous to itemize. They are contained in a growing body of work including van den Broeck *et al.* (1994), Fernàndez *et al.* (1997, 2000, 2002a, forthcoming), Griffin and Steel (2004), and Koop and Steel (2001). In addition, readers will note that we have restricted attention to log-linear frontiers that remain linear in parameters. When the frontier is non-linear, the algorithm is easily extended to include a Metropolis step in place of the usual Gibbs draw. Net of model comparison, the remaining steps in the basic algorithm go through without modification, and model comparison can be undertaken by implanting extensions outlined in Chib and Jeliazkov (2001). Also, the hierarchical structures are restricted to the inefficiency alone, begging an obvious question: What if the production frontier itself is hierarchical?—a matter taken up in some detail by Tsionas (2002). That contribution, similar in spirit to this Chapter, is particularly relevant. It showcases the extreme versatility of the hierarchical approach in empirical inquiry, guiding heterogeneous, interrelated sample units to a common whole and, ultimately, a model selection exercise using Chib's (1995) computations. Finally, a 2005 issue of the *Journal of Econometrics* (see Dorfman and Koop, 2005 for an overview) highlights many recent composed-error model developments, which natural-resource and environmental economists should find appealing.

Chapter 10

BAYESIAN APPROACHES TO MODELING STATED PREFERENCE DATA

David F. Layton
Daniel J. Evans School of Public Affairs
University of Washington
Box 353055 Seattle,
WA 98195-3055, U.S.A.
dflayton@u.washington.edu

Richard A. Levine
Department of Mathematics and Statistics
San Diego State University
5500 Campanile Drive San Diego
CA 92182, U.S.A.
ralevine@sciences.sdsu.edu

Abstract Bayesian econometric approaches to modeling non-market valuation data have not often been applied, but they offer a number of potential advantages. Bayesian models incorporate prior information often available in the form of past studies or pre-tests in Stated Preference (SP) based valuation studies; model computations are easily and efficiently performed within an intuitively constructed Markov chain Monte Carlo framework; and asymptotic approximations, unreasonable for the relatively small sample sizes seen in some SP data sets, need not be invoked to draw (posterior) inferences. With these issues in mind, we illustrate computationally feasible approaches for fitting a series of surveys in a sequential manner, and for comparing a variety of models within the Bayesian paradigm. We apply these approaches to a series of SP surveys that examined policies to conserve old growth forests, northern spotted owls, and salmon in the U.S. Pacific Northwest.

Keywords: Stated Preference, Bayesian, Marginal Likelihood, Owls, Salmon

1. Introduction

Bayesian econometric approaches to modeling non-market valuation data have not often been applied. A quick literature search turns up just a few papers including: León and León (2003) who model double bounded contingent valuation data, León *et al.* (2002) who consider benefits transfer from a Bayesian perspective, Layton and Levine (2003) who model stated preference (SP) data with complex response patterns, and Train (2003) who develops a Bayesian based algorithm for estimating mixed-logit models. There is a longer tradition in the marketing literature (e.g. Huber and Train (2001)). Applications in other areas of environmental economics are infrequent as well. Notable examples include Fernandez *et al.* (2002) who modeled fisheries data and Koop and Tole (2004) who considered mortality due to air pollution.

Bayesian approaches offer a number of advantages relative to the standard frequentist approach (e.g. maximum likelihood) for use in non-market valuation. First, with a wealth of prior studies, we often have useful prior information that can be used within the Bayesian approach to obtain more precise estimates of Willingness to Pay (*WTP*), or similar precision with a more economical sample size. This is simply not possible within the standard frequentist approach employed in nearly all non-market valuation studies. Further, even without prior studies to draw upon, pre-tests, which are almost always used in survey based approaches, might be used as priors. This can be considered even if the final survey instrument has been heavily edited in response to pre-test results (results here might mean actual estimates from pre-tests data, or focus group responses, peer reviewer comments, etc.). From a maximum likelihood perspective, researchers are forced to choose between including the pre-test data in the final analysis as if it were generated from an identical survey instrument (perhaps with accounting for differential variance) or omitting it. The Bayesian approach maps neatly into how we imagine most researchers view their pre-tests and final surveys – similar but different. Another issue raised by Train (2003), Huber and Train (2001), and in Layton and Levine (2003) is that sometimes the Bayesian approach based on Markov chain Monte Carlo methods may be computationally simpler than estimation by maximum likelihood. As suggested by the benefits-transfer application in León *et al.* (2002), Bayesian approaches offer alternative methods for utilizing the information in data generated by different collection mechanisms. Finally, the Bayesian approach does not require the researcher to invoke large sample results, e.g. asymptotic normality of the sampling distribution of the estimated coefficients in order to discuss the precision of the results. Many data sets are rather small, or small relative to the complexity of the models estimated, to reliably invoke asymptotic properties.

With all of the potential advantages, why the paucity of applications? One reason may be legitimate differences in statistical philosophy that makes the Bayesian approach unattractive to researchers. This no doubt occurs, but more importantly is the fact that even when Bayesian approaches afford some computational advantages, they still come with computational difficulties when applied to discrete choice data common in non-market valuation.

We illustrate computationally feasible approaches for handling a series of SP data sets that all relate to a common set of attributes and policy issues. We begin with a noninformative prior for the first SP set and then estimate the model. Using the results to form a prior for the next survey in the sequence we then find the next set of results and so on. This allows us to use the information contained in a series of SP data sets without having to resort to overly strong pooling assumptions. Another issue we tackle is how to estimate Bayes factors for model comparison. For discrete choice models this is extremely challenging given the need to compute the marginal likelihood for each model. For a small subset of models we illustrate how this can be accomplished.

In section 2 we outline our Bayesian econometric approach for estimating multinomial probit models and describe necessary algorithms. In section 3 we describe a series of SP surveys that examined policies to conserve old growth forests, northern spotted owls, and salmon in the Pacific Northwest. The data are ideal for this study as each survey data set is too small to reliably invoke asymptotic results, yet taken together we illustrate how one survey can be used as a prior for another, and how we base our inferences on the small sample posterior distribution of *WTP*. Section 4 presents our results. Section 5 concludes with a discussion of the implications of the results and interesting directions for future research.

2. The Econometric Model and Bayesian Estimation Approach

We follow the standard random utility formulation (McFadden, 1981) in that we relate the attributes of the alternatives to utilities that the subject perceives. Based on the subject's utilities, the subject makes some kind of choice or decision that only partially reveals their utilities to us, the researchers. In specifying the model we find it useful to state the model in what might appear to be somewhat non-standard notation for those accustomed to maximum likelihood estimation. Still, the basic ingredients of the model specification are identical to what has become known as the random parameters approach to discrete choice modeling. Our case is the Bayesian analog of a multinomial probit model with normally distributed random parameters first estimated via maximum likelihood by Hausman and Wise (1978) and first estimated in its

Bayesian form by McCulloch and Rossi (1994) and Geweke *et al.* (1994). Our model set-up and notation follow Layton and Levine (2003).[1]

2.1 Model Specification

Let U_{ij} denote the utility of alternative j for individual i. We denote the $J-1$ vectors of utilities over all J alternatives for individual i by U_i. We relate these utilities to the covariates of interest through the linear mixed model (see Searle, Casella, and McCulloch, 1992, Chapter 9). We will use boldface to indicate matrix quantities.

$$\begin{aligned} U_i &= \mathbf{X_i}\,\beta + \mathbf{R_i}\,\gamma_i + \varepsilon_i \\ \varepsilon_i &\sim N_J\,(0,\,\mathbf{\Sigma}) \end{aligned} \tag{2.1}$$

where $\mathbf{X_i}$ denotes a $J \times K$ design matrix over the "fixed" effects, $\mathbf{R_i}$ denotes a $J \times L$ design matrix over the "random" effects, β is a K-vector of coefficients that are common to all subjects and γ_i is an L-vector of subject specific coefficients. ε_i has a J dimensional normal distribution with a zero mean vector and dispersion matrix $\mathbf{\Sigma}$ (covariance matrix). In some, but not all cases, L will equal K. In non-market valuation, the most common specification is that $\mathbf{R_i}$ is a subset of $\mathbf{X_i}$, where all variables in $\mathbf{X_i}$ except for the price or cost variable are present in $\mathbf{R_i}$ so that $L = K - 1$.

We apply a hierarchical Bayesian model by specifying a hierarchy of prior distributions for the unknown parameters $(\beta, \gamma_i, \Sigma)$. In particular, we assume the following prior distributions.

$$\begin{aligned} \beta &\sim N_K\,\left(\bar{\beta},\,\mathbf{A}\right) \\ \gamma_i &\sim N_L\,(0,\,\mathbf{\Sigma_\gamma}) \\ \mathbf{\Sigma_\gamma^{-1}} &\sim Wishart\,(\nu_\gamma,\,\mathbf{V_\gamma})\,. \end{aligned} \tag{2.2}$$

The mean is a known K-vector $\bar{\beta}$, and $\mathbf{A}$ is a known $K \times K$dispersion matrix (covariance matrix) for the normal prior on β. γ_i is assumed normally distributed with mean zero and dispersion matrix $\mathbf{\Sigma_\gamma}$. The precision matrix, $\mathbf{\Sigma_\gamma^{-1}}$is specified as having a Wishart distribution with known parameters ν_γ and $\mathbf{V_\gamma}$ (see McCulloch and Rossi (1994)). The fact that γ_i is specified as being normally distributed and depending upon other deeper parameters that themselves are random variables, the $\mathbf{\Sigma_\gamma}$, leads to the term "hierarchical". It would be possible, but computationally demanding, to add still further layers to the hierarchy, but we do not pursue this here as it is doubtful that additional hierarchical levels will add much to our fit or interpretation. The Wishart prior

[1]For an excellent development of Bayesian Mixed (random parameter) Logit models see Train (2003).

on $\mathbf{H} = \mathbf{\Sigma}_{\boldsymbol{\gamma}}^{-1}$ has density:

$$p(\mathbf{H}|\nu_\gamma, \mathbf{V}_{\boldsymbol{\gamma}}) \propto |\mathbf{H}|^{(\nu_\gamma - L - 1)/2} \exp\left[tr(-0.5\mathbf{H}\mathbf{V}_{\boldsymbol{\gamma}})\right]. \tag{2.3}$$

The trace of the matrix is denoted by tr. We can rewrite (2.1) as:

$$\begin{aligned} U_i &= \mathbf{X_i}\,\beta + \tau_i \\ \tau_i &= \mathbf{R_i}\,\gamma_i + \varepsilon_i \sim N_J\left(0,\ \mathbf{\Sigma} + \mathbf{R_i}\mathbf{\Sigma}_{\boldsymbol{\gamma}}\mathbf{R_i^T}\right) \\ &\text{so that:} \\ U_i &\sim N_J\left(\mathbf{X_i}\beta, \mathbf{\Sigma} + \mathbf{R_i}\mathbf{\Sigma}_{\boldsymbol{\gamma}}\mathbf{R_i^T}\right). \end{aligned} \tag{2.4}$$

The matrix transpose is denoted by T. Finally, as is typical in non-market valuation based on random parameters models, we assume that the dispersion matrix, $\mathbf{\Sigma}$, of the alternatives specific errors, ε_i, is an identity matrix. Thus the model outlined in 2.1-2.4 above is the Bayesian interpretation of the random parameters multinomial probit model of Hausman and Wise (1978).

2.2 Bayesian Estimation

In order to obtain estimates of the parameters of interest $(\beta, \mathbf{\Sigma}_{\boldsymbol{\gamma}})$, we need to relate the utilities, U_i, to an observed response. For instance, in the most common elicitation task the response is a choice, which then implies that if alternative j is chosen by person i, that

$$U_{ij} > U_{ik}\ \forall k \neq j. \tag{2.5}$$

In maximum likelihood, one then finds the probability that the U_i fall into the region implied by (2.5) and inserts it into the likelihood. Denoting the observed responses of all subjects by Y and the entire set of parameters to be estimated by θ, the likelihood is $p(Y|\theta)$.[2] In the frequentist approach, one maximizes $p(Y|\theta)$ over θ. The Bayesian approach seeks the posterior density of the parameters, $p(Y|\theta)$, via Bayes' Rule (see for instance Gelman *et al.* (1995)):

$$p\,(\,\theta|Y) = \frac{p\,(\,Y|\theta)\ p\,(\,\theta)}{p\,(\,Y)}, \tag{2.6}$$

where $p(Y)$ is the marginal or integrated likelihood (see for instance Chib (1995)), and $p(\theta)$ is the prior so that

$$p\,(\,Y) = \int p\,(\,Y|\theta)\,p\,(\,\theta)\ d\theta. \tag{2.7}$$

[2] We use the notation $p()$ to denote densities or probabilities - the context should make it clear.

Given that $p(Y)$ is a constant of proportionality, albeit a critical one for model comparison, (2.6) is commonly utilized in the following form where the posterior is proportional to the likelihood times the prior:

$$p\,(\,\theta|Y)\;\propto\;p\,(\,Y|\theta)\;p\,(\,\theta)\,. \tag{2.8}$$

In both frequentist and Bayesian approaches, the presence of $p(Y|\theta)$ makes estimation computationally demanding. In the Bayesian approach, Albert and Chib (1993) and McCulloch and Rossi (1994) showed how one could use Monte Carlo Markov chain simulation methods combined with Data Augmentation (Tanner and Wong (1987)) to circumvent the computation of terms relating to $p(Y|\theta)$. When applied to the Multinomial Probit model, one applies a specific form of MCMC known as Gibbs sampling. The Gibbs sampling method involves sampling from the posterior density, $p(\theta|Y)$, via an iterative series of conditional densities, inducing a Markov chain with stationary density being the posterior density. Thus one can sample from the posterior density without having to know its form. This proves convenient for the multinomial probit model once one augments with unobserved utilities (Tanner and Wong's, 1987). Below we describe the form of the Gibbs sampling algorithm we apply for our model.

Analytical and numerical calculation of the posterior distribution from the model in (2.1) is infeasible due to the latent variable model subject to the linear inequality restrictions in (2.5). However, the prior specifications allow the application of the Gibbs sampler to draw Monte Carlo posterior inferences. The Gibbs sampler generates a sample $\theta^{(1)},\ldots,\theta^{(T)}$ from the posterior density $p(\theta|Y)$ without having to sample directly from this unknown posterior density. In our model, the sampler iteratively draws from the full conditional densities $p(\beta|U,\gamma,\mathbf{H},Y)$, $p(\gamma|U,\beta,\mathbf{H},Y)$, $p(\mathbf{H}|U,\gamma,\beta,Y)$, $p(U|\beta,\gamma,\mathbf{H},Y)$ so that the sample induces a Markov chain with stationary distribution $p(\beta,\mathbf{H}|Y)$. Here we have dropped the i subscripts for convenience and are viewing the relevant quantities in vector form.

In (2.4) we combine the random effects and the error terms. This formulation suggests that we may generate utilities U from a density dependent on the random effects γ only through the dispersion $\mathbf{\Sigma_\gamma}$ (or precision $\mathbf{\Sigma_\gamma^{-1}}$). Since the fixed effects β are drawn from a regression of utilities on the design matrix $X_\mathbf{i}$ with a known covariance matrix, this Gibbs step is also independent of the individual random effects γ. Thus once we know the error dispersion, $\mathbf{I}+\mathbf{R_i}\mathbf{\Sigma_\gamma}\mathbf{R_i}^T$, knowledge of the actual γ is redundant for updating the utilities and the fixed effect coefficients. This blocking of the Gibbs sampler is the Chib and Carlin (1999) approach of marginalizing the distribution of U over the random effects γ.

The conditional densities required follow. Let $\mathbf{\Omega_i}=\mathbf{I}+\mathbf{R_i}\mathbf{\Sigma_\gamma}\mathbf{R_i}^T$, and note that this density is before differencing with respect to the first choice.

Then each iteration of the Gibbs sampler makes a single draw from each of the following conditional densities in turn:

$$
\begin{aligned}
&\beta|U,\mathbf{H},Y \;\sim\; N_K\left(\hat{\beta},\hat{\Sigma}_{\boldsymbol{\beta}}\right)\\
&\gamma_i|U,\beta,\mathbf{H},Y \;\sim\; N_L\left(\hat{\gamma}_i,\hat{\mathbf{\Sigma}}_{\boldsymbol{\gamma_i}}\right)\\
&\mathbf{H}|U,\beta,\gamma,Y \;\sim\; Wishart\left(\nu_{\boldsymbol{\gamma}} \;+\; n,\mathbf{V}_{\boldsymbol{\gamma}} \;+\; \sum_{i=1}^{n}\left(\gamma_i-\bar{\gamma}\right)\left(\gamma_i-\bar{\gamma}\right)^T\right)\\
&U_{ij}|U_{i:-j},\beta,\mathbf{H},Y \;\sim\; Truncated\ Normal\ \left(m_{ij},\eta_{ij}^2;\tau_{ij}^{+},\tau_{ij}^{-}\right)
\end{aligned}
\tag{2.9}
$$

where $i=1,\ldots,n$ $j=1,\ldots,J$ and $U_{i:-j}$is the $(J-1)$-vector of $U_i$with the j^{th} element removed, $\bar{\gamma}$ is the sample mean of the γ_i, and

$$
\begin{aligned}
\hat{\beta} &= \hat{\Sigma}_{\boldsymbol{\beta}}\left\{\sum_{i=1}^{i=n}\left(\mathbf{X_i}^T\mathbf{\Omega_i}^{-1}U_i\right)+\mathbf{A}^{-1}\bar{\beta}\right\}\\
\hat{\Sigma}_{\boldsymbol{\beta}} &= \left(\sum_{i=1}^{i=n}\left(\mathbf{X_i}^T\mathbf{\Omega_i}^{-1}\mathbf{X_i}\right)+\mathbf{A}^{-1}\right)^{-1}\\
\hat{\gamma}_i &= \hat{\mathbf{\Sigma}}_{\boldsymbol{\gamma_i}}\left\{\mathbf{R_i}^T\left(U_i-\mathbf{X_i}\beta\right)\right\}\\
\hat{\mathbf{\Sigma}}_{\boldsymbol{\gamma_i}} &= \left(\mathbf{H}+\mathbf{R_i}^T\mathbf{R_i}\right)^{-1}\\
m_{ij} &= x_{ij}\beta+\mathbf{F}^T\left(U_{i;-j}-\mathbf{X_{i;-j}}\beta\right)\\
\eta_{ij}^2 &= \omega_i\left(j;j\right)-\omega_i\left(j;-j\right)\mathbf{F}\\
\mathbf{F} &= \mathbf{\Omega_i}^{-1}\left(-j;-j\right)\omega_i\left(-j;j\right).
\end{aligned}
\tag{2.10}
$$

Here x_{ij} is the j^{th} row of $\mathbf{X}_i$, $\mathbf{X}_{i;-j}$ is $\mathbf{X}_i$ with the j^{th} row removed, $\omega_i(j;j)$ is the (j,j) element of $\mathbf{\Omega_i}$. $\omega_i(j;-j)$ and $\omega_i(-j;j)$ are the j-th row and column of $\mathbf{\Omega_i}$ with the jth element removed, and $\mathbf{\Omega_i}^{-1}(-j;-j)$ is $\mathbf{\Omega_i}$ with the j^{th} row and column removed. The conditional means and dispersions above follow from general normal linear model theory.

The truncated normal distribution in (2.9) has mean m_{ij}, variance η_{ij}^2, and upper and lower truncation points τ_{ij}^{+} and τ_{ij}^{-} respectively. The truncation points allow us to account for the ordinal content of the SP data. Specifically, the truncations bound each utility according to the subjects' responses. For instance, for a choice, one would use (2.5) to define the truncation points. This can be implemented by first differencing each utility with respect to the most preferred choice, resulting in $(J-1)$ utility differences that are each less than 0. This also illustrates that in a given application τ_{ij}^{+} or τ_{ij}^{-} might equal $+\infty$ or $-\infty$. Drawing the utilities from the appropriate truncated normal

distributions is the Data Augmentation step, which allows the algorithm to avoid the computation of the likelihood.

Note that this form of the sampler handles any ordinal structure in the SP responses (choices, ranking, partial rankings, First/Worst, etc.). To model SP data with a different ordinal structure, one simply draws the utilities in (2.9) from the appropriate truncated normal distribution. For instance, if the structure of the model is that alternative A is revealed preferred to alternative B, but is A is worse than C and D, one simply draws truncated normal utilities U_A, U_B, U_C, U_D, that meet these conditions. Conditional on the utilities, the other distributions remain unaltered. In other words, to represent different types of SP tasks, one needs to adjust $p(U|\beta, \gamma, \mathbf{H}, Y)$, but not $p(\beta|U, \gamma, \mathbf{H}, Y)$, $p(\gamma|U, \beta, \mathbf{H}, Y)$, $p(\mathbf{H}|U, \gamma, \beta, Y)$.

The Gibbs sampler successively samples from each of the conditional distributions listed in (2.9). The utility simulation is in itself another application of a Gibbs sampler. The overall algorithm is thus a Gibbs sampler embedded within a Gibbs sampler requiring updates of $(n(J-1)+n+2)$ conditional densities (recalling that there are $(J-1)$ utilities after differencing). In our experience, the Gibbs sampler above requires a burn-in of only 10^2 or 10^3. For further details regarding the implementation of the sampler, see Layton and Levine (2003).

2.3 Model Comparison and Marginal Likelihood Computation

For a Bayesian approach to model comparison one can compute the Bayes factor (see Kass and Raftery (1995) for an in-depth discussion). The Bayes factor, BF_{12}, for comparing two models M_1 and M_2 (with equal prior probabilities) is:

$$BF_{12} = \frac{p\left(Y|M_1\right)}{p\left(Y|M_2\right)} = \exp\left\{\ln p\left(Y|M_1\right) - \ln p\left(Y|M_2\right)\right\}, \tag{2.11}$$

where $p(Y|M_s)$ is the marginal likelihood for model s (as in (2.7) with an explicit conditioning on a particular model). The second representation in (2.11) is useful since the marginal likelihood algorithm we will employ yields the log marginal likelihood.

There are a variety of possible approaches to computing $p(Y)$ (dependence upon the model indicator is suppressed for the time being). A number of these are developed in Newton and Raftery (1994). Unfortunately, some of the easiest approaches, such as forming a simulation estimate by drawing from the prior $p(\theta)$ and then averaging $p(Y|\theta)$, and a harmonic mean simulator using draws from the posterior, may perform poorly as discussed in Newton and Raftery (1994) and Chib (1995). Bos (2002) provides Monte Carlo evidence that in even a simple normal regression model these approaches may be com-

pletely unreliable. So instead we follow Chib (1995) and use the Gibbs output to estimate the marginal likelihood. While this is computationally demanding and time consuming to code, it appears to work well in practice (see Bos (2002) for Monte Carlo results).

Following Chib (1995) note that the marginal likelihood $p(Y)$ for the model in (2.9) can be written as:

$$p\left(Y\right) = \frac{p\left(Y|U,\beta,\gamma_1,\ldots,\gamma_n,\mathbf{H}\right)p\left(U,\beta,\gamma_1,\ldots,\gamma_n,\mathbf{H}\right)}{p\left(U,\beta,\gamma_1,\ldots,\gamma_n,\mathbf{H}|Y\right)}, \tag{2.12}$$

where $p\left(U,\beta,\gamma_1,\ldots,\gamma_n,\mathbf{H}\right)$ and $p(U,\beta,\gamma_1,\ldots,\gamma_n,\mathbf{H}|Y)$ are the prior and posterior densities, respectively, over these parameters. Extending on the results of Chib (1995) and Yu and Chan (2001), we may use our Gibbs algorithm to estimate this quantity. Note the following.

- Likelihood:

$$p\left(Y|U,\beta,\gamma_1,\ldots,\gamma_n,\mathbf{H}\right) = 1 \tag{2.13}$$

  since the responses are deterministic once the utilities are given and known.

- Prior:

$$\begin{aligned} p\left(U,\beta,\gamma_1,\ldots,\gamma_n,\mathbf{H}\right) = &\left\{\prod_{i=1}^{i=n} p\left(U_i|\beta,\gamma_1,\ldots,\gamma_n,\mathbf{H}\right)\right\} \times \\ &p\left(\beta\right) \times \left\{\prod_{i=1}^{i=n} p\left(\gamma_i|\mathbf{H}\right)\right\} \times p\left(\mathbf{H}\right), \end{aligned} \tag{2.14}$$

  where the prior distributions for $\beta,\gamma_1,\ldots,\gamma_n,\mathbf{H}$ are given in (2.2).

- Posterior:

$$\begin{aligned} p\left(U,\beta,\gamma_1,\ldots,\gamma_n,\mathbf{H}|Y\right) = &\, p\left(U|\beta,\gamma_1,\ldots,\gamma_n,\mathbf{H},Y\right) \times \\ &p\left(\beta|\gamma_1,\ldots,\gamma_n,\mathbf{H},Y\right) \times \\ &p\left(\gamma_1,\ldots,\gamma_n|\mathbf{H},Y\right) \times p\left(\mathbf{H}|Y\right), \end{aligned} \tag{2.15}$$

  where each conditional distribution on the right side may be estimated using a part of the Gibbs sampler output from (2.9).

We can thus estimate the marginal likelihood in (2.12) using these three pieces. Estimation of the posterior is complicated, and so we provide details in Appendix B. Using the marginal likelihood we can compare models and select a model using the Bayes factor in (2.11).

3. The Stated Preference Data Sets

The SP survey data sets we use represent three different but related surveys. All of the surveys were written in the early/mid 1990's and explored Washington State residents' values for protecting old (ancient) growth habitat and the northern spotted owl. This was a major policy issue at the time because protecting the northern spotted owl meant protecting an enormous amount of its old growth forest habitat, with an associated decrease in logging and timber production.[3] The basic design of the surveys consisted of information and context, and then a SP valuation question in which respondents were asked to provide their most and second most preferred choices from five different alternatives. The alternatives varied in terms of their costs (in $ per year for ten years), the amount of ancient forest protected and the probability that the northern spotted owl would survive in the wild for the next 150 years. The third of the three surveys included the percentage of salmon stocks within the northern spotted owl's range that would survive for 150 years. This salmon survival attribute was added in the third survey as protecting ancient forests also impacts salmon survival. Salmon conservation is an important and ongoing issue in the U.S. Pacific Northwest.

All of the surveys were in written form and self-administered in a controlled setting to randomly recruited residents of Seattle, Washington. Below we describe each survey and how we will include it in our analysis. For an example choice set see Appendix A.

Survey 1: The basic attributes were the probability of survival of the northern spotted owl, acres of old growth habitat conserved, and the cost. The survey presented five alternatives and the respondents picked their most and second most preferred alternatives. 108 useable surveys were obtained in Seattle in November 1993. We will use the results of this survey to form a prior for survey 2.

Survey 2: The same basic survey instrument as survey 1, but with a better experimental design for the attributes (recall that SP survey experimental design has evolved considerably in the last decade). This survey was administered during March of 1995, more than a year after survey 1. It was used to add more variability to the experimental design matrix. It has 46 useable surveys, which makes analysis of the data by itself via maximum likelihood fruitless, yet given that a year had elapsed, one might be concerned about whether the two surveys could be pooled. One could test this, but given the limited sample sizes one might question the results.

[3]The development team for these surveys was led by Gardner Brown of the University of Washington, with assistance and input from David Layton, Jeffrey Lazo, Bill Schulze, Mike Dekay, and Gary McClelland. For more detail than contained here, see Brown et. al. (1994), and Layton (1995).

Survey 3: This survey added an attribute (and necessary discussion) for the percentage of salmon stocks preserved. It was designed to test how and if values for the spotted owl and old growth forests changed with the inclusion of salmon. This survey was administered in Seattle during March of 1995 and has 161 useable responses. We will use the posterior from survey 2 to form priors for the common attributes in survey 3. These consist of price, habitat, and owl survival.

4. Results

We compare the following five models. Note that the sum total of parameters includes the number of fixed effects and number of covariance terms on the random effects.

- Model I ("main effects"): The simplest model has no random effects and no interactions. This model fits fixed effects on price, owl survival, salmon survival, and habitat preservation. Thus, β is 4×1, there are no γ_i parameters and consequently no dispersion matrix $\mathbf{H}$, yielding a grand total of 4 parameters.
- Model II ("main effects + environmental interactions"): We may consider adding interaction terms to Model I. This model fits fixed effects on price, owl survival, salmon survival, and habitat preservation as well as the three two-way interactions between the environmental variables. β is 7×1, yielding a grand total of 7 parameters.
- Model III ("main and quadratic effects + interactions"): We may consider adding quadratic terms for the environmental variables to Model III. This model contains ten parameters: four main effects, three interaction terms, and three quadratic terms. β is 10×1, yielding a grand total of 10 parameters.
- Model IV ("main + price interactions"): An alternative to Model II considers interactions with price. This model again fits fixed effects on price, owl survival, salmon survival, and habitat preservation as well as the three two-way interactions between price and environmental variables. β is 7×1, yielding a grand total of 10 parameters.
- Model V ("hierarchical model"): This model fits a fixed effect on price (one parameter), fixed (three parameters) and correlated random effects on owl survival, salmon survival, and habitat preservation variables. Thus, β is 4×1, the γ_i are 3×1 vectors parameterized by $\mathbf{H}$ a 3×3 matrix, for a grand total of 10 parameters.

These models cover a range of popular specifications, some of which allow *WTP* to be a non-linear function of habitat and species survival probabil-

ities. Model I represents the most limited and common type of model estimated. Models II and III add in interactions among the environmental quality attributes, Model IV includes price and environmental quality interactions as one way of capturing income effects, and Model V is the hierarchical analog of a random parameters multinomial probit model. In all cases we utilize the partial ranking data we have available in which the respondent provided their most and second most preferred alternatives. As discussed in Section 2, the Gibbs sampler is modified to handle this kind of data by altering the truncation points in the utility sub-sampler so that the utility of the most preferred alternative is always greater than the utility of the second most preferred alternative, and in turn that the utility of second most preferred alternative is greater than the remaining three alternatives.

Model I was first fit using survey data set 1 with a proper but noninformative prior. This prior is:

$$\beta_{data\,set\,1} \sim N_K \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1000 & 0 & 0 \\ 0 & 1000 & 0 \\ 0 & 0 & 1000 \end{bmatrix} \right). \tag{4.1}$$

The posterior distribution given this data set was then used to form a new prior distribution for data set 2. A normal distribution was fit to the posterior distribution for use as the prior distribution in fitting data set 2. As discussed in Train (2003), the posterior distribution is asymptotically normal. Given the size of our data sets, we do not feel comfortable invoking asymptotic results for our final estimates, but for use as a prior fitting a normal to the posterior seems reasonable. This prior is shown below.

$$\beta_{data\,set\,2} \sim N_K \left(\begin{bmatrix} -.0062 \\ .8872 \\ .1000 \end{bmatrix}, \begin{bmatrix} .0000 & -.0003 & -.0005 \\ -.0003 & .1109 & -.0104 \\ -.0005 & -.0104 & .0134 \end{bmatrix} \right). \tag{4.2}$$

We note that there is surprisingly little variance in the cost parameter (the values are rounded to four places). One could choose to inflate this variance in order to make the final results less sensitive to the prior, but we decided to use the prior unaltered from the data set 1 estimation. We then estimated Model I using data set 2 and the prior in (4.2). We again fit a normal distribution to the posterior distribution from the data set 2 for use in fitting models I-V using data set 3. This prior is in (4.3).

$$\beta_{data\,set\,3} \sim N_K \left(\begin{bmatrix} -.0039 \\ .9885 \\ .1257 \end{bmatrix}, \begin{bmatrix} .0000 & -.0000 & -.0000 \\ -.0000 & .0213 & -.0004 \\ -.0000 & -.0004 & .0008 \end{bmatrix} \right). \tag{4.3}$$

Again we note little variance in the cost parameter or covariance with other parameters. Note that data sets 1 and 2 do not have the Salmon survival attribute,

so for the Salmon survival attribute we choose a zero mean and variance of 1,000 as in the initial priors on Cost, Owls, and Habitat. Any other parameters, such as those on the interaction or quadratic terms are also given a mean 0 with variance of 1000. For model V we also need to choose a prior for $\mathbf{H}$. We chose $\nu_{\gamma} = 4$, and

$$\mathbf{V}_{\gamma} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5000 & 0 \\ 0 & 0 & 900 \end{bmatrix}. \tag{4.4}$$

The values in $\mathbf{V}_{\gamma}$ were chosen based upon inflating the variances obtained from some earlier maximum likelihood estimation of a random parameters multinomial probit model (Layton (1995)). The final results are not sensitive to further inflating $\mathbf{V}_{\gamma}$.

We then run Models I-V on data set 3, using the priors as outlined above. As noted in section 3, our blocking of the Gibbs sampler results in fast convergence, requiring only 10^2 or 10^3 burn-in iterations. The results of the models are shown in Table 10.1. Our results are based on a very conservative 10^4 burn-in, and we base our estimates on an additional 40,000 draws. The units for each main effect variable are: Cost is in dollars ($) per year; Owl and Salmon survival probability are both represented in decimals (55% is entered as 0.55); Habitat preservation is measured in millions of acres. For each parameter in Table 1, the posterior mean is provided, along with 95% interval estimates (equal-tailed) in parentheses. Note that these interval estimates do not assume asymptotic normality but are computed directly from the 40,000 draws of the posterior provided by the Gibbs sampler.

We use the approach described in section 2.3 for computing marginal likelihoods for each of the five models.[4] Table 10.1 shows that the hierarchical model performed best, displaying a larger marginal likelihood on the log scale. We compute Bayes factors for each of Models I-IV compared to Model V (the best fitting model), using (2.11). Using the scale suggested in Kass and Raftery (1995), any Bayes factor greater than 100 represents decisive evidence against the competing model (Models I-IV). With a minimum value of 1,225 for Model II, we find decisive evidence that Model V is the best model to use for policy analysis.

The hierarchical model V in Table 10.1 has mean estimates for Cost, Owls, and Salmon that are of the expected sign, and interval estimates that do not cross zero. The intervals are reasonably tight. The estimates for Habitat indicate a negative value which is unexpected if one perceives habitat protection

[4]Note the Gibbs sampler and the marginal likelihood computations are enormously simplified in Models I-IV compared to Model V. The marginal likelihood could be computed directly by the method of Chib (1995) with the addition of a likelihood simulator for models I-IV.

Table 10.1. Posterior Mean and 95% Interval Estimates

Parameter	*Model* I	*Model* II	*Model* III
Cost	–0.005	–0.009	–0.009
	(–0.009,–0.002)	(–0.014,–0.005)	(–0.014,–0.004)
Owls	1.759	–3.024	–6.572
	(0.577,2.974)	(–6.542,0.487)	(–13.826,0.581)
Salmon	2.603	1.901	12.225
	(0.348,4.939)	(–1.324,5.196)	(–9.324,33.869)
Habitat	0.065	0.178	0.24
	(–0.161,0.289)	(–0.291,0.652)	(–0.316,0.787)
owl×owl			–0.656
			(–5.018,3.726)
sal×sal			–9.503
			(–28.748,9.847)
hab×hab			–0.034
			(–0.104,0.037)
owl×sal		7.936	13.872
		(1.905,14.092)	(–0.938,28.735)
owl×hab		2.243	0.485
		(–29.554,34.250)	(–58.478,59.670)
sal×hab		–30.47	–1.455
		(–80.256,19.546)	(–61.042,58.641)
$\ln p = (y\|M_s)$	–424.667	–422.968	–432.603
Bayes Factors	6,701	1,225	18,736,799

Notes: The (equal tailed) interval estimates are in parenthesis.
$\ln p = (y|M_s)$ is the log marginal likelihood.
The Bayes Factors are for comparing Model V versus the other models.

as good, but is expected if one views it as imposing economic hardship. In any event, the interval estimate is wide, around plus or minus five times the mean and covers zero. In effect there is little evidence to support any value other than 0. The $\mathbf{H}$ matrix is not precisely estimated, but clearly contributes to the model's predictive power as shown by the Bayes factors. These results are qualitatively similar to those often seen in random parameters models estimated by maximum likelihood. A decrease in the scale of the error term is shown by the increase in the magnitude of the main effects and a large improvement in fit.

Table 10.2 provides *WTP* estimates for model V. These are easy to compute given the structure of the model and the availability of the posterior draws for the parameters. Draws for the *WTP* distribution for an incremental increase in one of the environmental attributes are made by dividing the draws of Owl, Salmon, and Habitat parameters by the corresponding draw of the negative of

Table 10.1 (continued)
Posterior Mean and 95% Interval Estimates

Parameter	*Model* IV	*Model* V
Cost	–0.013	–0.207
	(–0.032,0.006)	(–0.274,–0.151)
Owls	1.01	11.703
	(–0.431,2.486)	(2.178,22.630)
Salmon	2.814	165.073
	(–0.095,5.789)	(119.891,218.672)
Habitat	0.2	–0.354
	(–0.054,0.455)	(–2.238,1.570)
owl×cost	1.403	
	(0.292,2.507)	
hab×cost	1.324	
	(–0.575,3.230)	
sal×cost	–25.909	
	(–45.960,–5.852)	
H_{11}		3.844
		(0.002,14.497)
H_{21}		–0.432
		(–2.159,0.000)
H_{22}		–3.973
		(–27.788,5.214)
H_{31}		0.062
		(0.000,0.385)
H_{32}		0.654
		(–0.322,4.827)
H_{33}		9.192
		(0.028,67.496)
$\ln p = (y\|M_s)$	–429.962	–415.857
Bayes Factors	1,335,745	—

Notes: The (equal tailed) interval estimates are in parenthesis.
$\ln p = (y|M_s)$ is the log marginal likelihood.
The Bayes Factors are for comparing Model V versus the other models.

the Cost parameter (that is, element by element division times –1). This yields 40,000 *WTP* draws, which are summarized in Table 10.2. In interpreting Table 10.2, note that these *WTP*'s are per year for ten years. Finally, note that these *WTP*'s are based on a model which implies that *WTP* is a linear function of species survival probability. This would be an overly restrictive assumption to make at the outset, but we did estimate models that allowed for non-linear *WTP* in species survival, and used model selection techniques to choose our

final model. It is simply the case that the available data best supported a model with linear *WTP*.

Table 10.2. Posterior Mean and 95% Interval Estimates of *WTP* from Model V

	Owls	*Salmon*	*Habitat*
Mean	0.5696	8.0432	–1.9408
(95% interval estimate)	(0.109,1.058)	(6.378,10.014)	(–11.799,7.056)

Notes: Model V is the Hierarchical model. *WTP* is measured in 1995 $ per year for 10 years.
Owls units are per % improvement in the probability of species survival for 150 years.
Salmon units are per % of stocks surviving for 150 years.
Habitat units are per million acres of ancient forest preserved.

Space precludes a complete interpretation of the policy implications of our results, but we can briefly point out the assumptions necessary to put estimates such as these into terms useful for benefit-costs analysis, and what the likely outcome of such an analysis would be. Montgomery *et al.* (1994) considered the marginal cost of northern spotted owl preservation, where the margin is defined in terms of the probability of species survival for 150 years, as in our SP survey data. Indeed, the SP survey was designed to provide benefits suitable for comparison with the costs estimates of Montgomery *et al.* (1994). Montgomery *et al.* (1994) provided cost estimates for three programs under federal consideration at that time. These programs would have yielded a 95%, 91%, and 82% survival probability at a cost of 46, 33, and 21 billion 1990 US $ respectively, or about 53, 38, and 24 billion in March 1995 US$ (as in our survey). Our values are in a series of 10 year payments, so we need to convert them to a lump sum, which means picking a discount rate. Using a 10% rate (2%-3% percent higher than 10 year bonds or 1 year adjustable mortgages at that time), we find that our values for 95%, 91%, and 82% survival are respectively $325, $312, and $281 per household.[5] To cover the costs of the programs as listed above would require approximately 141 million, 106 million, and 75 million households. In 1995 there were about 97 million households in the U.S., so even heroically assuming that all U.S. households had values similar to those of Seattle Washington residents, only the 82% program would yield positive net benefits. Obviously, the extent of the market might be significantly smaller, say Washington, Oregon, and California households, in which case none of the programs would yield positive net benefits based on the *WTP* to

[5]Assuming 0% survival probability as the baseline. Our survey actually used 5% as the status quo baseline. One could adjust the numbers for the 5% baseline but it will not change the nature of the results. We assume that the payments begin at the end of year 1.

preserve the northern spotted owl (using 14 million households in these three states). If one were to use the upper bound of the interval estimate, which is 86% higher than the mean estimate, then the national population of households is sufficient to cover the costs, but still California, Oregon, and Washington, are not sufficient.[6]

The programs clearly preserve more than owls, but the habitat value is essentially zero in our results. Salmons are an important component, but present a challenge in scaling the effects of the programs to salmon stocks preserved, so we postpone such an analysis to future research. The above should give the reader a good feel for the nature of the results based on the owl survival values alone.

5. Conclusion

In this research we have shown how one can use Bayesian estimation to handle a series of surveys, or pre-tests, to form priors for the final survey. This allows one to use information from earlier efforts, even of very small sample size, without having to assert that the parameter estimates (to scale) would be the same. The point of survey pre-testing is to improve the instrument and one expects that values are likely to be conditional upon the instrument (at least the less-tested early versions). Further, by using the MCMC sampling approach, we obtain draws from the posterior which can be used to form point and interval estimates without resorting to normal approximations for our final estimates. We think this is important, as the sample sizes we used here were not large. We have seen a number of published studies with samples sizes of similar magnitude.

We also illustrated a method for computing the marginal likelihood of a model, a critical component in Bayesian model comparison. This calculation is complicated to say the least. We have found, as discussed in Train (2003), a number of computational advantages to Bayesian estimation of discrete choice models. However, the ability to avoid the computation of the likelihood has a hidden cost in that model comparison requires it, or some quantity related to it. In the MCMC set-up, the marginal likelihood calculation can be as time consuming (or more) as the original estimation, but is still preferable to some of the easiest approaches to computing the marginal likelihood, which may not perform well in practice. We believe that finding faster, and more intuitive, algorithms for the types of models environmental economist typically use will be crucial to their gaining wider acceptance. Model estimation without model comparison is justifiably unsatisfying to readers and reviewers.

[6]All household data is from U.S. Census Bureau estimates accessed on the internet on July 30, 2004, at: http://eire.census.gov/popest/archives/household/sthuhh6.txt.

Appendix A: Example Choice Set

1) $0 per year for ten years to preserve no additional ancient forests.

- Protect 1.5 million acres (21%) of ancient forest habitat.
- Protect 450 northern spotted owl pairs (13%).
- The northern spotted owl would have a 5% chance of surviving in the wild for the next 150 years (a 95% chance of becoming extinct).
- 50% of the current salmon stocks would survive for the next 150 years.

2) $60 per year for ten years to preserve some additional ancient forests.

- Protect 2 million acres (29%) of ancient forest habitat.
- Protect 1,000 northern spotted owl pairs (29%).
- The northern spotted owl would have a 55% chance of surviving in the wild for the next 150 years (a 45% chance of becoming extinct).
- 58% of the current salmon stocks would survive for the next 150 years.

3) $109 per year for ten years to preserve more of the remaining ancient forests.

- Protect 2.8 million acres (40%) of ancient forest habitat.
- Protect 1,600 northern spotted owl pairs (46%).
- The northern spotted owl would have a 63% chance of surviving in the wild for the next 150 years (a 37% chance of becoming extinct).
- 63% of the current salmon stocks would survive for the next 150 years.

4) $169 per year for ten years to preserve still more of the remaining ancient forests.

- Protect 4 million acres (57%) of ancient forest habitat.
- Protect 1,900 northern spotted owl pairs (54%).
- The northern spotted owl would have a 81% chance of surviving in the wild for the next 150 years (a 19% chance of becoming extinct).
- 70% of the current salmon stocks would survive for the next 150 years.

5) $241 per year for ten years to preserve almost all of the remaining ancient forests.

- Protect 5.5 million acres (79%) of ancient forest habitat.
- Protect 2,400 northern spotted owl pairs (69%).
- The northern spotted owl would have a 98% chance of surviving in the wild for the next 150 years (a 2% chance of becoming extinct).
- 76% of the current salmon stocks would survive for the next 150 years.

Note that owl pairs and survival probability completely co-vary, so in effect there is one attribute one could call "owl survival". Similarly, the percentage of ancient forest habitat preserved and the total amount of habitat completely co-vary as well.

Appendix B: Estimation of the Posterior Density for Computation of the Marginal Likelihood

Here we provide the details for computation of the marginal likelihood in (2.12). The posterior distribution as written in (2.15) requires estimation of four conditional distributions. The asterisked variables represent the posterior mean of these quantities as obtained from the output of the Gibbs sampler in (2.9).

- $p\,(\,\mathbf{H}|Y)$ can be found by marginalizing over $U_i, \beta,$ and $\gamma = \gamma_1, \ldots, \gamma_n$.

$$p\,(\,\mathbf{H}|Y) = \int p\,(\,\mathbf{H}|U, \beta, \boldsymbol{\gamma}, Y)\, p\,(\,U, \beta, \boldsymbol{\gamma}|Y)\, dU\, d\beta\, d\boldsymbol{\gamma} \qquad (7.1)$$

  This quantity may be estimated from a set of T Gibbs samples from (2.9) via

$$\hat{p}\,(\,\mathbf{H}^*|Y) \;=\; \frac{1}{T}\sum_{t=1}^{t=T} p\left(\,\mathbf{H}^*|U^{(t)}, \beta^{(t)}, \boldsymbol{\gamma}^{(t)}, Y\right) \qquad (7.2)$$

- $p\,(\,\gamma_1, \ldots, \gamma_n|\mathbf{H}, Y)$ can be found by first noting that the $\gamma_1, \ldots, \gamma_n$ are conditionally independent, so we can focus on a single γ_i without loss of generality. Marginalizing over U_i, β:

$$p\,(\,\gamma_i|\mathbf{H}^*, Y) \;=\; \int p\,(\,\gamma_i|U, \beta, \mathbf{H}^*, Y)\; p\,(\,U, \beta|\mathbf{H}^*, Y)\, dU\, d\beta \qquad (7.3)$$

  We may obtain a Monte Carlo estimate of this quantity given samples from $p\,(\,U, \beta|\mathbf{H}^*, Y)$. Such samples may be drawn by running the Gibbs sampler in (2.9), substituting $\mathbf{H}^*$ for $\mathbf{H}$ and thus not sampling values for $\mathbf{H}$ during this sampler. We thus obtain:

$$\hat{p}\,(\,\gamma_i^*|\mathbf{H}^*, Y) \;=\; \frac{1}{T}\sum_{t=1}^{t=T} p\left(\,\gamma_i^*|U^{(t)}, \beta^{(t)}, \mathbf{H}^*, Y\right) \qquad (7.4)$$

- $p\,(\,\beta|\gamma_1,\ldots,\gamma_n,\mathbf{H},Y)$ can be found by marginalizing over the U_i.

$$p\,(\beta|,\boldsymbol{\gamma}^*,\mathbf{H}^*,Y) = \int p\,(\beta|U,\boldsymbol{\gamma}^*,\mathbf{H}^*,Y)\,p\,(U|\boldsymbol{\gamma}^*,\mathbf{H}^*,Y)\;dU \quad (7.5)$$

We may obtain a Monte Carlo estimate of this quantity given samples from $p\,(\,U|\gamma_1^*,\ldots,\gamma_n^*,\mathbf{H}^*,Y)$. Such samples may be drawn by running the Gibbs sampler in (2.9), substituting $\mathbf{H}^*$ for $\mathbf{H}$ and γ_i^* for all γ_i and not sampling the $\mathbf{H}$ or any γ_i values during the sampler. We obtain:

$$\hat{p}\,(\,\beta^*|,\boldsymbol{\gamma}^*,\mathbf{H}^*,Y)\;=\;\frac{1}{T}\sum_{t=1}^{t=T} p\left(\,\beta^*|U^{(t)},\boldsymbol{\gamma}^*,\mathbf{H}^*,Y\right) \quad (7.6)$$

Note too that in our Gibbs sampler, the conditional distribution for β does not directly depend on any of the γ_i, so that in effect the γ_i^* do not actually enter the computation above.

- $p\,(\,U|\beta,\gamma_1,\ldots,\gamma_n,\mathbf{H},Y)$. We first recognize that the $U_i, i=1,\ldots,n$, are conditionally independent. We may thus focus on a single U_i without loss of generality. The Gibbs sampler in (2.9) works on each U_{ij}, so we estimate each of the corresponding univariate densities. Note too that in our Gibbs sampler, the conditional distribution for U_{ij} does not directly depend on any of the γ_i. We may thus eliminate them from the conditioning argument. At the j^{th} step in the Gibbs sampler for utilities, we have values for the first $j-1$ utilities. So we can plug in the posterior mean values for these quantities.

$$\begin{aligned} p\left(\,U_{ij}|U_{i1}^*,\ldots,U_{i(j-1)}^*,\beta^*,\mathbf{H}^*,Y\right)\;=\;& \\ \int p\left(\,U_{ij}|U_{i1}^*,\ldots,U_{i(j-1)}^*,U_{i(j+1)},\ldots,U_{iJ},\beta^*,\mathbf{H}^*,Y\right)& \\ \times\;\; p\left(\,U_{i(j+1)},\ldots,U_{iJ},|U_{i1}^*,\ldots,U_{i(j-1)}^*,\beta^*,\mathbf{H}^*,Y\right)& \\ \times\; dU_{i(j+1)}\times\ldots\times dU_{iJ}& \end{aligned} \quad (7.7)$$

We may obtain a Monte Carlo estimate of this quantity given samples from $p\left(\,U_{i(j+1)},\ldots,U_{iJ}|U_{i1}^*,\ldots,U_{i(j-1)}^*,\beta^*,\mathbf{H}^*,Y\right)$. Such samples may be drawn by running the Gibbs sampler in (2.9) substituting posterior mean values where appropriate. In particular, we sample only the utilities following U_{ij}, namely, $U_{i(j+1)},\ldots,U_{iJ}$. We obtain:

$$\begin{aligned} \hat{p}\left(\,U_{ij}^*|U_{i1}^*,\ldots,U_{i(j-1)}^*,\beta^*,\mathbf{H}^*,Y\right)\;=\;& \\ \frac{1}{T}\sum_{t=1}^{t=T} p\left(\,U_{ij}^*|U_{i1}^*,\ldots,U_{i(j-1)}^*,U_{i(j+1)}^{(t)},\ldots,U_{iJ}^{(t)},\beta^*,\mathbf{H}^*,Y\right)& \end{aligned} \quad (7.8)$$

Over all j we have

$$\hat{p}\,(\,U_i^*|\beta^*,\mathbf{H}^*,Y)\;=\;\prod_{j=1}^{j=J}\hat{p}\left(\,U_{ij}^*|U_{i1}^*,\ldots,U_{i(j-1)}^*,\beta^*,\mathbf{H}^*,Y\right) \quad (7.9)$$

The marginal likelihood may be estimated at any parameter values, but preferably a point of high density, Chib (1995). We chose the posterior mean. We evaluate the prior and posterior distributions in (2.14) and (2.15) at the posterior mean value for each parameter. The estimated posterior (2.15) is:

$$\begin{aligned}\hat{p}\,(\,U^*,\beta^*,\boldsymbol{\gamma}^*,\mathbf{H}^*|Y)\;=\;&\left\{\prod_{i=1}^{i=n}\hat{p}\,(\,U_i^*|,\beta^*,\boldsymbol{\gamma}^*,\mathbf{H}^*,Y)\right\}\\ &\times\hat{p}\,(\,\beta^*|\boldsymbol{\gamma}^*,\mathbf{H}^*,Y)\\ &\times\left\{\prod_{i=1}^{i=n}\hat{p}\,(\,\gamma_i^*|\mathbf{H}^*,Y)\right\}\;\times\;\hat{p}\,(\,\mathbf{H}^*|Y)\end{aligned} \quad (7.10)$$

The estimated log marginal likelihood is then

$$\ln\hat{p}\,(\,Y)\;=\;\ln\hat{p}\,(\,U^*,\beta^*,\boldsymbol{\gamma}^*,\mathbf{H}^*)\;-\;\ln\hat{p}\,(\,U^*,\beta^*,\boldsymbol{\gamma}^*,\mathbf{H}^*|Y) \quad (7.11)$$

Chapter 11

BAYESIAN ESTIMATION OF DICHOTOMOUS CHOICE CONTINGENT VALUATION WITH FOLLOW-UP

Jorge E. Araña
Department of Applied Economic Analysis
*University of Las Palmas de Gran Canaria**
jarana@daea.ulpgc.es

Carmelo J. León
Department of Applied Economic Analysis
University of Las Palmas de Gran Canaria
cleon@daea.ulpgc.es

Abstract Dichotomous choice contingent valuation involves a binary yes/no question that can be followed by a subsequent question in order to obtain more information from the respondent, and leading to more efficient welfare estimates. Estimation methods for these data have mainly focused on classical maximum likelihood. In this paper we study possible improvements utilising a Bayesian MCMC approach to model this type of data. The classical and Bayesian approaches are compared with a Monte Carlo simulation experiment. The results show that the Bayesian approach improves the performance of the model, particularly with relatively small samples.

Keywords: Bayesian methods, Contingent valuation, Data augmentation, Gibbs sampling.

*We are grateful to two anonymous referees and the editors of this book for helpful comments. This work has been done during the stay of the first author in the Department of Agricultural and Resource Economics (ARE) of the University of California, Berkeley.

1. Introduction

Contingent valuation (CV) is a method that is commonly used among economists and social scientists for measuring the benefits society receives from non-market or environmental goods. The method involves the use of survey work, in which individuals are asked an elicitation question inquiring their willingness to pay (*WTP*) or willingness to accept (*WTA*) for a particular change in their choice set. There are a number of ways of posing this question to the subject, going from the open ended format to the most recent multiple choice experiments. The dichotomous choice format requests a simple binary yes/no responses from the individual facing a particular definition of the change in the non-market good or policy proposal. This method was first proposed by Bishop and Herbelein (1979) and further developed by Hanemann (1984).

Dichotomous choice questions have several advantages over other approaches. First, they are straightforward and simple to answer. That is, the individual is posed with a simple decision about their preferences with respect to a particular profile of some policy proposal (or non-market good). Both open ended questions and multiple choice experiments are generally much harder for the subject to answer. Second, Hoehn and Randall (1987) show that dichotomous choice questions have the potential to be incentive compatible. With this format the subject perceives that the offered price is exogenous, and therefore she cannot influence the amount that would eventually pay if the policy proposal is approved. Obviously, this condition is not satisfied for open ended or multiple choice experiments, since the subject is given the opportunity to choose the characteristics of the good to be offered, including the price.

Nevertheless, it is known that by departing from dichotomous choice the researcher can increase the amount of information about the subject's preferences. An elicitation method which moves slightly from the single binary question is the double bounded (or dichotomous choice with follow up) format. It involves posing the subject a second binary *WTP* question for the same level of provision of non-market good asked in the first binary question. If the answer to the first price is yes (no) then the second price is accordingly increased (and viceversa). This method has been the subject of some controversy because it lacks the potential for incentive compatibility (e.g. Harrison and Kriström, 1995), and because it may generate responses which are internally inconsistent[1] (Bateman et al., 2001). However, it also increases the amount of information about the individual's preferences, and it has been shown to lead to more efficient estimators than the single bounded question format (Hanemann

[1]The term "internally inconsistent" refers to the fact that the distribution of underlying preferences implied by answers to the initial question may not be the same as that implied by the entire sequence of replies.

et al., 1991). Further elicitation questions offer more scope for bias effects and do not seem to be worth pursuing in terms of furthering the efficiency gains (Cooper and Hanemann, 1995; Scarpa and Bateman, 2000).

In spite of the potential trade-offs between single, double and multiple-bounded elicitation in terms of efficiency and other strategic response bias effects, there is a need to develop estimation methods which increase the precision and accuracy of the welfare estimates. In this chapter we focus on the contribution that can be provided by the application of Bayesian inference to dichotomous choice models in contingent valuation. Fernandez *et al.* (2004) show how Bayesian inference can be conduced for interval-data dichotomous choice models, such as double bounded, by demonstrating the propriety of the posterior distribution. Bayesian estimation methods are capable of incorporating prior information from experts and provide exact inference with relatively small samples.

In the next sections we consider the potential improvements of adopting a Bayesian estimation method. First, we outline a general Bayesian approach to dichotomous choice modelling. The method involves Gibbs sampling and data augmentation to overcome the technical problems encountered in simulating the posterior distribution. These techniques solve the problem of complex integration, thereby allowing the researcher to simulate the posterior distribution. Section 3 presents the empirical survey data for the application of the model. This section also discusses the results of the Bayesian modelling and compares them with the standard maximum likelihood approach. Finally, section 4 summarises the main findings and implications of the paper.

2. Modeling

The dichotomous choice elicitation method involves a binary yes/no question on the value of an environmental good for a given price. The response to this question is expected to be based on the individual's maximum willingness to pay (*WTP*), which is unobservable for the researcher. Following Cameron (1988), let us assume *WTP* is comprised of two components, a deterministic parameter μ and an unobservable random variable ε with zero mean and σ standard error. Thus, the response of individual *i* to binary question *j* is specified as $WTP_{ij} = \mu_i + \sigma\varepsilon_i$, where μ and σ are the location and scale parameters (mean and standard deviation respectively) of *WTP*. The location parameter can be expressed as a linear predictor associated with a $k \times 1$ regression parameter vector β and a covariate vector x_i, that is $\mu_i = x_i'\beta$.

By observing a subject's response (Yes/No) to an arbitrary first bid price B_{i1} it is assumed that her true valuation is higher ("Yes") or lower ("No") than B_{i1}.

Therefore, the probability of a " Yes" response is given by:

$$\Pr(WTP_{i1} \geq B_{i1}) = \Pr\left(\varepsilon_i \geq \frac{B_{i1} - \mu_i}{\sigma}\right) = F_\varepsilon\left(\frac{B_{i1} - \mu_i}{\sigma}\right) \tag{2.1}$$

As can be seen, the linear predictor for the location parameter is linked to the probability of a positive response by a known cumulative distribution function $F(\cdot)$, or link function.

For a given sample of n independent observations, Cameron and James (1987) show that the joint function for the data $f\left(Y|B_{i1}, X, \beta, \sigma\right)$ can be expressed as the following likelihood function:

$$L = \prod_{i=1}^{n}\left[F_\varepsilon\left(\frac{B_{i1} - x_i'\beta}{\sigma}\right)\right]^{y_{i1}} \times \left[1 - F_\varepsilon\left(\frac{B_{i1} - x_i'\beta}{\sigma}\right)\right]^{1-y_{i1}} \tag{2.2}$$

where the indicator variable $y_{i1} = 1$ if a positive response is observed, and 0 otherwise. In order to obtain more information about individuals preferences and account for potential heterogeneity, an increasing number of studies incorporate additional questions asking subjects if their *WTP* is strictly positive, zero or negative (Kriström, 1997; Haab, 1999, and Huhtala, 2000, among others).[2] The model can be extended to incorporate this additional information by incorporating a binary variable z_i, which takes the value of 1 if WTP is strictly positive and zero otherwise, thereby considering the negative part of the distribution, the contribution of the individual observation to the likelihood function would take the following form:

$$F_\varepsilon\left(\frac{B_{i1} - x_i'\beta}{\sigma}\right)^{y_{i1}z_i}\left[1 - F_\varepsilon\left(\frac{B_{i1} - x_i'\beta}{\sigma}\right)\right]^{(1-y_{i1})z_i} F_\varepsilon\left(\frac{-x_i'\beta}{\sigma}\right)^{1-z_i} \tag{2.3}$$

As Cameron and James (1987) pointed out, *"the presence of B_{i1} in the likelihood function makes possible that it can be maximized with respect to both the vector of coefficients (β) and the standard deviation (σ) by applying maximum likelihood methods"*. However, the properties of the ML estimators relay on asymptotic conditions and they may not be hold in finite samples (Griffith, Hill and Pope, 1987). The use of Bayesian methods is a reasonable alternative for small or finite samples, which may lead to more accurate estimation (McCulloch and Rossi, 1994).

Let us assume that some prior information is available about the parameters of the binary model. In this case, we can proceed by estimating the model in a Bayesian setting. For the sake of convenience, let us assume prior distributions

[2]The motivations of zero responses can be further separated in true zeros and protest responses (Strazzera *et al.* 2003).

$\beta/\sigma^2 \sim N(\beta_0, V_1)$ and $\sigma^2 \sim IG(a_1/2, b_1/2)$, where N is the normal distribution with mean vector β_0 and covariance matrix V_1; IG is the inverted gamma in where $a_1/2$ and $b_1/2$ are the scale and shape parameters respectively.

The posterior distribution can be evaluated by using a Gibbs sampling algorithm with data augmentation, similarly to Chib (1992) and Albert and Chib (1993). Data augmentation involves generating continuous variables defined in the censored intervals resulting from the discrete responses of the dichotomous choice method (Tanner and Wong, 1987). That is, a set of random variables WTP_{i1} is generated for positive WTP, such that $WTP_{i1} > B_{i1}$ if $y_{i1} = 1$ and $WTP_{i1} < B_{i1}$ if $y_{i1} = 0$. In addition, the negative part of the distribution is modelled by considering z_i to be a binary variable taking value one if the individual's WTP is strictly positive, and zero otherwise. Since the dependent variables follow a normal distribution, i.e. $WTP_{i1} \sim N(x_i'\beta, \sigma^2)$, the Gibbs algorithm involves sampling for $\left\{\beta, \sigma^2, WTP_{i1}, E\left(WTP_{i1}\right)\right\}$ iteratively from their posterior conditional density distributions, which are as follows:

$$f(W_{i1}\,|Y, Z, \theta) = \begin{cases} \phi\left(W_{i1} \mid x_i'\beta, \sigma^2\right) \mathrm{T}\left[B_{i1}, \infty\right] & \text{if } z_i y_{i1} = 1 \\ \phi\left(W_{i1} \mid x_i'\beta, \sigma^2\right) \mathrm{T}\left[0, B_{i1}\right] & \text{if } z_i(1 - y_{i1}) = 1 \\ \phi\left(W_{i1} \mid x_i'\beta, \sigma^2\right) \mathrm{T}\left[-\infty, 0\right] & \text{if } z_i = 0 \end{cases} \quad (2.4)$$

$$\pi\left(\beta|Y, Z, W_{11}, ...W_{n1}, \sigma^2\right) = \phi\left(\beta|\hat{\beta}_{WTP}, \tilde{V}\right)$$

$$\pi\left(\sigma^2|Y, Z, W_{11}, ...W_{n1}, \beta\right) = f_{IG}\left(\sigma^2|\frac{a_2}{2}, \frac{b_2}{2}\right) \quad (2.5)$$

where W_{i1} is WTP_{i1}; $Y = (y_{11}, ..., y_{n1})$; $Z = (z_1, ..., z_n)$; $\phi(\cdot)$ T$[a, b]$ is the density of a normal distribution truncated to the interval a, b, $f_{IG}$is the density function of an inverted gamma distribution, and

$$\tilde{V} = \left(\tfrac{1}{\sigma^2} \sum_{i=1}^{n} x_i x_i' + (V_1)^{-1}\right)^{-1},$$

$$\hat{\beta}_{WTP} = \tilde{V}\left(\tfrac{1}{\sigma^2} \sum_{i=1}^{n} x_i WTP_{i1} + (V_1)^{-1}\beta_0\right),$$

$$b_2 = b_1 + \sum_{i=1}^{n} \left(WTP_{i1} - x_i'\beta\right)^2, \qquad a_2 = a_1 + r, \text{ and } r = \sum_{i=1}^{n} z_i.$$

The Gibbs sampling algorithm can be summarized in the following steps:

0. Determine the initial values for β and σ. These values can be obtained from maximum likelihood or OLS estimation.

1. Generate sample values of W_{i1} from distribution (2.4) conditioning on the previous values of β and σ.

2. Obtain the arithmetic average of W_{i1}, *i=1,...,n*. This is a sample value of $E(WTP)$.

3. Sample β from distribution (2.5) conditioning on the most recent values of σ and W_{i1}obtained in step 1.

4. Sample from σ distribution (2.5) conditioning on the most recent values of β and W_{i1} obtained in step 1.

5. Repeat steps 1 to 4 until convergence is achieved.

After an initial "burn-in" period, the values thus generated may be regarded as drawn from the joint distribution of $E\left\{(WTP), \beta, \sigma^2 | Y\right\}$. These series of simulated values are utilized to estimate the posterior moments for the parameters after discarding the first d values in the chain.

As discussed in section 1, more information about WTP can be obtained by posing the subject with a second binary question, so as to increase the efficiency of the welfare estimates (Hanemann *et al.* 1991). The second bid offered (B_{i2}) is higher than B_{i1} if individual i answers positively to the first price and vice-versa. Thus, the probabilities of the possible bounds for *WTP* are defined as follows:

$$\begin{aligned}
\pi^i_{yy} &= \Pr\left(W_{i2} > B_{i2}\right) &&= \int_{B_{i2}}^{\infty} f_\varepsilon\left(\tfrac{W_{i2}-\mu_i}{\sigma}\right) dW_{i2} \\
\pi^i_{yn} &= \Pr\left(B_{i1} < W_{i2} \leq B_{i2}\right) &&= \int_{B_{i1}}^{B_{i2}} f_\varepsilon\left(\tfrac{W_{i2}-\mu_i}{\sigma}\right) dW_{i2} \\
\pi^i_{ny} &= \Pr\left(B_{i2} < W_{i2} \leq B_{i1}\right) &&= \int_{B_{i2}}^{B_{i1}} f_\varepsilon\left(\tfrac{W_{i2}-\mu_i}{\sigma}\right) dW_{i2} \\
\pi^i_{nn} &= \Pr\left(0 < W_{i2} \leq B_{i2}\right) &&= \int_{0}^{B_{i2}} f_\varepsilon\left(\tfrac{W_{i2}-\mu_i}{\sigma}\right) dW_{i2} \\
\pi^i_{0} &= \Pr\left(W_{i2} \leq 0\right) &&= \int_{-\infty}^{0} f_\varepsilon\left(\tfrac{W_{i2}-\mu_i}{\sigma}\right) dW_{i2}
\end{aligned} \tag{2.6}$$

where $f_\varepsilon(\cdot)$ is the density function of ε. Let y_{i2} be a second indicator variable taking values 0,1 for a "no" and "yes" answer to the second question, respectively. The contribution of the individual observation to the sample likelihood function is:

$$\left\{(\pi^i_{nn})^{(1-y_{i1})(1-y_{i2})}(\pi^i_{yy})^{y_{i1}y_{i2}}(\pi^i_{ny})^{(1-y_{i1})y_{i2}}(\pi^i_{yn})^{y_{i1}(1-y_{i2})}\right\}^{z_i}\left\{\pi^i_0\right\}^{(1-z_i)} \tag{2.7}$$

Parameters β and σ can be estimated by maximum likelihood methods applied to equation (2.7) (e.g. Hanemann *et al.* 1991). However, Bayesian estimation is feasible using Gibbs sampling and data augmentation, as in the case of

the single bounded model, and may have practical advantages in small samples. For estimation purposes, step 1 in the Gibbs sampling algorithm involves sampling for W_{i2} from the following conditional distribution:

$$f(W_{i2}|Y_{ij},\theta) = \begin{cases} \phi\left(W_{i2}|x_i'\beta,\sigma^2\right) \mathrm{T}\left[B_{i2,}\infty\right] & \text{if } y_{i1}y_{i2}z_i = 1 \\ \phi\left(W_{i2}|x_i'\beta,\sigma^2\right) \mathrm{T}[B_{i1},B_{i2}] & \text{if } y_{i1}(1-y_{i2})z_i = 1 \\ \phi\left(W_{i2}|x_i'\beta,\sigma^2\right) \mathrm{T}[B_{i2},B_{i1}] & \text{if } y_{i2}(1-y_{i1})z_i = 1 \\ \phi\left(W_{i2}|x_i'\beta,\sigma^2\right) \mathrm{T}\left[0,B_{i2}\right] & \text{if } (1-y_{i1})(1-y_{i2})z_i = 1 \\ \phi\left(W_{i2}|\ x_i'\beta,\sigma^2\right) \mathrm{T}\left[-\infty,0\right] & \text{if } z_i = 0 \end{cases} \tag{2.8}$$

Y_j =($Y_{1j}, Y_{2j}, ...Y_{nj}$) for $j = 1, 2$; x is a vector of covariates, and W_{i2} = (WTP_{12}, WTP_{22},..., WTP_{n2}), where n is the number of individuals in the sample.

Appendix 1 outlines a method to draw samples from the truncated normal distributions involved in (2.4) and (2.8).

3. Application and results

The empirical data to which the models are applied were obtained from a contingent valuation survey on the recreational value of Teide National Park in the island of Tenerife (Canary Islands, Spain). This park contains representative landscapes and ecosystems of the Canary Islands and the highest peak in Spain with 3,721 mts. The survey was carried out in 1997, interviewing local and foreign visitors. A total number of 1,045 on-site interviews were conducted. This number was reduced to 941 valid observations after screening from protest responses and excluding those cases with missing values in the covariates. Protest responses were attributed to those subjects who stated that the public authorities should pay for the natural area. There were also 96 individuals answering a zero value because they objected to pay for access to a public space for ethical reasons. The models developed in this paper allow us to specifically model zero and negative responses. In our case, these individuals can be treated either as zero or negative responses, thus they are not excluded for modelling purposes.

The valuation scenario considered a hypothetical entrance fee to be paid for access to the park. The utilisation of the money was justified with the aim of preservation. The valuation question reads as follows:

> "In this question we would like to know how much you value the experience you are having in Teide National Park. Let us consider you had to pay an entrance fee to enjoy the park. All visitors would have to pay and the money would be used for conserving the park in its current state. If you were asked to pay an entrance fee to enjoy the National Park, would you be willing to pay ____ euros?".

The elicitation process involved a first and a follow-up dichotomous choice questions. The starting prices were randomly distributed across the sample. The initial bid vector for the first dichotomous choice question was chosen from the analysis of the distribution for the responses to an open-ended question in the pre-test samples. These responses, together with the information from studies on the value of other natural areas in Spain, allowed us to determine a prior distribution for the responses to the *WTP* elicitation questions, from which the prior parameters were evaluated. A five bid vector design was derived from this joint empirical distribution following Cooper (1993). In order to generate equal intervals across the sample, the second upper and lower bids were set equal to the successive bid in the first vector. Table 11.1

Table 11.1. Explanatory variables in the models

BEFORE	= 1 if the subject had been before in the NationalPark (NP) = 0 otherwise
PINFOR	= 1 if the subject had information about the NP before the arrival to the island = 0 otherwise
MIMP	= 1 if subject said that visit the NP was very important in the choice of travel to the island. = 0 otherwise
YEDU	Years of education of the subject
PINC	Annual personal income of the subject

displays the various socioeconomic covariates found to be significantly associated with *WTP*. For the Bayesian estimation method we have assumed very non-informative or diffuse priors, as a convenient setting for estimation process. That is, since we have no prior information on model parameters we assumed $a_1 = 0$, $b_1 = 0$, $V_1 = 10^9 \times I_k$, where I_k is the identity matrix of dimension k (number of covariates including the intercept). The starting values for Gibbs sampling were taken from maximum likelihood estimation, and the number of iterations was 10,000. The computation of the posterior moments involved a burn-in of the first 2,000 observations, i.e. the results of these iterations were discarded for the final estimates. Convergence checks as in Raftery and Lewis (1992) and Geweke (1992) do not indicate any problems, and this is corroborated by visual inspection of trace plots and rerunning the samplers many times from different starting values. All the computations were undertaken with GAUSS.

Table 11.2 presents the results for the first stage in the elicitation process together with those for the second stage, or double bounded results, by using the

Table 11.2. Bayesian estimates for single and double bounded models.

Variables	*Single bounded*	*Double bounded*
Intercept	2.7097	3.3472
	(1.5915)	(0.955)
BEFORE	-3.382	-3.5975
	(1.2327)	(0.8129)
PINFOR	2.1819	1.701
	(1.0088)	(0.6312)
MIMP	1.6777	1.3424
	(0.9423)	(0.5848)
YEDU	0.2347	0.1854
	(0.974)	(0.0595)
PINC	0.0008	0.0008
	(0.0005)	(0.0003)
σ	9.532	6.893
	(0.815)	(0.272)
Mean *WTP* (Euros)	8.58	8.00
95% HDI.	[8.02, 9.12]	[7.72, 8.26]
Median *WTP* (Euros)	8.55	7.81
95% HDI	[7.99, 9.11]	[7.52, 8.09]

Standard errors in brackets.

Bayesian estimation method. The results for maximum likelihood estimation were quite similar for both models, and therefore are not reported here. The functional form for the valuation function is assumed to be linear with a normal distribution for *WTP*. This assumption was preferred over the lognormal transformation because it allowed us to model the negative part of the distribution and also provided slightly better results according to standard goodness of fit measures. The explanatory variables presented the expected results for both models, with *WTP* rising with income, education, the access to previous information, and the importance of the visit to the park, while declining for those subjects who had been before in the park.

As it is often observed in these comparisons, mean and median *WTP* are slightly larger for the single bounded method. Although there is some overlapping in the confidence intervals of these welfare estimates, it can be shown that they are significantly different at the 80 percent level. Nevertheless, the differences amount to a 7% (8%) reduction in the mean (median) in the double model with respect to the single. As expected, the double bounded assumption leads to more efficient results, as shown by the narrower confidence intervals.

The relative performance of the two estimation methods can be compared with a Monte Carlo experiment. True dichotomous choice data were simulated

under the assumption that the true parameters for the data generating process β^T and σ^T were coming from the single bounded model estimates reported above. The simulated data for the second stage responses are also based on the *WTP* from this data generating process, and are therefore absent from collateral biases such as "yea saying" or strategic behavior. The Monte Carlo experiment can be summarized in the following steps:

1 For each individual, true *WTP* is defined by applying parameters $\left(\beta^T, \sigma^T\right)$ to the valuation function, i.e. $WTP_i^T = x_i'\beta^T + \sigma^T\varepsilon_i$, where i is a random realization from a standard normal distribution.

2 From the sample observations, 150 random subsamples were taken of three alternative sizes, 150, 300 and 700 observations.

3 The responses for each individual in the subsample for the first question were generated by comparing the bid price received with WTP_i^T, i.e. if $WTP_i^T > B_{i1}$ then $y_{i1} = 1$.

4 The responses for each individual in the subsample for the second question were generated by comparing the second bid price received with WTP_i^T, i.e. if $WTP_i^T > B_{i2}$ then $y_{i2} = 1$.

5 The single and double bounded models were estimated for each subsample utilizing both the maximum likelihood and Bayesian methods.

6 The performances of the models were evaluated with the root mean squared error statistic (RMSE) as a percentage of the true mean WTP.

Table 11.3. Monte Carlo simulation results of root mean squared error (RMSE) as a percentage of true mean *WTP*

	Sample size		
	n = 150	n = 300	n = 700
Estimation method	*Single bounded*		
Maximum Likelihood	43.61	34.82	26.57
Bayesian	29.76	23.78	19.62
	Double bounded		
Maximum Likelihood	42.00	31.12	20.34
Bayesian	27.36	19.38	17.51

Table 11.3 presents the results of the Monte Carlo simulations for both estimation methods. For either the single or double bounded models, the results show that the Bayesian estimation method leads to a better performance according to

the RMSE at all sample sizes considered. These size cover the normal range of CV applications. The Bayesian improvements over maximum likelihood are substantially larger at small sample models. Double bounded estimates deliver larger improvements than single bounded one at small sample sizes: for n=300, the reduction in RMSE amounts to 31.7% with single bounded and 37.7% with double bounded.

Thus, the Bayesian approach to dichotomous choice provides a better representation of the true data generating process even at relatively large samples in the context of contingent valuation. This ultimately leads to a more accurate estimation of the parameters, hence to a more exact estimation of the welfare measures. The implication is that the Bayesian estimation through Gibbs sampling and data augmentation enhances the test comparisons commonly undertaken in CV research which often relies on small subsamples with different treatments.

These results are of difficult generalization because they are conditioned on the bid design and the assumption of the true distribution of *WTP*. Yet, they are promising. Our bid design was based on the methodology proposed by Cooper (1993) for a predetermined number of bids. However, several alternative designs are available (Boyle et al, 1988, Nyquist, 1990, Duffield and Paterson, 1991, Kanninen, 1993 and Alberini, 1995, among others). Some of these designs have been compared by using Monte Carlo simulations, finding that both welfare estimations and efficiency levels are sensitive to the design of the bid vector (Kanninen, 1995; Alberini, 1995; Scarpa and Bateman, 2000). The Bayesian estimation method could affect the trade-offs to be obtained from more efficient bid designs and the use of more efficient elicitation methods.

On the other hand, we have also assumed that the true conditional *WTP* distribution is a normal distribution, which can be unrealistic for some empirical data. There is evidence in some contexts that the *WTP* distribution can be multimodal and/or asymmetric, which yields maximum likelihood estimations based on symmetric distributions to be biased (Carson, 1994). Previous studies testing the effects of potential *WTP* misspecifications on welfare estimations can be found in Cooper and Hanemann (1995) and Crooker and Herriges (2004), among others. The Bayesian estimation methods could be compared with maximum likelihood under alternative assumptions for the true distribution and for the one utilized by the researcher. Potential trade-offs can be found from this analysis, in which misspecification errors are to be reduced by the utilization of more accurate Bayesian methods, particularly with small samples.

4. Conclusions

Dichotomous choice models based on binary yes/no questions constitute a straightforward and efficient method of estimating willingness to pay for environmental or non-market goods using a stated preference approach. Estimation with these models has been mostly based on maximum likelihood methods. Bayesian techniques allow the researcher to incorporate prior information and provide exact inference at small sample sizes. In this chapter we have shown how Bayesian estimation methods can be useful for conducting inference with single and double bounded dichotomous choice models. The methods are based on Markov chain Monte Carlo techniques combining Gibbs sampling and data augmentation, which allow the research to bypass the common problem of integrating over complex integrals to evaluate the posterior distribution in a Bayesian framework.

We feel that the comparison between conventional maximum likelihood methods and their Bayesian counterparts could help researches to decide as to what methods are more appropriate in each particular situation. Our simulation results show that the Bayesian estimation approach may increase the performance of the double bounded model (dichotomous choice with follow up) when compared with the single bounded (or single binary question) model, especially with small and relatively small samples. In any case, the Bayesian estimation approach always delivered a better representation of the empirical data than the maximum likelihood estimation, with either single or double bounded model. The improvements in performance with the Bayesian methods are larger and more important at small sample sizes, but we feel that they also deserve some consideration at relatively large sample sizes.

A practical implication of the results in this chapter is that stated preference researchers using CV techniques could find it useful to conduct inference with Bayesian methods when their sample sizes are not sufficiently large, which is commonly the case in most research studies. In particular, the comparisons between treatments which are common with methodological studies of CV and stated preference methods could benefit substantially from conducting a Bayesian inference based approach.

However, in order to fully understand the potential advantages of using Bayesian estimation techniques in CV studies more research is needed. Of particular interest would be to test the sensitivity of the results presented here to variations in bid design and in assumptions on implicit distribution for conditional *WTP*. To disentangle these effects maximum likelihood and Bayesian estimation should be compared under alternative bid designs and distribution assumptions. It is well known that welfare estimates are sensitive to bid designs, and there could be scope for some trade-offs between improved bid designs and the use of more efficient estimation methods as explored earlier

in the context of maximum likelihood (Alberini 1995; Scarpa and Bateman 2000). Similarly, since the error in the distribution assumption might bias welfare estimates, it should also be explored whether more accurate estimation techniques are capable of overcoming misspecification bias. Some relevant works in this line are Kriström (1991), Li (1996), Chen and Randall (1997), Creel and Loomis (1997) and Araña and Leon (2005).

Appendix

A draw from the truncated normal distribution 2.4 can be obtained using the inverse distribution method proposed by Devroye (1986). By applying this method, the truncated normal distribution can be sampled as follows:

$$W_{i1} = \begin{cases} x_i'\beta + \sigma\Phi^{-1}(U_{11}) & \text{if } z_i y_{i1} = 1 \\ x_i'\beta + \sigma\Phi^{-1}(U_{12}) & \text{if } z_i(1-y_{i1}) = 1 \\ 0 & \text{if } z_i = 0 \end{cases} \tag{A1}$$

where U_{11} is generated from a uniform density distribution in interval $\left[\Phi\left(\frac{-x_i'\beta}{\sigma}\right), \Phi\left(\frac{B_{i1}-x_i'\beta}{\sigma}\right)\right]$, and U_{21} from a uniform density distribution in interval $\left[\left(\Phi\left(\frac{B_{i1}-x_i'\beta}{\sigma}\right), 1\right)\right]$. A draw from the truncated distribution in 2.8 can be obtained in a similar way,

$$W_{i2} = \begin{cases} x_i'\beta + \sigma\Phi^{-1}(U_{12}) & \text{if } z_i y_{i1} y_{i2} = 1 \\ x_i'\beta + \sigma\Phi^{-1}(U_{22}) & \text{if } z_i y_{i1}(1-y_{i2}) = 1 \\ x_i'\beta + \sigma\Phi^{-1}(U_{32}) & \text{if } z_i y_{i2}(1-y_{i1}) = 1 \\ x_i'\beta + \sigma\Phi^{-1}(U_{42}) & \text{if } z_i(1-y_{i1})(1-y_{i2}) = 1 \\ 0 & \text{if} z_i = 0 \end{cases} \tag{A2}$$

where U_{12} is generated from a uniform density distribution in the interval

$$\left[\left(\Phi\left(\frac{B_{i2}-x_i'\beta}{\sigma}\right), 1\right)\right];$$

U_{22} is generated by a uniform density distribution in the interval

$$\left[\left(\Phi\left(\frac{B_{i1}-x_i'\beta}{\sigma}\right), \Phi\left(\frac{B_{i2}-x_i'\beta}{\sigma}\right)\right)\right];$$

U_{32} is generated by a uniform density distribution in the interval

$$\left[\left(\Phi\left(\frac{B_{i2}-x_i'\beta}{\sigma}\right), \Phi\left(\frac{B_{i1}-x_i'\beta}{\sigma}\right)\right)\right];$$

and U_{42} is generated by a uniform density distribution in the interval

$$\left[0, \left(\Phi\left(\frac{B_{i2}-x_i'\beta}{\sigma}\right)\right)\right].$$

Chapter 12

MODELING ELICITATION EFFECTS IN CONTINGENT VALUATION STUDIES

A Monte Carlo Analysis of the Bivariate Approach

Margarita Genius
Department of Economics, University of Crete, Greece
genius@ermis.soc.uoc.gr

Elisabetta Strazzera
DRES and CRENoS, University of Cagliari, Cagliari, Italy
strazzera@unica.it

Abstract A Monte Carlo analysis is conducted to assess the validity of the bivariate modeling approach for detection and correction of different forms of elicitation effects in double bound contingent valuation data. Alternative univariate and bivariate models are applied to several simulated data sets, each one characterized by a specific elicitation effect, and their performance is assessed using standard selection criteria. The bivariate models include the standard bivariate probit model, and an alternative specification, based on the Copula approach to multivariate modeling, which is shown to be useful in cases where the hypothesis of normality of the joint distribution is not supported by the data. It is found that the bivariate approach can effectively correct elicitation effects while maintaining an adequate level of efficiency in the estimation of the parameters of interest.

Keywords: Double Bound, Elicitation effects, Bivariate models, Probit, Joe Copula

1. Introduction

In the conclusions of their extensive overview of the state of the art of the contingent valuation (CV) method, Carson, Flores and Meade (2001) remark that:

> " at this point in the development of CV, the key objective in terms of methodological development should shift to trying to determine how to reduce the cost

> of conducting CV studies while still maintaining most of the quality of the very best studies now being conducted."

Since costs are mainly driven by the survey administration, the crucial question is how to obtain valid and reliable estimates for the population *WTP* from smaller samples than those employed in benchmark high quality CV studies. A substantial stream of research has been conducted in the last decade aimed at finding an optimal method for elicitation of *WTP*, apt to combine the two properties of unbiasedness and statistical efficiency.

There seems to be a fairly general agreement in the literature (one notable exception is Green, Jacowitz, Kahneman and McFadden, 1998) that the single bound method is potentially valid in terms of incentive compatibility, i.e. in terms of incentive for the respondent to truthfully reveal her preferences. Unfortunately, the single bound method is inefficient in terms of information conveyed by the elicitation process, and small size surveys can be particularly affected by this problem. As is well known after the seminal paper by Hanemann, Loomis and Kanninen (1991), a follow-up question helps to improve efficiency of the estimates: the double bound procedure shows a dramatic increase in the precision of the estimates.[1] The problem is that, as pointed out by Carson, Groves and Machina (1999) and discussed more thoroughly in the following of this chapter, an iteration of the elicitation question may lead respondents to misrepresent their true preferences.

Yet, the double bound elicitation method may still be a preferred choice if the statistical analysis could detect the presence of elicitation effects in a reliable manner; and if, after the correction of the estimates for such effects, the statistical efficiency of the method were preserved. A recent trend is to base the statistical analysis of double bound data on the use of bivariate probit models. This line of research was initiated by Cameron and Quiggin (1994), who compare estimates from several competing specifications based on the bivariate probit to those from the univariate double bound model. The approach is further pursued by Alberini, Carson and Kanninen (1997), who use bivariate probit specifications to model different behavioral hypotheses induced by the reiteration of the *WTP* elicitation question. Such hypotheses are then tested as competing models by means of standard specification tests.

In a neoclassical theoretical framework, such as that considered by Carson, Groves and Machina, elicitation effects are deemed to affect the response to the second bid offer, no matter whether the observed response to the first bid is positive or negative. A different approach has been taken by DeShazo (2002), who theorizes that respondents who reject the first bid tender have no incentive to misrepresent their valuation when facing the second bid, while the opposite

[1]The addition of further follow-ups is relatively less beneficial in terms of efficiency gains, as shown by Scarpa and Bateman (2000).

holds for the others. For data affected by this form of distortion ("framing effects"), DeShazo proposes that only the single bound response should be used. If needed, information from the second elicitation question should be limited to the sub-sample of respondents who say "no" to the first price offer. In an application, DeShazo models the resulting censored data by means of a bivariate probit with sample selection.

The present chapter is aimed at assessing the validity of the bivariate approach to modeling double bound data for detection and correction of different forms of elicitation effects. The analysis is based on Monte Carlo methods, involving generation of several simulated data sets, each one characterized by a specific form of elicitation problem. Alternative univariate and bivariate models are applied to the simulated data, and their performance is assessed using standard selection criteria. The univariate models considered in the present work are the single bound and the double bound estimators, plus a univariate censored model, applied to data characterized by DeShazo's framing effects. The bivariate models include the standard Bivariate Probit model, based upon the Bivariate Normal distribution, and the bivariate probit with sample selection proposed by DeShazo.

It is well known that the normality assumption for the distribution of *WTP* (or its logarithm) is often not supported by the data. In these cases the use of a bivariate probit would result in biased estimates: we therefore extend our analysis to alternative bivariate models, namely Copula models, which are characterized by a great flexibility in the distributional shape of their marginals, and in their dependence structure. In particular, the Joe Copula, which is characterized by asymmetry and fat tails, is applied to the double bound data generated in our experiments, and its performance is compared to the models mentioned above.

The paper is organized as follows: in section 2 we examine the behavioral hypotheses underlying the decision to respond strategically to *WTP* questions. Section 3 presents the bivariate modeling approach to fit double bound data, which can be based on the conventional Bivariate Probit, or, alternatively, on Copula models, here introduced. Section 4 describes the experimental design of the Monte Carlo analysis: underlying behavioral assumptions, data construction, and statistical models applied to the simulated data; in section 5 we discuss results; section 6 concludes the paper.

2. Behavioral hypotheses

Carson, Groves and Machina (1999, henceforth CGM) set a theoretical framework to analyze the incentive properties of different formats for elicitation of *WTP* in contingent valuation surveys. They deem the single bound protocol as valid, as long as some conditions hold: the valuation procedure

should be presented as a referendum, involving the supply of a new public good with a coercive contingent payment (for example, a tax), or the choice between two new public goods. Moreover, the survey should be seen by respondents as *consequential*, i.e. having the potential effect of influencing the agency decisions.

Unfortunately, the incentive properties of the single bound are not shared by the double bound method. CGM list several reasons as to why a follow-up question may produce elicitation effects. For example, if the initial bid is considered by the respondent as informative about the real cost of provision of the good, the second bid may induce uncertainty about the cost distribution. For risk averse agents this would result in a lower reservation price (because of the risk premium), so that the *WTP* elicited after the second question would be shifted downward [2].

The same consequence on the final *WTP* value could also be produced by a strategic type of reaction to the new bid tender: agents may consider the iterative procedure as a bargaining game, where it would be strategically optimal not to reveal a higher *WTP* than the second price offered.

A third behavioral hypothesis considered by CGM is that individuals take the offered bids as informative data on the true cost of provision of the good: if, as is often the case in contingent valuation studies, the respondent does not possess some *a priori* information, the bid offered in the first elicitation question can be taken as conveying information on the price distribution of the good, and the second bid may be used as a Bayesian update of that information. CGM observe that this could have either positive or negative effects on the underlying *WTP* distribution, according to the sign of the answer to the first elicitation question. This problem, often referred to as "anchoring" or "starting point bias," is analyzed for applications to double bound data by Herriges and Shogren (1996), who propose an econometric model that incorporates a Bayesian updating mechanism to detect and correct anchoring bias. They show in an application that correcting for starting point bias reduces the difference between the double bound and the single bound estimator, both for point estimates and confidence intervals. Analogous results are obtained by McLeod and Bergland (1999) in two other applications of the updating model, and they conclude that :

> "the increased precision in the estimated *WTP* by asking a follow-up question is not as large, or even non-existent, when Bayesian updating is accounted for in the estimation".

[2] By the same argument, some doubts may be cast on the One and One Half bound method, recently proposed by Hanemann, Loomis and Signorello (2002), which confronts the individual with an initial bid distribution, rather than a single first bid.

This point of view seems largely subscribed by Whitehead (2002), but with an important qualification: the double bound model should be used, with appropriate corrections for anchoring or other kinds of bias, when the initial bids are poorly chosen, i.e. they do not represent adequately the population's *WTP* distribution. In such a case the double bound still allows relevant efficiency gains, even after controlling for elicitation effects.

The updating econometric models described above are specified as univariate distributions. Cameron and Quiggin (1994) argue that when the underlying *WTP* values from the two elicitation questions are different, the assumption of a unique distribution is unduly restrictive, and propose a bivariate modeling approach, based on the bivariate probit model, which is also applied by Alberini (1995) and Alberini, Carson and Kanninen (1997) to several data sets. In general, these papers empirically support the view that the two underlying *WTP* values obtained through the double bound method are not identical, and while for the most part the data seems unaffected by a systematic form of bias, some (Alaska study, in Alberini, Carson, Kanninen) show evidence of a downward shift in the *WTP* elicited through the follow-up question.

An alternative theoretical framework is proposed by DeShazo (2002). Building upon theoretical results from Prospect Theory (Kahneman and Tversky, 1979), he argues that strategic behavior may only emerge for respondents who answer "yes" to the first price offer, while incentives for the others would be unaltered across elicitation questions. The reason is that if the respondent's true *WTP* is above the first bid, she expects to gain some consumer surplus, which may be taken as a reference point. Conversely, no consumer's surplus is expected, and no reference point is created, by people whose *WTP* is below the first bid tender. Under the double bound protocol, a "yes" response to the first bid question leads to a higher second bid (ascending sequence), while the converse holds in the case of a negative answer (descending sequence). According to DeShazo's theory, ascending sequences are susceptible to strategic behavior, induced by the creation of a reference point (framing effect), which does not affect descending sequences. He devises a test to verify if such assumptions are tenable: if so, the suggestion is to use single bound data, or to use the double bound data from descending sequences only, using the single bound response for the rest of the sample.

3. Bivariate Models

The bivariate model for discrete dependent variables is a two-equation system:

$$\begin{aligned} Y_{1i} &= x_{1i}\prime\beta_1 + u_{1i} \\ Y_{2i} &= x_{2i}\prime\beta_2 + u_{2i}, \end{aligned} \tag{3.1}$$

the dichotomous dependent variables are $y_{1i} = 1$ if $Y_{1i} > 0$; $y_{2i} = 1$ if $Y_{2i} > 0$; $x_{.i}$ are vectors of exogenous variables; β_1, β_2 are vectors of unknown parameters; and u_{1i}, u_{2i} are error terms with zero means, variances σ_1^2, σ_2^2, marginal distribution functions F_1, F_2 and with a joint distribution function H.

The bivariate probit model as applied in the contingent valuation literature is defined by equation 3.1 and the following:

$$y_{1i} = 1 \text{ if } Y_{1i} > t_{1i}; y_{2i} = 1 \text{ if } Y_{2i} > t_{2i}$$

where t_1, t_2 are the bids proposed and H is a bivariate normal with zero vector of means, unit variances and correlation coefficient ρ. Denoting the bivariate normal with zero means, unit variances and correlation ρ by $\Phi(.,.,\rho)$ and defining $\frac{t_{1i}-x_{1i}\prime\beta_1}{\sigma_1}$ and $\frac{t_{2i}-x_{2i}\prime\beta_2}{\sigma_2}$, the log-likelihood for the bivariate probit is given by:

$$\begin{aligned}\ln L(\beta_1, \beta_2, \rho) = \sum_{i=1}^{n} \quad & (1-y_{1i})(1-y_{2i}) \ln\left[\Phi(a_{1i}, a_{2i}, \rho)\right] + \\ & y_{1i}y_{2i} \ln\left[\Phi(-a_{1i}, -a_{2i}, \rho)\right] + \\ & (1-y_{1i})y_{2i} \ln\left[\Phi(a_{1i}, -a_{2i}, -\rho)\right] + \\ & y_{1i}(1-y_{2i}) \ln\left[\Phi(-a_{1i}, a_{2i}, -\rho)\right]. \end{aligned} \tag{3.2}$$

The four terms in the log-likelihood correspond to "no-no", "yes-yes", "no-yes" and "yes-no" responses to the two bid tenders respectively.

The double bound model results from 3.2 if $\rho = 1$ and the parameters are the same across equations. If the two error terms are not perfectly correlated then the responses to the two bids are governed by a bivariate model, with parameters that may or may not be the same across equations.

The assumption of normality for the *WTP* distribution is often not supported by the data, and this may give rise to serious misspecification problems. When the model is univariate, the analyst can pick from a wide range of possible distributions the one that better fits the data. When the model is bivariate, the choice is usually constrained to the bivariate normal distribution, and a bivariate probit as above is applied. In practice, *WTP* is often assumed to have a log-normal distribution, which accounts for the skewness that generally characterizes *WTP* distributions, and the bivariate normal is applied to the logarithm of *WTP*. Unfortunately, also the log-normal assumption may not be supported by the data, which implies that the bivariate probit would not be a valid estimator (as also seen in Alberini, 1995): distributional misspecification of the marginals will, in general, result in inconsistent estimates of the parameters since it implies misspecification of the model for the conditional mean of the binary dependent variable (see Ruud, 1983). On the other hand, alternative bivariate distributions, such as the bivariate logistic or the bivariate extreme

value, are not as flexible, in terms of correlation allowed between marginals, as the bivariate normal.

As suggested by Hanemann and Kanninen (1999) a possible solution to the problem could be the following: even if the stochastic parts of the two equations are specified as non-normal, they can be transformed into random variables that are characterized by the bivariate normal distribution. This transform, which involves the use of the inverse standard normal distribution, is a special case of a bivariate copula function, and is known in econometrics after Lee's (1982, 1983) applications to sample selection models. A general definition for bivariate copulas is:

Definition: A *2-dimensional copula* is a function $C : [0,1]^2 \rightarrow [0,1]$, with the following properties:

(a) For every $u \in [0,1], C(0,u) = C(u,0) = 0$;

(b) For every $C(u,1) = u$ and $C(1,u) = u$;

(c) For every $(u_1,\nu_1),(u_2,\nu_2) \in [0,1] \times [0,1]$ with $u_1 \leq u_2$ and $\nu_1 \leq \nu_2$: $C(u_2,\nu_2) - C(u_2,\nu_1) - C(u_1,\nu_2) + c(u_1,\nu_1) \geq 0$.

The last condition is the two-dimensional analogue of a nondecreasing one-dimensional function.

The theoretical basis for multivariate modeling through copulas is provided by a theorem due to Sklar (1959).

Sklar's Theorem *Let H be a joint distribution function with margins F_1 and F_2, which are, respectively, the cumulative distribution functions of the random variables x_1 and x_2. Then, there exists a function C such that:*

$$H(x_1,x_2) = C\left(F_1(x_1),F_2(x_2)\right), \text{for every } \{x_1,x_2\} \in \bar{\mathbb{R}},$$

where $\bar{\mathbb{R}}$ represents the extended real line. Conversely, if C is a copula and F_1 and F_2 are distribution functions, then the function H defined above is a joint distribution function with margins F_1 and F_2.

Since the copula function "links a multidimensional distribution to its one-dimensional margins" (Sklar, 1996), the name "copula" (connection) is explained. The parametric copula approach ensures a high level of flexibility to the modeler, since the specification of the margins F_1 and F_2 can be separated from the specification of the dependence structure through the function C and an underlying parameter θ governing the intensity of the dependence.

Although the Lee copula allows flexibility in the choice of the margins, it maintains some restrictive properties (for example, symmetry) of elliptical distributions. More interesting for applied work is the class of Archimedean copulas. These are functions generated by an additive continuous, convex decreasing function φ, with $\varphi(1) = 0$. In general, Archimedean copulas have the following form:

$$\varphi(C_\theta(u,\nu)) = \varphi(u) + \varphi(\nu).$$

The additive structure of Archimedean copulas makes maximum likelihood estimation and calculation of the score function relatively easy. Furthermore, the family is sufficiently large so as to allow a wide range of distributional shapes (right or left skewness, fat or thin tails, etc.). A particular feature of most Archimedean copulas is monotonicity, i.e. they cannot accommodate negative dependence, and this may limit their application in some contexts. In the present application, where the margins represent the underlying *WTP* distributions elicited by the double bound method, it is realistic to exclude negative dependence, and use of Archimedean copulas is warranted. Specifically, drawing from previous work (Genius and Strazzera, 2004), we choose the Joe copula, which is defined as follows:

$$C(u,v) = 1 - \left((1-u)^{\theta} + (1-\nu)^{\theta} - (1-u)^{\theta}(1-\nu)^{\theta}\right)^{1/\theta}, \theta \in [1,\infty),$$

where u and ν are univariate distributions, and θ is a dependency parameter.

A relevant part of our analysis deals with the estimation of the dependency between equations. When dealing with elliptical copulas (such as the BVN, or the Lee Copula) a valid measure of dependence is linear correlation; however, this does not hold when the bivariate distribution is not elliptical (see figure 12.1 for a comparison of distributional shapes: the Joe copula is not elliptical). Alternative measures of dependence include Kendall's $\tau(K_{\tau})$ which is a measure of concordance. It is defined as follows:

$$K_{\tau} = \Pr\left((X-\tilde{X})(Y-\tilde{Y}) > 0\right) - \Pr\left((X-\tilde{X})(Y-\tilde{Y}) < 0\right),$$

where (X,Y) and $(\tilde{X},\tilde{Y})$ are two independent random vectors with a common distribution function H whose margins are F and G. Kendall's τ can also be expressed in terms of copulas (see Nelsen, 1999):

$$K_{\tau} = 4\int\int_{[0.1]^2} C(u,\nu)dC(u,\nu) - 1$$

For continuous random variables the above measure is a measure of concordance, which implies that it takes values in $[-1,1]$, taking the value zero when we have independence. We recall that the linear (or Pearson) correlation is not a measure of dependence: for example, $\rho(x,y) = 0$ does not imply independence of the two variables.

Since our estimations involve both elliptical (Normal) and not elliptical (Joe) bivariate distributions, for comparison purposes we report results for the Kendall's τ rather than for the correlation parameter or the dependence parameter θ.

4. Experimental Design

The experiments presented in this chapter are aimed at analyzing the performance of competing models when some specific forms of bias affect the

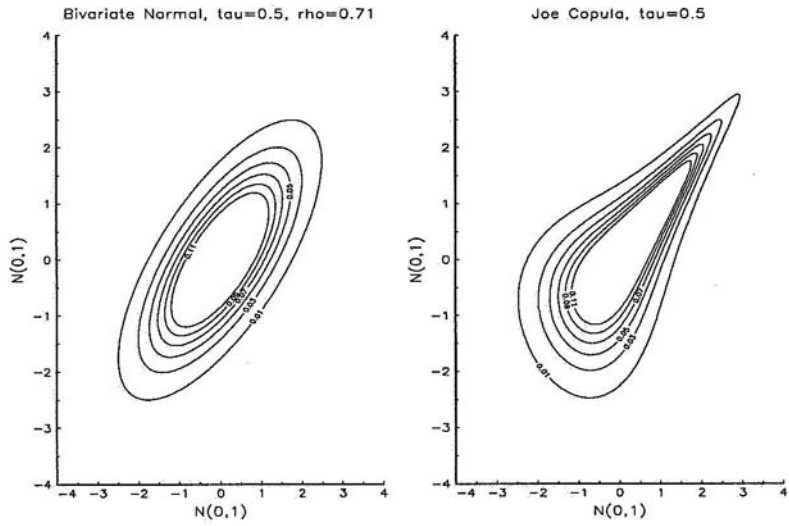


Figure 12.1. Contour plots of Bivariate Normal and Joe Copula when marginals are $N(0,1)$.

responses to the second bid question in the double bound format. Each form of response effect implies a specific bivariate data generating process (DGP) for the two *WTP*s. The first DGP we consider introduces a mild form of elicitation effect, where the underlying *WTP* elicited after the second question is the same as the first, but because of some disturbance in the elicitation process they are not perfectly correlated:

$$\begin{aligned} Y_{1i} &= x_i\prime\beta + u_{1i} \\ Y_{2i} &= x_i\prime\beta + u_{2i} \\ 0 < \rho &< 1 \end{aligned} \tag{4.1}$$

where ρ is the correlation parameter, and u_{1i} and u_{2i} are identically distributed random variables with mean zero, and variance σ^2. If the random errors are assumed to be distributed as a Normal, this specification gives rise to a BVN model, where parameters of the two equations are constrained to be equal, while the correlation parameter is unconstrained. In Cameron and Quiggin this was deemed as the best specification to fit their double bound data (a well known Australian study for the Kakadu area), which were mod-

eled by means of the univariate interval data (or double bound) model, and several alternative specifications of the bivariate probit model. The simulated data are constructed using the following specification: intercept parameter α =10, slope coefficient β =3, standard deviation σ =5, and a BVN distribution, with correlation ρ =0.7. The corresponding value of τ is given by $\tau = (2/\pi)\arcsin(\rho)$ =0.493. The variable x is generated from a uniform, with mean 3.95 and standard deviation 2.05.

The same BVN and interval data specifications used by Cameron and Quiggin were analyzed again by Alberini (1995) on a slightly different sample from the same data, and in this case the preferred specifications was as in 3.1, but with different estimated variances for the two latent dependent variables. The underlying behavioral hypothesis could be that the cognitive process after the second elicitation question is more "disturbed" — and indeed Alberini (1995) finds that the estimated standard deviation of the second equation is substantially higher than the first. This more general structure has been analyzed again in Alberini, Carson and Kanninen (1997). Our Monte Carlo analysis considers both cases, which can be respectively referred to as restricted and unrestricted random effects models, with an equality restriction imposed on the standard deviation parameters of the former (experiment A). In the latter DGP, the standard deviation of the second equation is set at 7 (experiment B).

Another experiment (C) studies the performance of different models when the double bound elicitation method produces more serious forms of bias, leading to a downward shift of the second equation *WTP*: CGM indicate several possible causes of this effect, briefly reported in section 2 of this chapter. In its simplest form the bivariate model with shift is structured as follows:

$$\begin{aligned} Y_{1i} &= x_i\prime\beta + u_{1i} \\ Y_{2i} &= \delta + x_i\prime\beta + u_{2i} \\ 0 < \quad \rho &< 1, \quad \delta < 0, \end{aligned} \tag{4.2}$$

i.e. the shift is simply a leftward translation of the *WTP* distribution. While more complex specifications may model the shift effect as dependent on some covariates, in our experiment we hold to this basic model, setting δ =–2, the other data being constructed as in experiment B.

The Framed model proposed by DeShazo is relatively new, and, to our knowledge, has not been as yet studied by means of Monte Carlo methods. Here the structure of the DGP is somewhat more complex, since the model involves a mechanism of sample selection. Theoretically responses to both questions should be dictated by model B but because of framing effects a percentage of respondents belonging to the "yes,yes" class produce responses in the "yes,no" class. In DeShazo's proposed method to estimate such data affected by framing effects, follow-up responses from individuals who faced a downward sequence of bids enter the second equation, while for individuals

facing an upward sequence only the first response is considered, as if it were a single bound elicitation. The bivariate model for descending sequences proposed by DeShazo is the following:

$$\begin{aligned} Y_{1i} &= x_i\prime\beta + u_{1i} \\ Y_{2i} &= x_i\prime\beta + u_{2i} \\ 0 < \quad \rho &< 1 \end{aligned} \tag{4.3}$$

where Y_2 is modeled for respondents in a descending sequence only (i.e. individuals who responded No to the first elicitation question). The parameter values are constructed as in experiment B above but we switch randomly a percentage of "yes,yes" responses to "yes,no". In order to evaluate the performance of the bivariate estimator versus its univariate counterpart, we also estimate a univariate Framed model: it is a censored double bound model, with the second bid included in the equation for respondents in the descending sequence only.

In another experiment (E) we study the performance of different estimators when the initial bids are poorly chosen: in particular, the initial bids of this experiment leave uncovered the left tail of the true *WTP* distribution, being placed at the 45, 65, and 85 percentile of the distribution. As discussed in section 2, this is a case where the double bound method is generally deemed superior to the single bound, since the follow-up question allows inspection of the previously excluded part of the distribution. All other aspects of the data generation process are as in experiment B.

Finally, in experiments F and G we analyze the case where the DGP departs from the Bivariate Normal distribution. *WTP* distributions are commonly specified as non normal: logistic or extreme value (for *WTP* or its log) or gamma distributions are typical choices in CV data analysis. Application of bivariate models to *WTP* data may induce a misspecification problem, if the standard bivariate probit is adopted. In this experiment we are especially interested in analyzing the performance of the bivariate probit estimator when the assumption of bivariate normality is wrong. We first simulate a bivariate distribution with normal marginals, but with a dependency structure different from linear correlation: i.e. the Joe copula distribution with normal marginals (JOE-N: experiment F). In experiment G, we use as a DGP a bivariate model with extreme value marginals, again linked by a Joe copula (JOE-E). The dependency parameter τ is set at the value 0.499 which corresponds to a value of θ equal to 2.85. All other settings are as in experiment B. The reader is referred to the Appendix for the algorithm used to generate the data by means of the JOE copula.

Given the above scenarios we analyze the performance of different estimators: the univariate single and double bound models; a univariate model for descending sequences (in experiment D only); the bivariate probit model;

the bivariate model based on the Joe Copula with normal marginals for experiments A-F; a bivariate Joe Copula with extreme value marginals for experiment G; and the bivariate probit with sample selection for descending sequences. The respective likelihoods for each experiment are given in Table 12.1, where $a_{1i} = \frac{t_{1i} - x_i\prime\beta}{\sigma}$ and $a_{2i} = \frac{t_{2i} - x_i\prime\beta_2}{\sigma_2}$.

Table 12.1. Likelihood of Estimated Models

MODELS	LIKELIHOOD TERMS: Experiments (A) to (F)	LIKELIHOOD TERMS: Experiment (G), where the cdf for extreme value is given by, $G(a) = \exp(\text{-}\exp(\text{-}a))$
SINGLE	P(yes): 1-$\Phi(a_{1i})$ P(no): $\Phi(a_{1i})$	P(yes): 1-$G(a_{1i})$ P(no): $G(a_{1i})$
DOUBLE	P(n,n): $\Phi(a_{2i})$ P(n,y): $\Phi(a_{1i})$-$\Phi(a_{2i})$ P(y,n): $\Phi(a_{2i})$-$\Phi(a_{1i})$ P(y,y): 1-$\Phi(a_{2i})$	P(n,n): $G(a_{2i})$ P(n,y): $G(a_{1i})$-$G(a_{2i})$ P(y,n): $G(a_{2i})$-$G(a_{1i})$ P(y,y): 1-$G(a_{2i})$
FRAMED1 (experiment D only)	P(n,n): $\Phi(a_{2i})$ P(n,y): $\Phi(a_{1i})$-$\Phi(a_{2i})$ P(yes): 1-$\Phi(a_{1i})$	P(n,n): $G(a_{2i})$ P(n,y): $G(a_{1i})$-$G(a_{2i})$ P(yes): 1-$G(a_{1i})$
BVN	P(n,n): $\Phi(a_{1i}, a_{2i}, \rho)$ P(n,y): $\Phi(a_{1i}$,-a_{2i},-$\rho)$ P(y,n): $\Phi($-a_{1i}, a_{2i},-$\rho)$ P(y,y): $\Phi($-a_{1i},-$a_{2i}, \rho)$	
JOE COPULA	P(n,n): $C(\Phi(a_{1i}), \Phi(a_{2i}), \theta)$ P(n,y): $\Phi(a_{1i})$-$C(\Phi(a_{1i}), \Phi(a_{2i}), \theta)$ P(y,n): $\Phi(a_{2i})$-$C(\Phi(a_{1i}), \Phi(a_{2i}), \theta)$ P(y,y): 1-P(n,n)-P(n,y)-P(y,n)	P(n,n): $C(G(a_{1i}), G(a_{2i}), \theta)$ P(n,y): $G(a_{1i})$-$C(G(a_{1i}), G(a_{2i}), \theta)$ P(y,n): $G(a_{2i})$-$C(G(a_{1i}), G(a_{2i}), \theta)$ P(y,y): 1-P(n,n)-P(n,y)-P(y,n)
FRAMED2	P(n,n): $\Phi(a_{1i}, a_{2i}, \rho)$ P(n,y): $\Phi(a_{1i}$,-a_{1i},$-\rho)$ P(yes): 1-$\Phi(a_{1i})$	P(n,n): $C(G(a_{1i}), G(a_{2i}), \theta)$ P(n,y): $G(a_{1i})$-$C(G(a_{1i}), G(a_{2i}), \theta)$ P(yes): 1-$G(a_{1i})$

5. Results

The results of the experiments are based on the parameter estimates of the models reported in Table 12.1, under the different scenarios considered in the

previous section. The Monte Carlo analysis involved 400 replications, and was performed using a Gauss 386-i code.[3]

The maximum likelihood estimation required imposition of constraints on the dependence parameters for the bivariate models: specifically, for bivariate models based on the bivariate normal distribution $-1 < \rho < 1$ is imposed while for bivariate models based on the Joe copula the constraint is given by $\theta \geq 1$. Summary statistics of these results are reported in Tables 12.2-12.8. In particular, we show results for bias, standard deviation and empirical size of the α, β, and σ estimates, and the bias and standard deviation of the dependency parameter τ. Bivariate models are in general susceptible to convergence problems, so we also report the number of successful replications: it will be seen that convergence failures can be a serious problem in some circumstances. In our study the definition of failure includes cases where the maximum number of iterations was exceeded and cases where the Hessian failed to invert. For the BVN model the first type of failure was very often associated with the estimate of ρ being on the boundary of the parameter space.

No equality constraints are imposed on the α, β, and σ coefficients across equations, since this would be the first approach that a modeler would take to check for those forms of elicitation bias that would result in different parameters across equations, in addition to imperfect correlation.

The results of experiment A show that even a minor elicitation effect, resulting in imperfectly correlated distributions of the error terms, produces some bias in the double bound estimates. On the other hand they are characterized by standard errors so small that the Mean Squared Error (MSE) criterion would rank the double bound estimator best for all sample sizes (as also found in Alberini, 1995).

However, it should be noted that the empirical size of the parameters of the double bound model worsens for large samples.

The BVN model, which is the correct specification, estimates correctly the correlation coefficient, and proves to be a valid instrument to detect an elicitation effect problem. For experiments A and B we also report the estimates obtained from the more parsimonious BVN model, where the α and β parameters are constrained to be equal across equations for both experiments while σ is constrained only in experiment A (so corresponding to the true DGP). It can be observed that the constrained BVN model brings a substantial gain in efficiency, such that the MSE of the BVN estimates gets close to that of the DOUBLE model. Unfortunately, the BVN model seems susceptible to convergence failure, especially in the large sample experiment. Using a Joe copula when estimating the bivariate model (JOE-N) produces good results in terms

[3] We consider sample sizes 200, 400 and 1,000 but report only the latter two in this chapter. Results for the 200 case are available from the authors upon request.

Table 12.2. Experiment A: $\alpha = \alpha_2 = 10$, $\beta = \beta_2 = 3$, $\sigma = \sigma_2 = 5$, $\rho = 0.7$, $\tau = 0.493$.

Size 400

	α	β	σ	α_2	β_2	σ_2	τ	reps
	–0.047	0.016	0.008	–	–	–	–	–
SINGLE	1.240	0.310	0.583	–	–	–	–	–
	(0.068)	(0.068)	(0.065)	–	–	–	–	–
	0.201	–0.060	–0.186	–	–	–	–	–
DOUBLE	0.705	0.174	0.297	–	–	–	–	–
	(0.083)	(0.078)	(0.135)	–	–	–	–	–
	–0.104	0.031	0.037	–0.169	0.041	0.113	–0.030	370
BVN	1.216	0.304	0.581	0.955	0.240	0.600	0.211	–
	(0.057)	(0.054)	(0.054)	(0.032)	(0.032)	(0.068)	–	–
	–0.107	0.026	0.056	–	–	–	–0.007	382
BVN	0.797	0.199	0.396	–	–	–	0.207	–
Restricted	(0.045)	(0.037)	(0.052)	–	–	–	–	–
	0.021	–0.008	–0.111	–0.059	0.003	–0.032	0.242	377
JOE-N	1.173	0.299	0.588	0.900	0.225	0.534	0.249	–
	(0.078)	(0.069)	(0.098)	(0.061)	(0.061)	(0.111)	–	–
	0.025	0.001	–0.049	-1.721	0.283	1.443	–0.184	344
FRAMED2	1.157	0.295	0.566	4.182	0.919	3.651	0.429	–
	(0.058)	(0.061)	(0.073)	(0.049)	(0.052)	(0.055)	–	

Size 1,000

	α	β	σ	α_2	β_2	σ_2	τ	reps
SINGLE	–0.018	–0.001	–0.034	–	–	–	–	–
	0.862	0.204	0.379	–	–	–	–	–
	(0.050)	(0.050)	(0.055)	–	–	–	–	–
DOUBLE	0.200	–0.055	–0.165	–	–	–	–	–
	0.490	0.112	0.171	–	–	–	–	–
	(0.110)	(0.105)	(0.160)	–	–	–	–	–
BVN	–0.072	0.012	–0.022	–0.068	0.017	0.060	0.027	357
	0.810	0.193	0.366	0.667	0.162	0.374	0.187	–
	(0.042)	(0.045)	(0.056)	(0.076)	(0.064)	(0.050)	–	–
BVN	–0.069	0.014	0.026	–	–	–	0.029	363
CONST.	0.557	0.130	0.252	–	–	–	0.180	–
	(0.066)	(0.066)	(0.055)	–	–	–	–	–
JOE-N	0.058	–0.020	–0.109	0.051	–0.019	–0.068	0.277	379
	0.791	0.193	0.388	0.640	0.156	0.342	0.217	–
	(0.061)	(0.071)	(0.110)	(0.090)	(0.087)	(0.108)	–	–
FRAMED2	0.004	–0.004	–0.053	–0.835	0.124	0.699	0.082	339
	0.857	0.202	0.377	2.072	0.507	1.675	0.372	–
	(0.050)	(0.050)	(0.068)	(0.032)	(0.050)	(0.053)	–	–

[a] The three values in each block are: Bias, Standard Deviation and Empirical Size (Nominal size 5%).
[b] Number of replications where convergence is achieved and the Hessian is invertible.

of relatively small bias and standard errors, but empirical sizes are higher than nominal, and the dependency parameter is overestimated. This is a general feature that can be observed in all experiments where the Joe copula is fitted instead of the correct BVN distribution. On the other hand, the copula model is often more robust to convergence problems, and may be useful for preliminary exploration of the data.

A much higher rate of failures than the BVN characterizes the Bivariate Framed model proposed by DeShazo, which we assess also in the experiments with "unframed" data. The convergence problem is observed in all experiments, including experiment D with "framed" data: this fragility may be seen as a limit to its practical usefulness, although, as is discussed below for other experiments, it generally performs quite well in the estimation of the first equation parameters.

In experiment B variances are different across equations, the second being higher than the first. The relative performance of the double bound model decreases compared to experiment A and the empirical size is above 5% (reaching 0.998 for σ in the larger sample). Overall, the estimates of the bivariate models for the dependence and the first equation parameters are more correct and precise than the corresponding estimates in experiment A and are superior to the single bound model estimates in terms of the MSE criterion. As in the previous experiment, some convergence problems are observed, but they seem to reduce when the constrained model is estimated. Again, it can be noticed how the adoption of the constrained BVN model can effectively correct elicitation effects, while maintaining a high precision in the estimates of the first equation parameters.

Experiment C simulates a response bias characterized by a downward shift, which means that the intercept in the second *WTP* equation is smaller; and, as in B, by higher variance in the same equation. As discussed in section 2, a downward shift of the *WTP* distribution elicited after the second bid offer can be due to strategic behavior or risk aversion, behavioral attitudes that can be thought of as heterogeneously varied among individuals, giving rise to higher disturbance in the second *WTP* model. The double bound estimates are now more evidently biased, as it can also be seen from the empirical size values, which in the large sample are 32% for the constant, and 99.3% for the scale parameter. Both the bivariate probit and the JOE-N model perform quite well in the estimates of the first equation, and are superior to the SINGLE in terms of MSE for the relevant parameters.

As expected, given the data generating process used for this experiment, the BVN is superior to the copula model in estimating the dependence parameter, but it is more susceptible to convergence problems, and such vulnerability seems independent of sample size. As usual, the FRAMED2 model is even more fragile in this respect than the others, with 26% of the replications lost

Table 12.3. Experiment B: $\alpha = \alpha_2 = 10$, $\beta = \beta_2 = 3$, $\sigma = 5$, $\sigma_2 = 7$, $\rho = 0.7$, $\tau = 0.493$

Size 400

	α	β	σ	α_2	β_2	σ_2	τ	reps
SINGLE	–0.047	0.016	0.008	–	–	–	–	–
	1.240	0.309	0.583	–	–	–	–	–
	(0.068)	(0.068)	(0.065)	–	–	–	–	–
DOUBLE	0.313	–0.097	0.809	–	–	–	–	–
	0.796	0.195	0.344	–	–	–	–	–
	(0.1)	(0.115)	(0.678)	–	–	–	–	
BVN	–0.073	0.021	0.005	–0.173	0.040	0.177	0.005	362
	1.192	0.308	0.557	1.195	0.304	1.210	0.171	–
	(0.058)	(0.061)	(0.058)	(0.044)	(0.033)	(0.127)	–	–
BVN	–0.077	0.016	0.020	–	–	0.128	–0.002	372
Restricted	0.961	0.240	0.474	–	–	1.010	0.156	–
	(0.046)	(0.056)	(0.046)	–	–	(0.116)	–	–
JOE-N	0.145	–0.032	–0.074	–0.097	0.003	–0.087	0.255	359
	1.136	0.292	0.524	1.500	0.387	1.693	0.250	–
	(0.089)	(0.095)	(0.072)	(0.081)	(0.075)	(0.301)	–	–
FRAMED2	–0.062	0.019	–0.022	–1.346	0.216	1.639	–0.101	340
	1.263	0.315	0.576	3.811	1.018	4.358	0.309	–
	(0.059)	(0.065)	(0.068)	(0.071)	(0.071)	(0.103)	–	–

Size 1,000

	α	β	σ	α_2	β_2	σ_2	τ	reps
	–0.018	–0.001	–0.034	–	–	–	–	–
SINGLE	0.862	0.204	0.379	–	–	–	–	–
	(0.050)	(0.050)	(0.055)	–	–	–	–	–
	0.259	–0.075	0.837	–	–	–	–	–
DOUBLE	0.549	0.124	0.219	–	–	–	–	–
	(0.108)	(0.100)	(0.998)	–	–	–	–	–
	–0.063	0.011	–0.011	–0.108	0.026	0.081	0.011	367
BVN	0.819	0.194	0.355	0.837	0.201	0.772	0.133	–
	(0.049)	(0.046)	(0.057)	(0.06)	(0.044)	(0.125)	–	–
	–0.064	0.012	–0.009	–	–	0.047	0.007	375
BVN	0.643	0.148	0.310	–	–	0.623	0.120	–
Restricted	(0.059)	(0.051)	(0.053)	–	–	(0.096)		–
	0.185	–0.046	–0.069	0.102	–0.044	–0.372	0.313	371
JOE-N	0.762	0.185	0.345	1.094	0.261	1.126	0.208	–
	(0.078)	(0.086)	(0.078)	(0.119)	(0.127)	(0.553)	–	–
	0.001	–0.006	–0.050	–0.654	0.083	0.750	–0.050	343
FRAMED2	0.864	0.202	0.374	2.209	0.566	2.510	0.259	–
	(0.064)	(0.067)	(0.052)	(0.087)	(0.067)	(0.108)	–	–

[a] The three values in each block are: Bias, Standard Deviation and Empirical Size (Nominal size 5%).
[b] Number of replications where convergence is achieved and the Hessian is invertible.

in the large sample experiment. It can also be observed that the shift has produced an increase in bias for the estimates of the second equation, especially for FRAMED2.

Table 12.4. Experiment C: α_1 =10, α_2 =8, $\beta = \beta_2$ =3, σ =5, σ_2 =7,ρ =0.7, τ =0.493

Size 400

	α	β	σ	α_2	β_2	σ_2	τ	reps
SINGLE	–0.047	0.016	0.008	–	–	–	–	–
	1.240	0.309	0.583	–	–	–	–	–
	(0.068)	(0.068)	(0.065)	–	–	–	–	–
DOUBLE	–0.789	–0.099	0.802	–	–	–	–	–
	0.849	0.199	0.361	–	–	–	–	–
	(0.130)	(0.105)	(0.648)	–	–	–	–	–
BVN	–0.087	0.031	0.031	–0.206	0.039	0.150	0.001	348
	1.190	0.300	0.571	1.619	0.340	1.099	0.154	–
	(0.049)	(0.060)	(0.057)	(0.052)	(0.049)	(0.083)	–	–
JOE-N	0.005	–0.008	–0.082	0.167	–0.037	–0.162	0.242	371
	1.162	0.299	0.561	2.134	0.433	1.632	0.233	–
	(0.092)	(0.094)	(0.097)	(0.143)	(0.129)	(0.332)	–	–
FRAMED2	–0.04	0.015	–0.032	–1.528	0.088	4.411	–0.097	309
	1.234	0.307	0.577	4.419	0.834	4.173	0.312	–
	(0.071)	(0.068)	(0.065)	(0.104)	(0.068)	(0.123)	–	–

Size 1,000

	α	β	σ	α_2	β_2	σ_2	τ	reps
SINGLE	–0.018	–0.001	–0.034	–	–	–	–	–
	0.862	0.204	0.379	–	–	–	–	–
	(0.050)	(0.050)	(0.055)	–	–	–	–	–
DOUBLE	–0.869	–0.072	0.841	–	–	–	–	–
	0.590	0.129	0.220	–	–	–	–	–
	(0.320)	(0.110)	(0.993)	–	–	–	–	–
BVN	–0.068	0.009	0.005	–0.093	0.014	0.086	0.005	348
	0.841	0.198	0.355	1.060	0.209	0.731	0.123	–
	(0.052)	(0.055)	(0.057)	(0.063)	(0.055)	(0.086)	–	–
JOE-N	0.004	–0.016	–0.042	0.434	–0.077	–0.325	0.266	381
	0.798	0.188	0.347	1.474	0.278	1.186	0.206	–
	(0.063)	(0.058)	(0.089)	(0.228)	(0.189)	(0.430)	–	–
FRAMED2	0.017	–0.006	–0.041	–1.310	0.151	1.177	–0.074	295
	0.838	0.199	0.370	3.713	0.678	3.317	0.247	–
	(0.051)	(0.044)	(0.061)	(0.105)	(0.088)	(0.136)	–	–

[a] The three values in each block are: Bias, Standard Deviation and Empirical Size (Nominal size 5%).
[b] Number of replications where convergence is achieved and the Hessian is invertible.

The data created for experiment D incorporate framing effects in some of the second responses. In this case it could very well happen that the joint distri-

Table 12.5. Experiment D: $\alpha = \alpha_2$ =10, $\beta = \beta_2$ =3, σ =5, σ_2 =7,ρ =0.7,τ =0.493

Size 400

	α	β	σ	α_2	β_2	σ_2	τ	reps
	–0.047	0.016	0.008	–	–	–	–	–
SINGLE	1.240	0.309	0.583	–	–	–	–	–
	(0.068)	(0.678)	(0.065)	–	–	–	–	–
	1.884	–0.904	0.134	–	–	–	–	–
DOUBLE	0.663	0.149	0.260	–	–	–	–	–
	(0.770)	(-1.000)	(0.073)	–	–	–	–	–
	0.210	–0.030	0.598	–	–	–	–	–
FRAMED1	0.910	0.258	0.461	–	–	–	–	–
	(0.080)	(0.053)	(0.218)	–	–	–	–	–
	–0.067	0.032	0.077	-6.588	–0.375	13.738	–0.666	337
BVN	1.186	0.300	0.577	5.453	0.792	9.002	–	–
	(0.056)	(0.053)	(0.039)	(0.018)	(0.166)	(0.018)	–	–
	–0.014	0.002	–0.024	–1.792	–0.854	5.146	–0.385	395
JOE-N	1.236	0.315	0.585	2.214	0.368	2.876	–	–
	(0.078)	(0.084)	(0.084)	(0.048)	(0.570)	(0.638)	–	–
	–0.055	0.019	–0.038	–1.362	0.211	1.612	–0.107	342
FRAMED2	1.259	0.314	0.571	3.800	1.018	4.320	–	–
	(0.061)	(0.067)	(0.067)	(0.064)	(0.067)	(0.096)	–	–

Size 1,000

	α	β	σ	α_2	β_2	σ_2	τ	reps
	–0.018	–0.001	–0.034	–	–	–	–	–
SINGLE	0.862	0.204	0.379	–	–	–	–	–
	(0.050)	(0.050)	(0.055)	–	–	–	–	–
	1.964	–0.844	0.058	–	–	–	–	–
DOUBLE	0.458	0.096	0.152	–	–	–	–	–
	(0.985)	(-1.000)	(0.035)	–	–	–	–	–
	0.197	–0.036	0.587	–	–	–	–	–
FRAMED1	0.600	0.155	0.262	–	–	–	–	–
	(0.083)	(0.075)	(0.578)	–	–	–	–	–
	–0.120	0.04	0.071	-5.581	–0.428	12.395	–0.670	328
BVN	0.888	0.209	0.395	4.060	0.537	6.648	–	–
	(0.052)	(0.055)	(0.049)	(0.076)	(0.308)	(0.598)	–	–
	–0.003	–0.005	–0.041	–1.062	–0.865	4.307	–0.456	392
JOE-N	0.858	0.206	0.382	1.229	0.222	1.371	–	–
	(0.054)	(0.054)	(0.061)	(0.069)	(0.949)	(0.980)	–	–
	–0.022	0.001	–0.03	–0.609	0.076	0.734	–0.050	341
FRAMED2	0.865	0.205	0.383	2.117	0.547	2.531	–	–
	(0.056)	(0.065)	(0.050)	(0.080)	(0.067)	(0.106)	–	–

[a] The three values in each block are: Bias, Standard Deviation and Empirical Size (Nominal size 5%).
[b] Number of replications where convergence is achieved and the Hessian is invertible.

bution of observed responses has changed with respect to the joint distribution of "unframed responses" or that the dependence structure has changed. The average estimate of Kendall's τ is negative for the BVN model (the average ρ estimate is also negative), and smaller than in previous experiments for the JOE-N — while this does not apply to the sample selection FRAMED2 model.

Two salient features of this experiment are that now the DOUBLE model is the worst one in terms of MSE for α and β, with empirical sizes often above 98% and that the estimate of τ is highly biased especially for BVN. For this experiment we tested the univariate censored model FRAMED1 described in the previous section: although its estimates for α and β are satisfactory, we observe that the estimate of σ is strongly biased, and that empirical sizes increase with sample size and are beyond 5%. All bivariate models perform quite well in estimating the first equation parameters, with JOE-N generally slightly better than BVN, which may suggest that the underlying distribution of observed responses might have changed. More importantly, the rate of successful replications of the JOE-N model in experiment D, compared to that of the other two bivariate models confirms the robustness of this copula model, and its usefulness at least at the exploratory stage of the analysis.

In experiment E the initial bids have been changed, while it is maintained the behavioral hypothesis underlying experiment B, resulting in imperfect correlation of the *WTP* equations plus different associated standard deviations. Except for the DOUBLE, whose results are fairly close to the results in experiment B, the estimates of all other models have worsened in terms of bias, inefficiency and empirical sizes for all models, especially for the small sample case. The JOE-N model performs relatively better than the others, even though its empirical sizes tend to increase for larger samples. In any case, all bivariate models, and especially the correctly specified BVN, give a good estimate of the dependence parameter, and therefore are able to signal the presence of some response effect. The DOUBLE model is more robust to poorly chosen starting bids but it is biased, as can be seen from the estimate of the scale parameter which remains biased for large samples.

In experiments F and G we analyze two cases of departure from the assumption of bivariate normality for the *WTP* distributions. In F the data are generated by a bivariate distribution derived from a Joe copula with normal marginals, while in G the Joe copula links two extreme value distributions. All other settings are as in experiment B, with the qualification that the correct specification now is the JOE copula: in experiment F the Joe copula with normal marginals (JOE-N), in experiment G the Joe copula with extreme value marginals (JOE-E). It can be seen from Tables 12.7 and 12.8 that the two copula models estimate correctly all the parameters. While the BVN performs relatively well with normal marginals, its application to a bivariate with non normal marginals results in biased estimates of the first equation parameters.

Table 12.6. Experiment E: $\alpha = \alpha_2 = 10$, $\beta = \beta_2 = 3$, $\sigma = 5$, $\sigma_2 = 7$, $\rho = 0.7$, $\tau = 0.493$

Size 400

	α	β	σ	α_2	β_2	σ_2	τ	reps
	–0.189	0.044	0.030	–	–	–	–	–
SINGLE	1.802	0.375	0.653	–	–	–	–	–
	(0.040)	(0.038)	(0.053)	–	–	–	–	–
	0.276	–0.058	0.914	–	–	–	–	–
DOUBLE	0.876	0.193	0.367	–	–	–	–	–
	(0.073)	(0.073)	(0.770)	–	–	–	–	–
	–0.280	0.063	0.051	–0.485	0.104	0.402	–0.022	366
BVN	1.711	0.356	0.636	1.807	0.392	1.644	0.211	–
	(0.041)	(0.044)	(0.050)	(0.052)	(0.057)	(0.010)	–	–
	0.065	–0.004	–0.021	–0.405	0.071	0.187	0.265	376
JOE-N	1.608	0.339	0.628	2.693	0.555	2.553	0.254	–
	(0.074)	(0.072)	(0.080)	(0.090)	(0.093)	(0.253)	–	–
	–0.160	0.044	–0.008	–1.230	0.178	1.255	–0.130	356
FRAMED2	1.798	0.373	0.660	3.955	1.005	3.748	0.306	–
	(0.053)	(0.048)	(0.059)	(0.062)	(0.065)	(0.070)	–	–

Size 1,000

	α	β	σ	α_2	β_2	σ_2	τ	reps
	–0.036	0.004	–0.010	–	–	–	–	–
SINGLE	1.250	0.252	0.450	–	–	–	–	–
	(0.043)	(0.040)	(0.053)	–	–	–	–	–
	0.260	–0.054	0.936	–	–	–	–	–
DOUBLE	0.597	0.132	0.231	–	–	–	–	–
	(0.090)	(0.098)	(0.990)	–	–	–	–	–
	–0.070	0.010	0.009	–0.231	0.047	0.249	0.007	342
BVN	1.199	0.243	0.436	1.081	0.234	0.989	0.177	–
	(0.067)	(0.053)	(0.050)	(0.044)	(0.044)	(0.155)	–	–
	0.220	–0.044	–0.054	–0.014	–0.007	–0.136	0.296	375
JOE-N	1.147	0.233	0.413	1.528	0.317	1.535	0.232	–
	(0.080)	(0.077)	(0.075)	(0.109)	(0.123)	(0.469)	–	–
	0.022	–0.007	–0.050	–0.742	0.106	0.730	–0.065	343
FRAMED2	1.216	0.245	0.434	2.282	0.558	2.097	0.274	–
	(0.044)	(0.044)	(0.055)	(0.032)	(0.032)	(0.050)	–	–

[a] The three values in each block are: Bias, Standard Deviation and Empirical Size (Nominal size 5%).
[b] Number of replications where convergence is achieved and the Hessian is invertible.

Table 12.7. Experiment F: $\alpha = \alpha_2 = 10, \beta = \beta_2 = 3, \sigma = 5, \sigma_2 = 7, \rho = 0.7, \tau = 0.499$

Size 400

	α	β	σ	α_2	β_2	σ_2	τ	reps
	–0.107	0.028	0.008	–	–	–	–	–
SINGLE	1.239	0.306	0.589	–	–	–	–	–
	(0.053)	(0.055)	(0.070)	–	–	–	–	–
	0.371	–0.098	0.797	–	–	–	–	–
DOUBLE	0.837	0.201	0.366	–	–	–	–	–
	(0.088)	(0.110)	(0.635)	–	–	–	–	–
	–0.224	0.038	0.039	–0.179	0.091	0.470	–0.049	379
BVN	1.211	0.298	0.558	1.391	0.352	1.326	0.169	–
	(0.050)	(0.055)	(0.050)	(0.050)	(0.045)	(0.077)	–	–
	–0.132	0.031	0.003	–0.259	0.074	0.269	0.026	397
JOE-N	1.205	0.299	0.554	1.425	0.362	1.280	0.209	–
	(0.060)	(0.073)	(0.060)	(0.055)	(0.053)	(0.101)	–	–
	–0.055	0.015	–0.017	-3.360	0.527	3.305	–0.358	304
FRAMED2	1.279	0.316	0.596	4.573	1.097	5.137	0.342	–
	(0.0658)	(0.069)	(0.082)	(0.039)	(0.023)	(0.069)	–	–

Size 1,000

	α	β	σ	α_2	β_2	σ_2	τ	reps
	–0.035	0.010	0.015	–	–	–	–	–
SINGLE	0.859	0.209	0.371	–	–	–	–	–
	(0.053)	(0.050)	(0.043)	–	–	–	–	–
	0.377	–0.084	0.834	–	–	–	–	–
DOUBLE	0.498	0.121	0.221	–	–	–	–	–
	(0.113)	(0.113)	(0.98)	–	–	–	–	–
	–0.192	0.026	0.030	0.018	0.043	0.253	–0.027	364
BVN	0.830	0.202	0.362	0.844	0.215	0.924	0.138	–
	(0.047)	(0.055)	(0.049)	(0.058)	(0.047)	(0.113)	–	–
	–0.078	0.018	0.012	–0.123	0.039	0.169	0.009	398
JOE-N	0.798	0.194	0.355	0.822	0.203	0.770	0.155	–
	(0.043)	(0.053)	(0.048)	(0.038)	(0.050)	(0.090)	–	–
	–0.034	0.007	0.015	-4.327	0.743	4.405	–0.415	305
FRAMED2	0.864	0.206	0.372	8.079	1.808	9.123	0.257	–
	(0.056)	(0.039)	(0.052)	(0.056)	(0.036)	(0.059)	–	–

[a] The three values in each block are: Bias, Standard Deviation and Empirical Size (Nominal size 5%).
[b] Number of replications where convergence is achieved and the Hessian is invertible.

Table 12.8. Experiment G: $\alpha = \alpha_2 = 10, \beta = \beta_2 = 3, \sigma = 5, \sigma_2 = 7, \tau = 0.499$

Size 400

	α	β	σ	α_2	β_2	σ_2	τ	reps
	–0.070	0.019	–0.027	–	–	–	–	–
SINGLE	1.177	0.286	0.596	–	–	–	–	–
	(0.065)	(0.055)	(0.065)	–	–	–	–	–
	0.224	–0.073	0.937	–	–	–	–	–
DOUBLE	0.795	0.186	0.400	–	–	–	–	–
	(0.075)	(0.098)	(0.715)	–	–	–	–	–
	–0.463	0.039	–0.315	–0.950	0.154	–0.100	0.012	343
BVN	1.183	0.291	0.538	1.312	0.312	1.003	0.166	–
	(0.055)	(0.052)	(0.149)	(0.052)	(0.044)	(0.195)	–	–
	–0.090	0.017	–0.059	–0.223	0.067	0.196	0.039	394
JOE-EXT	1.135	0.271	0.539	1.142	0.295	0.989	0.225	–
	(0.069)	(0.056)	(0.053)	(0.056)	(0.051)	(0.084)	–	–
	–0.008	0.005	–0.037	–1.466	0.342	0.804	0.023	345
FRAMED2	1.155	0.274	0.588	2.950	0.889	2.579	0.220	–
	(0.070)	(0.052)	(0.064)	(0.023)	(0.038)	(0.049)	–	–

Size 1,000

	α	β	σ	α_2	β_2	σ_2	τ	reps
	0.017	–0.003	0.012	–	–	–	–	–
SINGLE	0.794	0.185	0.382	–	–	–	–	–
	(0.065)	(0.050)	(0.048)	–	–	–	–	–
	0.212	–0.058	0.989	–	–	–	–	–
DOUBLE	0.501	0.116	0.240	–	–	–	–	–
	(0.063)	(0.080)	(0.992)	–	–	–	–	–
	–0.299	–0.019	–0.350	–0.588	0.067	–0.398	0.080	360
BVN	0.729	0.178	0.325	0.767	0.181	0.624	0.164	–
	(0.056)	(0.067)	(0.264)	(0.081)	(0.039)	(0.419)	–	–
	–0.037	0.004	–0.009	–0.068	0.025	0.077	0.041	393
JOE-EXT	0.736	0.174	0.359	0.707	0.181	0.621	0.182	–
	(0.056)	(0.051)	(0.059)	(0.038)	(0.048)	(0.092)	–	–
	0.047	–0.008	–0.001	–0.978	0.169	0.466	–0.007	352
FRAMED2	0.781	0.184	0.381	1.799	0.471	1.367	0.219	–
	(0.060)	(0.054)	(0.048)	(0.006)	(0.031)	(0.028)	–	–

[a] The three values in each block are: Bias, Standard Deviation and Empirical Size (Nominal size 5%).
[b] Number of replications where convergence is achieved and the Hessian is invertible.

The bad performance of the BVN in this experiment is particularly evident for smaller sample sizes, but it can also be observed that the bias of the variance estimate increases with sample size. Note also that in experiment G the BVN is at high risk of convergence failure: over 10% of the replications for the large sample size are lost.

6. Conclusions

The need to obtain valid and reliable estimates for the value of non market goods under tighter budget constraints than those characterizing most benchmark CV studies has spurred research to a quest for an elicitation method that maximizes the amount of information obtained from respondents. More efficient elicitation methods allow smaller sample sizes, and this in turn results in less expensive surveys. The double bound method meets this efficiency requirement, but at the cost of potentially inducing elicitation effects, and hence unreliable estimates; on the other hand, the single bound method may not induce response effects, but it is statistically inefficient, and requires large samples in order to give reliable estimates. A solution could be at hand if appropriate statistical analysis could enable the modeler to detect and correct response bias from double bound data, so preserving the property of unbiasedness of the single bound method, and efficiency of the double bound. In this paper we have analyzed the performance of the bivariate modeling approach to the treatment of double bound data affected by elicitation bias. Several experiments were conducted, addressing different types of elicitation problems, and alternative distributional assumptions. We find that the bivariate approach can effectively detect the presence of elicitation effects, and produce correct estimates while maintaining a satisfactory level of efficiency. We have shown that application of a correctly specified bivariate model can lead to efficiency levels close to those reported for the double bound model. Clearly, the bivariate approach should be based on well specified models, since application of, say, a bivariate probit to non normal bivariate distributions would result in biased estimates. This was clearly shown in one experiment, where the bivariate probit was applied to a bivariate distribution with extreme value marginals.

We propose the adoption of copula models as a flexible instrument to fit non-normal bivariate distributions. Specifically, in this analysis we adopted the Joe copula, which turns out to be quite robust to convergence problems, and therefore can be used in preliminary analysis of the data to check for presence of response effects.

Finally, we found that convergence problems often affect the bivariate probit model, and even more seriously the bivariate probit with sample selection proposed by DeShazo to model data affected by framing effects. We showed that even when data are characterized by this type of bias, the sample selection

model does not perform better than the bivariate models with no selection, especially the Joe copula: it is possible that framing produces some effect on the dependency structure, which can be better modeled by asymmetric bivariate distributions, such as the Joe copula.

Obviously, our results only apply to the forms of response bias considered in the present work: further research is called for to explore the effectiveness of the bivariate approach for correction of other relevant sources of bias, such as anchoring effects.

Appendix

The algorithm below can be used to generate pairs of pseudo random numbers with arbitrary marginal distribution functions but with a Joe copula as a joint distribution function: see Embrechts *et al.* (2003). The Joe copula is defined by,

$$C(u,\nu) = 1 - \left[(1-u)^\theta + (1-\nu)^\theta - (1-u)^\theta(1-\nu)^\theta\right]^{1/\theta}, \theta \in [1,\infty)$$

and has a generator given by $\varphi(t) = -\ln[1-(1-t)^\theta]$. For any Archimedean copula the function K_c defined as, $K_c(t) = t - \varphi(t)/\varphi'(t)$ is the distribution function of the copula C. The algorithm goes through the following steps:

1 draw a pair of independent random variables $s, q \sim U[0,1]$.

2 choose the value of θ and set $t = K_c^{-1}(q)$. Since there is no closed form expression for the inverse of K_c, the equation $t - \varphi(t)/\varphi'(t) - q = 0$ has to be solved numerically using a root finding procedure.

3 set $u = \varphi^{-1}[s\varphi(t)]$ and $\nu = \varphi^{-1}[(1-s)\varphi(t)]$, where $\varphi^{-1}(t) = 1-(1-e^{-t})^{1/\theta}$ for the Joe copula. The pseudo random numbers u, ν are uniformly distributed on [0,1] and have a Joe copula as a joint distribution function.

4 for arbitrary distribution functions F_1 and F_2 define,

$$r_1 = F_1^{-1}(u) \text{ and } r_2 = F_2^{-1}(\nu).$$

The pseudo random numbers r_1, r_2 have marginal distributions given by F_1 and F_2 respectively while their joint distribution is given by a Joe copula. If we use the inverse normal transformation in both cases above, then we have generated pseudo random numbers with normal marginal distributions and a Joe copula as a joint distribution. If the inverse of the extreme value distribution is used instead, then the pseudo random numbers have extreme value marginals with a Joe copula as a joint distribution.

Chapter 13

PERFORMANCE OF ERROR COMPONENT MODELS FOR STATUS-QUO EFFECTS IN CHOICE EXPERIMENTS

Riccardo Scarpa*
Environment Department
University of York
York, U.K.
rs24@york.ac.uk

Silvia Ferrini
Department of Statistics
University of Florence, Florence, Italy
ferrini@ds.unifi.it

Kenneth Willis
School of Architecture and Planning
University of Newcastle University
Newcastle upon Tyne, U.K.
kenneth.willis@newcastle.ac.uk

Abstract Environmental economists have advocated the use of choice modelling in environmental valuation. Standard approaches employ choice sets including one alternative depicting the status-quo, yet the effects of explicitly accounting for systematic differences in preferences for non status-quo alternatives in the econometric models are not well understood. We explore three different ways of addressing such systematic differences using data from two choice modelling

*We acknowledge constructive comments on previous drafts by Giovanni Baiocchi, Elisabetta Cherchi, Joseph Herriges, Stephane Hess, Andreas Kontoleon, Luis Rizzi, two anonymous reviewers and participants in the Budapest 2004 EAERE meeting. All remaining errors are entirely our responsibility.

studies designed to value the provision of environmental goods. Preferences for change versus status-quo are explored with standard conditional logit with alternative-specific constant for status-quo, nested logit and a less usual mixed logit error component specification (kernel logit). Our empirical results are consistent with the hypothesis that alternatives offering changes from status-quo do not share the same preference structure as status-quo alternatives, as found by others in the marketing literature, in the environmental economic literature and in food preference studies. To further explore the empirical consequences of such mis-specification we report on a series of Monte Carlo experiments. Evidence from the experiments indicates that the expected bias in estimates ignoring the status-quo effect is substantial, and—more interestingly—that error component specifications with status-quo alternative specific-constant are efficient even when biased. These findings have significant implications for practitioners and their stance towards the strategies for the econometric analysis of choice modelling data for the purpose of valuation.

Keywords: choice-modelling, stated-preference, environmental valuation, status-quo bias, Monte Carlo simulations, water resources.

1. Introduction

Since their early appearance in the environmental economics literature in the middle-to-late nineties (Roe *et al.* 1996; Boxall *et al.*, 1996; Garrod and Willis, 1997; Adamowicz *et al.*, 1998) "choice experiments"[1] have enriched and further diversified the non-market valuation applications based on stated preferences. The number of studies on this methodology has been rapidly growing (Layton 2000; Morrison *et al.*, 2002; Foster and Mourato, 2002; Garrod *et al.* 2002) with applications covering many non-market valuation contexts. Overall, the role of this approach in diversifying the field of non-market valuation has been eloquently praised (Randall, 1998).

The basic method requires respondents to indicate a preference ordering — by ranking, rating or identifying a preferred choice — over a set of experimentally designed alternatives. Although the inclusion or exclusion of the status-quo (henceforth abbreviated in SQ) in the choice-set depends on the objective of the survey (see Breffle and Rowe, 2002 for a discussion), to increase realism (Ortúzar and Willumsen, 2001) most studies in transportation and environmental economics are based on survey designs that include a SQ alternative. This is often described to respondents in terms of the attribute values that are experienced and associated with the SQ. An issue that remains little explored to date is whether or not respondents "perceive"—and as a consequence evaluate—the alternatives associated with change from the SQ somewhat differently from the SQ alternative. This asymmetry would be consistent with reference-dependent

[1] Or "choice modelling", or "conjoint choice analysis" or more generally multi-attribute stated preferences.

utility theories (Kahneman and Tversky, 1979; Hartman *et al.* 1991; Samuelson and Zeckhauser, 1988; Bateman *et al.* 1997). This is a key issue because if such a difference exists, then it raises, amongst others, three problems relevant for choice experiments. First, how should one explicitly account for this effect in the analysis? Second, what are the consequences of accounting for it in an incorrect fashion, that is, the consequences of mis-specification? Third, what difference should one expect when using different sample sizes within the conventional range?

We find that these issues have yet to be satisfactorily addressed in the multi-attribute stated preference literature, and so, the main objective of this chapter is to explore such problems focusing on the finite sample properties of welfare estimates using experiments based on empirical results.

We compare three different ways of modelling diversity in perception of SQ versus alternatives involving change. All three are based on the conventional random utility framework.

In principle, one can argue that there are two kind of effects when a SQ alternative is present in all choice-sets. The first is a systematic effect which can be easily estimated by means of an alternative-specific SQ constant in the utility function. The second is an effect on the stochastic error structure postulated by the researcher. For example, designed alternatives involving change from the SQ one can share an error structure with a stochastic behaviour that is more similar to each other than it is to the error associated with the SQ alternative. In other words, designed alternatives involving change are correlated, and their error component is hence not independent. Such correlation can be accommodated within a nested logit framework.

In the literature SQ effects are typically dealt with by two specifications: the conditional logit with alternative-specific SQ constants and the nested logit. The first addresses systematic SQ effects, the second the correlation across utilities of designed alternatives. Both have been tested and found statistically significant. In this paper we propose a third specification, which flexibly and simultaneously addresses both types of effects by means of an error component mixed logit specification with alternative-specific SQ constant. This flexible model induces a correlation pattern in the utility of alternatives involving change, as well as capturing a systematic effect due to the SQ in the indirect utility.[2]. In one of the two datasets employed here is found to be a significant improvement over its competitors.

As a backdrop to such an investigation we report on findings from a large-scale survey designed to value the public good provision associated with water supply to the residents of the counties of Yorkshire, in the U.K. The analysis

[2]We are grateful to Joseph Herriges for suggesting this last specification during the 2004 EAERE meeting in Budapest, where a previous version of this paper was presented.

of these data provide evidence in support of the hypothesis of a systematic difference in customers' evaluations of SQ and non-SQ alternatives across the three specifications.

The finite sample properties of each of these three estimators are then analyzed using Monte Carlo experiments. These are conducted at three sample sizes that cover the range most commonly employed in the literature. The results of these experiments provide valuable information on the potential size of the bias in the estimates in the presence of SQ effects. They also provide suggestions on the relative efficiency of various estimators under misspecification.

The remainder of the paper is organized as follows. The next section briefly outlines the motivations for the investigation. Section 3 defines a common notation for the various models. In section 4 we describe the studies from which we draw inspiration for the Monte Carlo experiment. The results of the studies are presented in section 5, while the evidence from the Monte Carlo experiments is discussed in section 6. In section 7 we conclude.

2. Motivation for the study

2.1 The nature of SQ effects

For the purpose of this paper we define "SQ effects" as the systematic inclination of respondents to display a different attitude towards SQ alternatives from those reserved to alternatives involving some change, over and beyond what can be captured by the variation of attributes' levels across alternatives.

In the context of public economics, of which environmental economics is a sub-discipline, we are often concerned with a SQ resulting from previous policy outcomes, and incorporating public views on property rights, institutional arrangements etc. This extends to endowment of passive use and non-use values, which overall represent a bundle of issues conceptually quite different from those embedded in the SQ alternative employed in transportation and market research choice experiments, where the emphasis is on *use* values.[3]

However, regardless of its prevailing nature in environmental economics, this effect seems quite general. In their much quoted paper on consumer rationality and SQ effects (or "bias") Hartman *et al.* (1991) write:

> This analysis suggests, for example, that consumers attach "undue" importance to their current commodity bundle, demonstrating "apparently irrational" reluctance to alternative bundles. (Page 141)

The explanatory nature of such a general phenomenon is quite complex. For example, in a previous contribution on the topic Samuelson and Zeckhauser

[3]In transportation the tendency to systematically prefer the SQ alternative is termed "attrition", e.g. Bradley and Daly, 1997; Cantillo and Ortúzar, 2004.

(1988) identified and validated with evidence three major categories of explanations for this type of behaviour:

1 rational decision making in the presence of transition costs and/or uncertainty;

2 cognitive mis-perceptions (e.g. loss aversion and prospect theory as proposed by Kahneman and Tversky (1979))

3 psychological commitment stemming from misperceived sunk costs, regret avoidance, or a drive for consistency.

In their conclusions from the study Samuelson and Zeckhauser write:

> In choosing among alternatives individuals display a bias toward sticking with the status quo.[...] Assuming the status quo bias proves important, rational models will present excessively radical conclusions, exaggerating individual's responses to changing economic variables and predicting greater probability than observed in the world.

In our experience with SQ effects in choice experiments such effects can show both, a predilection for the SQ or a reluctance to stick with it.[4]

Because of this multiple causes of SQ bias, we do not find fruitful to elaborate on a conceptual model, which inevitably will leave some explanations unaccounted for. Hence, in what follows we maintain the conceptual definition quite general, yet we specifically focus on the analysis of multinomial discrete choices for environmental valuation under a random utility framework.

2.2 SQ effects in choice experiments

Above we have referred to some evidence from psychology and experimental economics suggestive that people evaluate what they know, and are familiar with (i.e. the SQ), in a systematically different fashion from how they evaluate hypothetical alternative scenarios. This has a direct bearing in the application of choice experiments to non-market valuation of public goods. Whether the technique is used to expand the set of modelling approaches existing in practice (as is often argued in support of mixing stated and revealed preference data (e.g. Hensher *et al.*, 1999), the so called "data fusion" approach), or is employed as a way to elicit trade-offs that will eventually lead to a richer description of people's preference for environmental goods (e.g. when respondents are asked to evaluate SQ scenarios against hypothetical changes), a check for what we loosely call "SQ effects" should be performed. The specification that best addresses such an effect will depend on circumstances.

[4]For further reference to status quo effects in power outages discrete choice elicitation studies see section 2 in this volume.

If a "generic" SQ effect exists in the evaluation of alternatives of choice-experiments for non-market goods, then an adequate practical understanding of the various econometric approaches accounting for such bias needs to be developed. By investigating the finite sample properties of common estimators we contribute to the on-going research on the understanding of the implications of modelling choices for the derivation of welfare estimates from choice-experiments investigating non-market goods.[5]

In particular, a systematic investigation of standard RUM-based modelling approaches to SQ effects in non-market valuation seems to be missing, and this is what we set out to provide here.

Some work on this issue is to be found in market research (Haaijer 1999; Haaijer *et al.* 2001), but this is limited to comparing Nested logit and conditional (or multinomial)[6] logit with alternative-specific constant for the SQ. Furthermore, the study is prevalently concerned with technical aspects proper of market research (coding effects, brand effects, market shares etc.). Their results suggest that the violation of the independence of irrelevant alternative makes the use of models not reliant on such restriction appealing, an argument that is often used in promoting the use of random parameter specifications for the indirect utility (Layton, 2000; Garrod *et al.*, 2002; Kontoleon and Yobe, 2003). This is a suggestion that we explore in more detail here by using more flexible models, but focussing on a basic error components specification, rather than on one with random parameters.

In practice we report the results of an investigation comparing multinomial logit with an alternative specific constant for the SQ (MNL-Asc), nested logit (NL), and an error component mixed logit (MXL-ε) which also includes an an alternative specific constant for the SQ. These encompass those models to which most practitioners would turn to, at least in the first instance, when trying to account for SQ effects in econometric specifications. In general our results show that the conventional practice of using simply a MNL-Asc may often be unsound. In particular, the proposed error component model with alternative-specific constant for the SQ, that would appear to be novel in this literature, seems to perform better than others in most of the circumstances examined here.

In the following we focus on the case in which the choice set contains only three alternatives: the SQ and two other alternatives, all of which are described

[5]For example, an original avenue of investigation based on hurdle models and more broadly centred on non-participation is to be found in von Haefen and Adamovicz (2004). Their results, like ours, show the sensitivity of welfare estimates to the treatment of SQ and non-participation choices.

[6]Somewhat loosely, the terms *conditional* and *multinomial* logit are used as synonymous in this chapter.

on the basis of attribute levels, and from which respondents are asked to select the one they prefer.[7]

Although there are many variations on the theme, the prevalent set-up for choice experiments in environmental economics tends to present respondents with a choice task that involves the identification of a preferred alternative from a choice set including the SQ and often two (e.g. Boxall *et al.*, 1996; Hanley *et al.*, 1998; Rolfe *et al.*, 2000; Foster and Mourato, 2003; Scarpa *et al.*, 2003; Lehtonen *et al.*, 2003) or sometimes few more (e.g. Kontoleon and Yabe, 2003) experimentally designed alternatives. This set-up is often argued on the basis of a lower cognitive burden on respondents than that associated with other choice contexts in which the complexity of choice task is higher. In experiments involving ranking – for example – especially with many experimentally designed alternatives, it has long been noticed that the hypothesis of identical preference across decisions at different ranks is empirically violated (Hausman and Ruud 1987; Ben-Akiva *et al.* 1992).

3. Econometric specifications accounting for SQ and their rationale

3.1 Common notation

It is useful to start by defining a common notation for the various models, referring as much as one can to convention. The reference structure is the case where the analytical objective is to obtain maximum likelihood estimates of a $1 \times k$ row vector of utility weights $\boldsymbol{\beta}$ for a column vector $\mathbf{x}$ of $k \times 1$ attributes for the individual linear indirect utility function V_j. The available data are choices from choice tasks including a SQ (indexed as sq) and a minimum of two experimentally designed alternatives (indexed with subscripts c_1, c_2). This basic implementation is often encountered in the published literature, but it can be extended without loss of generality. For the purpose of valuation, welfare estimates can be obtained as (non-linear) functions of the estimates, using the usual difference between the logsums weighted by the inverse of the cost coefficient.

3.2 Conditional logits with and without SQ-Asc

The basic random utility consistent model for analyzing choice experiment data is the conditional logit, which we consider here as a baseline.

When u_{nj}, the stochastic component of utility for respondent n and alternative j, is identically and independently Gumbel distributed across all alterna-

[7]The main results of our study can, however, be generalised to many other contexts in which respondents are asked to rank any set of alternatives including the one depicting the SQ.

tives, then the choice probability is expressed by the well-known formula:

$$P_n(i) = \frac{\exp(\lambda V_i)}{\sum_j \exp(\lambda V_j)},\ j = sq, c_1, c_2 \tag{3.1}$$

where λ is the scale parameter of the unobserved stochastic component. This is the conventional conditional logit model, which we refer to as MNL. Here any diversity in preference for the alternatives different from the SQ may be explicitly made part of the non-stochastic component of utility, for example in the form of an alternative specific constant (Asc) which takes the form of an indicator function $Asc = 1$ iff $j = sq$.

For a simple example of a positive value on the SQ Asc consider the tendency for respondents who perceive the cognitive task of assessing all the alternatives as too daunting, to fall back on the familiar SQ, rather than engaging into a costly and unrewarded cognitive task.

In practice, significance of Asc parameter represents the most immediate SQ test. Although often the ASCs are associated with the designed alternatives, and the SQ is left as a baseline (Hanley and Wright 2003), we prefer here to use a dummy variable for the alternative describing the SQ, rather than one for each of the alternatives involving change, as advocated in Adamovicz *et al.* (1998). This specification allows the analyst to account for diversity in the probability of choice between hypothetical alternatives and experienced SQ. Notice, however, that this solution does not change the stochastic structure of the model as it only enters the *deterministic* component of utility, leaving the *stochastic* error structure unchanged. As such it does not allow for a varying correlation structure across alternatives, which instead we find to be quite a plausible hypothesis in behavioural terms.

Under linear indirect utility $V_j = \boldsymbol{\beta}'\mathbf{x}_j$ our parsimonious specification is therefore:

$$P(c_1) = \frac{e^{\boldsymbol{\beta}\mathbf{x}_{c_1}}}{e^{(Asc+\boldsymbol{\beta}\mathbf{x}_{sq})} + e^{\boldsymbol{\beta}\mathbf{x}_{c_1}} + e^{\boldsymbol{\beta}\mathbf{x}_{c_2}}} \tag{3.2}$$

$$P(c_2) = \frac{e^{\boldsymbol{\beta}\mathbf{x}_{c_2}}}{e^{(Asc+\boldsymbol{\beta}\mathbf{x}_{sq})} + e^{\boldsymbol{\beta}\mathbf{x}_{c_1}} + e^{\boldsymbol{\beta}\mathbf{x}_{c_2}}} \tag{3.3}$$

$$P(sq) = \frac{e^{(Asc+\boldsymbol{\beta}\mathbf{x}_{sq})}}{e^{(Asc+\boldsymbol{\beta}\mathbf{x}_{sq})} + e^{\boldsymbol{\beta}\mathbf{x}_{c_1}} + e^{\boldsymbol{\beta}\mathbf{x}_{c_2}}} \tag{3.4}$$

where—for convenience—the subscript n denoting individuals is ignored, and the scale parameter λ is standardized to 1 and hence it is omitted.

Note that if hypothetical changes are expected to increase utility, then the sign of Asc will be negative, and positive if the effect has the opposite direction. This is the alternative specific constant conditional logit model, which we refer to as MNL-Asc.

3.3 Nested logit

The different nature of the SQ alternative vis-à-vis the two experimentally designed ones may translate into a difference in the substitution patterns, and hence in a different correlation structure of the unobserved components of the individual utilities. One assumption consistent with this case is when the stochastic component of utility is distributed according to a generalized extreme value (GEV) distribution, then different patterns of correlation across the utility of alternatives can be generated, although these are subject to considerable restrictions (Train, 2003). In fact, correlations are imposed to be similar within nests, but for alternatives in different nests the unobserved components are uncorrelated, and indeed independent. This is the case of the nested logit in which the unobserved components of utility have the GEV cumulative distribution:

$$\exp\left[-\sum_{g=1}^{G}\left(\sum_{j\in J}\exp(-u_{ig}/\eta_g)\right)^{\eta_g}\right] \tag{3.5}$$

where g denotes nests. In the set-up we consider here, with one SQ alternative and two experimentally designed ones, the assumption that the correlation amongst unobserved stochastic components differs between the two sets of alternatives generates two nests. The first is a degenerate one associated with the SQ alternative. The second is associated with changes from the SQ and contains both the experimentally designed ones. This gives rise to the following probability structure for the first decision stage:

$$P(change) = \frac{e^{\eta VI}}{e^{\eta VI} + e^{\boldsymbol{\beta}'\mathbf{x}_{sq}}} \text{ and } P(sq) = 1 - P(change)$$

While for the second decision stage, which is given the decision of embracing some change, is:

$$P(c_1|change) = \frac{e^{\boldsymbol{\beta}\mathbf{x}_{c_1}}}{e^{\boldsymbol{\beta}\mathbf{x}_{c_1}} + e^{\boldsymbol{\beta}\mathbf{x}_{c_2}}} \text{ and } P(c_2|change) = \frac{e^{\boldsymbol{\beta}\mathbf{x}_{c_2}}}{e^{\boldsymbol{\beta}\mathbf{x}_{c_1}} + e^{\boldsymbol{\beta}\mathbf{x}_{c_2}}} \tag{3.6}$$

$$P(c_j) = P(change)P(c_j|change) = \frac{e^{\eta VI}}{e^{\eta VI} + e^{\boldsymbol{\beta}\mathbf{x}_{sq}}} \times \frac{e^{\boldsymbol{\beta}\mathbf{x}_{c_j}}}{e^{\boldsymbol{\beta}\mathbf{x}_{c_j}} + e^{\boldsymbol{\beta}\mathbf{x}_{c\neq j}}} \tag{3.7}$$

where $VI = \ln\left[\exp\left(\boldsymbol{\beta}\mathbf{x}_{c_1}\right) + \exp\left(\boldsymbol{\beta}\mathbf{x}_{c_2}\right)\right]$ and can be interpreted as a measure of the expected utility of accessing the nest with the alternatives associated with change. The reader is reminded that the coefficient η is a measure of dissimilarity between alternatives in the various nests, while the value $1-\eta$ is a proxy for correlation for alternatives within the same nest. In this context a higher value of η can be intuitively interpreted as higher utility weight of moving away from the SQ.

A number of recent choice experiment studies in environmental economics have used nested logit models to account for SQ effects and found them superior in terms of fit to their MNL counterparts (Blamey *et al.* 2002; Hanley and Wright, 2003; Lehtonen *et al.*, 2003; Li *et al.*, 2004).

It is noteworthy that although this model maintains the independence of irrelevant alternatives (IIA) property across alternatives belonging to the same nest, it allows for differences in cross-elasticities across nests.

3.4 Error components via mixed logit

Notice that neither the MNL-Asc nor the NL specifications simultaneously identify both the *systematic* and *stochastic* components of the SQ effect, nor do they allow for taste-heterogeneity, or break completely[8] away from the IIA assumption. A specification that may overcome all these limitations is the mixed logit with error components. It does so by allowing flexible patterns of substitution via an induced correlation structure across utilities.

This is, of course, a special case of the large family of mixed logit, which—as described in McFadden and Train (2002)—with adequate data quality, may in principle be used to approximate any type of RUM.

The richness and flexibility of mixed logit models have been shown to generate a large variety of correlation patterns (Brownstone and Train, 1999; Train, 2003; Munizaga and Alvarez, 2001; Herriges and Phaneuf, 2002). Train (2003, page 156) discusses eloquently how mixed logit can give rise to two quite different interpretations, the random parameter and, under some restrictions, the error component one (or kernel logit (Ben-Akiva *et al.* 2001)). Further considerations, more specific to transportation applications, can be found in Cherchi and Ortúzar (2004), and some potential drawbacks are discussed in Hensher and Greene (2003).

Specifications using random utility parameters are well-known and often employed in choice experiments designed for the valuation of environmental goods in their panel form, so as to account for repeated choices, break away from the IIA assumption and address unobserved heterogeneity. However, in our study we wish to maintain comparability across the underlying assumptions of the MNL-Asc and NL specifications, which do not allow taste-heterogeneity. We hence focus on the decomposition of the *unobservable* component of utility, rather than on random effects in the indirect utility, and adopt only an error component interpretation, something that is less frequently seen in this kind of literature. We exploit the fact that the inclusion of additional zero-mean error components in the structure of utility of each nest induces correlation patterns (Herriges and Phaneuf, 2002). In the presence of SQ effects

[8]The nested logit maintains it within each nest.

different correlation patterns exists between the unobservable components of utility of the SQ alternative, and those in alternatives involving change.

For example, in our choice experiment the error component approach takes the following basic utility form[9]:

$$\begin{aligned} U(c_1) &= \boldsymbol{\beta}\mathbf{x}_{c_1} + \tilde{u}_{c_1} = \boldsymbol{\beta}\mathbf{x}_{c_1} + \varepsilon_{c_1} + u_{c_1}, \\ U(c_2) &= \boldsymbol{\beta}\mathbf{x}_{c_2} + \tilde{u}_{c_2} = \boldsymbol{\beta}\mathbf{x}_{c_2} + \varepsilon_{c_2} + u_{c_2}, \\ U(sq) &= Asc + \boldsymbol{\beta}\mathbf{x}_{sq} + u_{sq} \end{aligned} \tag{3.8}$$

where, in our case, $\varepsilon_{c_1} = \varepsilon_{c_2} \sim N(0, \sigma^2)$ are additional error components to u_{c_1} and u_{c_2}, which are Gumbel-distributed with variance $\pi^2/6$, thereby leading to the following error covariance structure :

$$Cov(\tilde{u}_{c_1}, \tilde{u}_{c_2}) = \sigma^2, \;\; Var(\tilde{u}_{c_1}, \tilde{u}_{c_2}) = \sigma^2 + \pi^2/6, \tag{3.9}$$

$$Cov(\tilde{u}_{c_j}, \tilde{u}_{sq}) = 0, \;\; Var(\tilde{u}_{c_j}, \tilde{u}_{sq}) = \pi^2/6, \;\; j = 1, 2; \tag{3.10}$$

where $\tilde{u}_{c_j} = \varepsilon_{c_j} + u_{c_j}$. Note that this is an analog of the nested logit model in the sense that it allows for correlation of utilities across alternatives in the same nest, but different correlation for those across nests. However, there is no IIA restriction, and the Asc captures any remaining systematic effect on the SQ alternative.

Conditional on the presence of the error component ε_j the choice probability is logit, and the assumption above leads to the following expression for each marginal choice probability:

$$\begin{aligned} P(i) &= \int_{\varepsilon} P(i|\varepsilon) f(\varepsilon|\boldsymbol{\theta}) d\varepsilon \\ P(i) &= \int_{-\infty}^{+\infty} \frac{e^{\boldsymbol{\beta}\mathbf{x}_i + \varepsilon_i}}{\sum_j e^{\boldsymbol{\beta}\mathbf{x}_j + \varepsilon_j}} \phi(0, \sigma^2) d\varepsilon, \;\; j = c_1, c_2, sq \end{aligned} \tag{3.11}$$

where $\phi(\cdot)$ is the normal density, and $\varepsilon_j = 0$ when $j = sq$.

Notice, however, that the additional error component can be either independent across choices (for example in a non-panel structure) or it can be the same for all choices made by the same individual (in a panel implementation). This is relevant in choice experiments as it breaks away from assumption of independence in the error structure across choices by the same respondent, which is implicit in both conditional and nested logit assumptions. While in random parameter specifications it is more plausible to assume fixity of parameter across choices by the same respondent by means of panel estimators, it is less clear

[9] In fact, as expanded upon by Brownstone and Train (1999) and Herriges and Panheuf (2002), more general forms than this may be empirically appealing.

that this is the case for error components. Ultimately this remains an empirical issue to be assessed case-by-case, and within the same category of model and estimation procedure it can be assessed on the basis of log-likelihood values. In this paper we focus on a non-panel application and hence dwell on the conventional assumption of independence of error across choices by the same respondent. This assumption ensures comparability of the error component results with those from the other two specifications which also implies independence across choices.

In what follows, we refer to this error component mixed logit model with Asc as MXL-ε.

It is important to note that such model nests the other three, in the sense that a restriction of $\sigma = 0$ is consistent with the MNL-Asc model, a restriction of $Asc = 0$ is consistent with an analog of the NL model. Both restrictions return the MNL model.

3.5 Estimation issues

All models are estimated in GAUSS 3.6 by maximum likelihood methods, except for equation (3.11), which is estimated by maximum simulated likelihood (MSL) with Halton draws[10] (Train, 2000, 2003). The choice probability for alternative i is approximated by:

$$P(i) \approx \tilde{P}(i) = \frac{1}{S}\sum_{s=1}^{S} \frac{e^{\boldsymbol{\beta}\mathbf{x}_i+\varepsilon_i^s}}{\sum_j e^{\boldsymbol{\beta}\mathbf{x}_j+\varepsilon_j^s}} \tag{3.12}$$

where $\varepsilon_j^s = 0$ when $j = sq$, and s denotes simulation draws.

4. The Yorkshire Water choice modelling studies

In spring and summer 2002, as a part of a large-scale investigation into the preference structure of its customers, Yorkshire Water (YW) conducted a set of choice experiments. The aim was to characterize the preference for fifteen different attributes related to water provision, called here service factors (SFs). As a result of focus-group activities and discussion with the management, these SFs were separated into five groups, giving rise to five separate choice experiments. The first three were mostly concerned with SFs of a private good nature, and are ignored here.[11] In this chapter we are concerned with

[10]Model estimates were found to be stable at 50 Halton draws and obtained by using the GAUSS code made available by Kenneth Train. However, error component models can be estimates also in Nlogit by formulating adequate dummy variables and using the subcommand "dummy(n,*,0)" which restricts the mean of normally distributed parameter to be equal to zero.

[11]For a more extensive report the interested reader is referred to Willis and Scarpa, 2002 or Willis *et al.* 2004.

the two choice experiments that addressed attributes of the service that can be commonly interpreted as 'public goods'.

4.1 Study 1

The first choice experiment, defined here as 'study 1', looked at four service factors as attributes: area flooding by sewage (AF); river quality (RQ); nuisance from odour and flies (OF); and cost of service (change in water bill payment). There were eight levels of payment expressed as either increases or decreases on the current bill, while all other attributes were expressed at four levels as reported in Table 13.1. The design chosen was an orthogonal main effect factorial with a total of 32 profiles, which were split into sequences of four choices for each respondent. The design was obtained using SAS (for a survey of experimental designs for logit models using SAS see Kuhfeld, 2004).

The expected signs for the coefficient estimates were as follows. The percent of areas protected from sewage escape is indicated as AF (area flooding) and it is expected to show a positive sign. The percent of river length capable of supporting healthy fisheries is indicated as RQ (river quality) and it is also expected to show a positive sign. Finally, the number of households and business affected by odour and flies (OF) is expected to show a negative sign. Notice that this is more a club good than a public good, but it certainly has public good characteristics.

4.2 Study 2

The second choice experiment looked at three service factors as attributes: water amenities for recreation (AM) expected to be positive, quality of bathing water (BB) also expected to be positive, and cost of service obviously expected to be negative. There were seven levels of payment always expressed as increases on the current bill, while all other attributes were expressed at three levels (Table 13.1). The orthogonal main effect factorial design was obtained with SAS and gave a total of 27 cards, which were also split in sequences of four choices for each respondent.

4.3 Sampling method

The survey instrument was tested in a pilot study and further refined as a consequence. It was administered in person, by enumerators experienced with stated-preference questionnaires through a computer-assisted survey instrument. A representative sample of 767 Yorkshire Water residential customers completed the sequence of choices in study 1, for a total of 2,921 choices. A representative sample of 777 residential customers completed the sequence for study 2 experiment with a total of 3,108 choices. More detailed information

Table 13.1. Service levels (SF) for residential and business customers.

Abbreviation	*Factor*	*Description*	*Scaling*	*levels* : –1	*levels* : 0	*levels* : +1	*levels* : +2
AF	Sewage escape to land	% of areas protected from sewage escape in gardens, roads, paths and open areas	1	Coded =20%	Coded =35%	Coded =50%	Coded =100%
RQ	Ecological quality of rivers	% of river length capable of supporting healthy fisheries and other aquatic life in the long term	1	Coded =60%	Coded =75%	Coded =85%	Coded =90%
OF	Odour and flies	Number of households and businesses affected by odour and high numbers of flies from sewage treatment works	0.01	2000 Coded =20	600 Coded =60	300 Coded =30	150 Coded =15
AM	Ability to use inland waters for recreational use	Number of areas with waste water discharges designed to allow recreational activities on rivers	1		0	4	12
BB	Bathing beaches water quality	Sewage works and disinfections designed to meet government standards for bathing water	1		Meets current existing gov'nt standards coded =100%	Improvement: 50% better than gov'nt standard coded =150%	Improvement: 100% better than gov'nt standard coded= 200%

on the sampling methodology and the samples employed is available from the report to the water company (Willis and Scarpa, 2002).

5. Results and discussion

5.1 Estimates for study 1

The estimates for study 1 are reported in Table 13.2. Notice that the utility weights all have the anticipated signs for the attributes of the alternatives, and are statistically significant in all models. The inclusive value estimate in the nested logit model is in the (0-1] interval, and hence is consistent with utility maximization. The estimated spread of the error component (σ) is virtually zero in the MXL-ε model, which basically is equivalent to the MNL-Asc and NL models.

Table 13.2. Estimates for study 1, SFs: AF, RQ, OF. N = 2,921.

Coefficient	MNL	MNL-Asc	NL	MXL-ε
AF	0.011 (10.1)	0.017 (13.5)	0.018 (13.3)	0.017 (13.5)
RQ	0.057 (17.5)	0.070 (19.0)	0.075 (19.3)	0.07 (19.0)
OF	–0.125 (–19.91)	–0.130 (–18.8)	–0.137 (–18.4)	–0.130 (–18.8)
Cost	–0.159 (–25.7)	–0.135 (–20.5)	–0.142 (–20.7)	–0.135 (–20.5)
σ_ε				0.040 (0.121)
η			0.899 (112.9)	
SQ-Asc		0.604 (10.7)		0.604 (10.7)
MRS_{AF}	0.07 (0.06,0.08)	0.13 (0.10,0.15)	0.13 (–0.16,0.36)	0.13 (0.10,0.15)
MRS_{RQ}	0.36 (0.32,0.39)	0.52 (0.46,0.59)	0.53 (0.06,0.08)	0.52 (0.46,0.59)
MRS_{OF}	–0.79 (–0.87,–0.71)	–0.96 (–1.10 ,–0.85)	–0.96 (–1.17,–0.09)	–0.96 (–1.10,–0.085)
ln-L or ln-Sim$L$	–2,245	–2,185	–2,185	–2,185
AIC	4,498	4,500	4,500	4,502

Confidence intervals around marginal rates of substitution obtained with Krinsky and Robb (1986) method.

We observe that all three models accounting for SQ achieve a very similar fit according to the Akaike Information Criteria (AIC=-2ln$L$+2p).[12] The lowest log-likelihood is fitted by the conventional conditional logit.

The evidence is consistent with the hypothesis that there is a systematic and significant difference in perception and substitutability between *experimentally designed* alternatives and *experienced* SQ.

From the viewpoint of policy evaluation it is clear that customers of YW, feel strongly for the public goods associated with various water provision strategies. For example, the implicit *WTP* for a one percent increase in the area protected from sewage escape is valued by the average household between 0.07 and 0.13 pence.

Relatively more valuable is the percent increase in the length of river capable of supporting long-term fisheries, which gives a value ranging from 0.36 to 0.53 pence. A reduction of one hundred properties suffering nuisance from odour and flies is valued between 0.79 and 0.96 pence if we consider the point estimates across specifications.

In this sample, it is evident that choosing estimates that account for SQ bias in some form, does make a substantial difference, as the MNL model provides lower estimates than the other three models. However, within those accounting for SQ effects, the welfare estimates are of similar magnitude, with the exception of the NL estimates for AF.

5.2 Estimates for study 2

A similar pattern of considerations can be made for study 2 in which the experimentally designed alternatives never allowed for a decrease in public good provision, something that – instead – was allowed for in study 1, and that undoubtedly may increase the likelihood of violation of the IIA property.

Again, the estimates are consistent with the hypothesis that respondents perceived experimentally designed alternatives and SQ *differently*. Neglecting this fact would lead the analyst to infer lower *WTP* estimates for the public goods examined and to select models that were significantly worse in terms of AIC value. In study 2 (Table 13.3), however, there seems to be support for the hypothesis that the difference in perception between SQ and change should be incorporated in the *stochastic* component of utility, rather than in the systematic one. In fact, allowing for different correlation patterns (NL) improves the AIC by a much larger amount than allowing only for a systematic SQ effect in the deterministic component of utility (MNL-Asc).

[12]This criterion can be used to discriminate between un-nested models by placing a penalty on the number of parameters p, since NL is nested neither in the MNL nor in the MXL-ε. The model associated with the minimum value is to be considered the best (Akaike, 1973).

Table 13.3. Estimates for study 2, SFs: AM, BB. N = 3,180.

Coefficient	MNL	MNL-Asc	NL	MXL-ε
AM	0.067 (13.0)	0.079 (13.4)	0.095 (13.8)	0.114 (14.2)
BB	0.132 (21.0)	0.148 (20.0)	0.170 (19.9)	0.210 (19.3)
Cost	–0.161 (–22.2)	–0.158 (–21.5)	–0.167 (–21.2)	–0.196 (–19.9)
σ_ε				3.702 (7.5)
η			0.833 (44.9)	
SQ-Asc		0.290 (4.3)		–1.024 (3.1)
MRS_{AM}	0.42 (0.36,0.47)	0.50 (0.43,0.57)	0.57 (0.49,0.64)	0.58 (0.51,0.65)
MRS_{BB}	0.82 (0.6,1.01)	0.94 (0.8,1.03)	1.02 (0.8,1.14)	1.07 (0.8,1.26)
ln-L or ln-Sim$L$	–2,776	–2,766	–2,748	–2,719
AIC	5,558	5,540	5,504	5,450

Confidence intervals around marginal rates of substitution obtained with Krinsky and Robb (1986) method.

When the IIA property is not imposed (MXL-ε) the model fits the data best, and its estimates identify substantial positive correlation (0.9) amongst non SQ alternatives: the estimated total variance for non SQ utilities is 15.335, much larger than the Gumbel error variance of $\pi^2/6 \simeq 1.645$. This large variance is only in part surprising, as the public goods components in the attributes under valuation are of much more pervasive interest to the population of customers in study 2 than in study 1. Public goods are known to be subject to much larger variation in individual valuations than private goods. A negative Asc reveals that respondents are in fact more inclined to support change from the SQ. This attitude would be consistent with a perception of under-provision of the public goods under valuation. A similar finding is reported in Lehtonen *et al.* 2003.

In terms of the policy implications for amenity provision and quality of bathing waters the estimates imply the following. An average *WTP* per household between 0.42 and 0.57 pence for an increase in one unit in the number of areas with waste water discharges designed to allow recreational activities on rivers; and 0.81 and 1.07 pence per household for a one percent increase in the current government standards for bathing waters.

WTP estimates do not vary much in magnitude across models, with the notable exception of the obviously mis-specified MNL model, which provides lower point estimates. Considering the confidence intervals — obtained using the Krinsky and Robb method (1986) — a significant difference is observed only for the value estimates of AM.

6. Monte Carlo experiments

6.1 Monte Carlo Design

The analyses of the two data-sets lead to results with contrasting interpretations. The first set of results indicates that the three SQ specifications are statistically equivalent. The second highlights that differences across SQ specifications can be statistically significant, although—at this sample size—they are not so for the implied *WTP* estimates. This issue raises the question of evaluating the relative finite-sample performance of the three SQ specifications.

To explore such an issue we focussed on the effects of reciprocal misspecifications in these models, and their sensitivity to sample size, by means of Monte Carlo (MC) experiments.

We ran a series of systematic experiments with GAUSS (routines are available from the authors) aimed at describing selected features of the finite sample properties of each of these specifications. The experiments were run using sample sizes (number of choices) that reflect those frequently encountered in the literature (N = 700, 1,400 and 2,900) so as to provide practitioners with some guidance about the expected efficiency gains achievable by increasing the sample size under different specifications and data generating processes.

Without loss of generality, we employ the data matrix of study 1 because it includes decreases of some valuable attribute levels. We use as data generating processes (DGPs) the set of estimates from this sample (Table 13.2), the only exception being the variance for the error component in MXL-ε. In this case the estimated σ_ε was not significant, so we choose to use a larger, yet realistic value (the estimated value from study 2, in Table 13.3). The steps involved are:
1) Compute the deterministic part of the utility for each alternative by using the maximum-likelihood (ML) or simulated-ML estimates reported in Table 13.2 and the original matrix of attribute levels $\mathbf{X}$.
2) Generate the unobserved stochastic component of the utility of each alternative by using pseudo-random draws (with seed) from the inverse cumulative distribution function suitable for each model.[13]
3) Derive an indicator of choice $\mathbf{y}^r$ from the alternative associated with the highest computed utility.

[13]Gumbel errors for models MNL, MNL-Asc and for the u_j of equation (3.8) the MXL-ε; GEV for the NL as for equation (3.5); and a re-scaled standard Normal for the ε_{c_1} and ε_{c_2} of equation (3.8) for MXL-ε.

4) Proceed to the estimation of the parameters of all models based on the simulated choice responses $\mathbf{y}^r$ and matrix of attribute levels $\mathbf{X}$, and save the relevant results of estimation (parameter estimates, t-values, log-likelihood at convergence etc.).
5) Repeat the previous steps for R=550 times.

Given the results in chapter 16 in this volume by G. Baiocchi, we report results of the MC experiment by using pseudo-random draws obtained with R and loaded into GAUSS. Overall the results were qualitatively similar to those previously obtained in GAUSS. The results presented here are from draws obtained from the free software R (those obtained in GAUSS are available from the authors upon request).

As a criterion to evaluate the performance of the various estimators we focus our attention on the expected difference between squared errors:

$$\Delta_{SE} = E\left((\hat{\gamma} - \gamma_0)^2 - (\tilde{\gamma} - \gamma_0)^2\right) \tag{6.1}$$

where $\hat{\gamma}$ is the estimator under mis-specification, $\tilde{\gamma}$ is the estimator correctly specified, and γ_0 is the true value from the DGP. The larger this value the worse is the consequence of mis-specification.

We prefer Δ_{SE} to the more frequently employed difference between mean squared errors:

$$E(\hat{\gamma} - \gamma_0)^2 - E(\tilde{\gamma} - \gamma_0)^2 \tag{6.2}$$

because of the lower variance associated with its estimator, which we estimate by:[14]

$$\bar{\Delta}_{SE} = \frac{1}{R}\sum_{r=1}^{R}\left((\hat{\gamma}_r - \gamma_0)^2 - (\tilde{\gamma}_r - \gamma_0)^2\right) \tag{6.3}$$

Since only relative values matter in the coefficient estimates in random utility models, we focus on the marginal rates of substitution (γ = MRS), which are computed relative to the money coefficient. These measures—under certain conditions—can be interpreted as marginal *WTP* values, and hence are meaningful *per se*. Further, parameter estimates are asymptotically normally distributed, but MRS are non-linear functions and as such they do not have a well-defined sampling distribution.

[14] As Davidson and MacKinnon (1993) point out (page 740), the variance of equation (6.2) is: $R^{-1}V(\hat{\gamma}) + R^{-1}V(\tilde{\gamma}) - 2R^{-1}Cov(\hat{\gamma}, \tilde{\gamma})$ and for a positive covariance this variance is inferior to the variance associated with the difference of the mean squared errors. A positive covariance across estimates is very likely in our implementation because the estimates are obtained using the same pseudo-random draws.

Four types of concise measures are reported from the Monte Carlo experiments.[15]

First, we report the mean of the differences of the squared errors as from equation (6.2). Mis-specified models associated with large values of these are troublesome. Negative values indicate that the mis-specification is on average less biased than the correct estimator at that sample size, which can be explained by a compensating higher efficiency. To give a more readily interpretable measure of efficiency we also report the values of the inter-quartiles of these differences. The smaller these intervals the more efficient the mis-specification.

Secondly, we report the percent of cases in which the mis-specified estimator produces an estimate which is closer to the true value than the correct estimator. We report this in two forms, one for each MRS, $I(AF), I(RQ)$, and $I(OF)$ and one reporting the percent of cases in which this happens for *all three* attributes $I(AF, RQ, OF)$.

Thirdly, we report the mean of the relative absolute error:

$$\overline{RAE} = \frac{1}{R} \sum_{r}^{R} \left| \frac{\hat{\gamma}_r - \gamma_0}{\gamma_0} \right|. \tag{6.4}$$

This measure gives an idea of the relative magnitude of the bias of the estimate.

Finally, we report the fraction of MC experiments in which the estimated MRS is placed within a 5% interval around the true value, as a measure of efficiency computed as:

$$\Gamma_{0.05} = \frac{1}{R} \sum_{r}^{R} 1(\hat{\gamma}_r \in \gamma_0 \pm \gamma_0 \times 0.05). \tag{6.5}$$

Where $1(\cdot)$ is an indicator function. This count gives an idea of how clustered estimates are around the true values.

In addition, select points are illustrated using plots of the kernel smoothing of the obtained distributions of estimates, using the normal kernel with optimal bandwidth [4].

6.2 Monte Carlo Results

The results reported in tables 13.4-13.6 indicate that the values for $\bar{\Delta}_{SE}$[16] and their dispersion—as described by the size of the inter-quartile intervals in

[15] We omit to report the simulation performance of the AIC as a selection criterion for the correct specification. In brief, the simulation results showed that AIC was a stable indicator of performance and performed extremely well at all sample sizes and across all models.

[16] These values are scaled up by 1,000.

brackets —decrease as the sample size increases. Notice that in some cases—as evidenced by non-positive values of $\bar{\Delta}_{SE}$—the mis-specified model outperforms the true model in terms of the size of the expected squared-bias. This happens at all sample sizes and for all attributes when the true DGP is MNL-Asc and the specification is MXL-ε (Table 13.4). Under this DGP the specification MXL-ε seems to perform at least as well as the NL one, except at small sample sizes, and limited to $\Gamma_{0.05}$ and to individual $I(\cdot)$ values.

In terms of expected squared bias, when the DGP is NL the MXL-ε (Table 13.5) performs either as well (AF), or better than the correct specification at small sample sizes, but not at medium to high. Interestingly, at this sample size the MNL-Asc specification outperforms the true one for one attribute (OF). However, for this attribute the mis-specification MXL-ε gives more accurate estimates than the true specification 16% of the times, versus a 4 and 11% for the MNL and MNL-Asc, respectively. In terms of cases within the 5% interval around the true values, MXL-ε performs very similarly to the true specification at all sample sizes.

Notice, though, that the results in Table 13.6 show that when the true DGP is MXL-ε the mis-specifications *never* outperform the true specification in all the criteria, across all sample sizes. When, instead, MXL-ε was not the true DGP the mis-specifications never substantially outperform it. This is suggestive that, in the absence of a strong a-priori information on the true specification, the MXL-ε is preferable across the board.

In figure 13.1 we present a kernel plot of the distributions of the RAE for the *WTP* for Area Flooding when the true model is a nested logit, with N=2,900. From this figure it is evident how the real choice is between MNL, and the group MNL-Asc, NL and MXL-ε. Similar patterns emerge when the DGP is MNL-Asc, suggesting that these three models are effectively interchangeable. A stronger difference across specifications accounting for SQ emerges when the DGP is MXL-ε, as shown in figure 13.2. Here the true specification (dot-dashed line) shows a distribution of RAE values that outperforms the other two (dotted and continuous line) in that it is much more tightly concentrated on zero, while the MNL (dashed line) remains strongly biased.

7. Conclusions

Our empirical results from the analysis of the preferences of customers of Yorkshire Water are consistent with the fact that they are willing to pay for environmental improvements via an increase on their water bill. The estimated amount of *WTP* for quasi-public goods and pure public goods is plausible, and it is quite stable across the specifications used.

The models providing best statistical fit are found amongst those accounting for SQ effects, that was loosely defined as a systematic effect to choose the

Table 13.4. Summary statistics of Monte Carlo distributions of *WTP* estimates for DGP MNL-Asc.

	MNL			MNL-Asc			NL			MXL-ε		
Sample size	700	1,400	2,900	700	1,400	2,900	700	1,400	2,900	700	1,400	2,900
$\bar{\Delta}_{SE}^{AF}$	3.33	2.49	2.8				0.12	0.15	0.16	–0.01	0	0
	(6.39)	(3.86)	(3.71)				(0.07)	(0.05)	(0.03)	(0.02)	(0.01)	(0)
$\bar{\Delta}_{SE}^{AF}$	5.39	26.09	26.52				0.96	0.87	1.02	–0.01	–0.04	–0.01
	(19.07)	(40.07)	(34.76)				(0.97)	(0.7)	(0.46)	(0.38)	(0.21)	(0.1)
$\bar{\Delta}_{SE}^{AF}$	–11.97	21.83	29.98				–1.56	–0.8	–0.4	–0.29	–0.02	–0.02
	(17.09)	(49.53)	(47.01)				(2.94)	(1.88)	(1.63)	(1.6)	(0.48)	(0.26)
$I(AF)$	9	4	0				56	53	48	55	53	51
$I(RQ)$	24	4	0				50	51	51	47	50	53
$I(OF)$	41	16	9				57	51	47	51	54	49
$I(AF, RQ, OF)$	4	1	0				18	22	21	20	22	21
$\overline{RAE}(AF)$	0.49	0.41	0.43	0.17	0. 10	0.07	0.17	0.1	0.08	0.17	0.1	0.06
$\overline{RAE}(RQ)$	0.19	0.32	0.32	0.13	0.08	0.05	0.13	0.08	0.06	0.13	0.08	0.05
$\overline{RAE}(OF)$	0.11	0.18	0.18	0.14	0.08	0.06	0.13	0.08	0.06	0.14	0.08	0.06
$\Gamma_{0.05}^{AF}$	0	0	0	17	34	46	19	32	45	18	34	45
$\Gamma_{0.05}^{RQ}$	9	0	0	25	38	55	25	39	54	24	37	55
$\Gamma_{0.05}^{OF}$	24	3	0	22	39	49	24	43	49	23	40	50

True absolute values of *WTP* are AF = 0.13, RQ = 0.52, OF = 0.97. The $\bar{\Delta}_{SE}$ values are multiplied by 1,000.

Table 13.5. Summary statistics of Monte Carlo distributions of *WTP* estimates for DGP NL.

	MNL			MNL-Asc			NL			MXL-ε		
Sample size	700	1,400	2,900	700	1,400	2,900	700	1,400	2,900	700	1,400	2,900
$\bar{\Delta}_{SE}^{AF}$	3.38	2.57	2.85	0.02	0.01	0				0	0	0
	(6.23)	(4.02)	(3.83)	(0.12)	(0.04)	(0.03)				(0.07)	(0.45)	(0.05)
$\bar{\Delta}_{SE}^{RQ}$	5.14	27.19	27.1	0.07	0.05	0.02				–0.01	0.01	0.03
	(19.6)	(41.03)	(36.06)	(0.79)	(0.69)	(0.33)				(0.72)	(0.56)	(0.31)
$\bar{\Delta}_{SE}^{OF}$	–10.82	22.55	28.53	0.87	0.77	0.32				–0.44	0.07	0.07
	(16.25)	(50.21)	(48.16)	(5.58)	(2.98)	(1.68)				(3.3)	(1.9)	(1.11)
$I(AF)$	1	2	0	43	43	43				50	47	45
$I(RQ)$	24	1	0	47	51	47				48	47	46
$I(OF)$	42	15	5	43	41	45				50	47	49
$I(AF, RQ, OF)$	4	0	0	11	11	12				16	18	18
$\overline{RAE}(AF)$	0.48	0.41	0.43	0.16	0.09	0.07	0.16	0.09	0.07	0.16	0.1	0.07
$\overline{RAE}(RQ)$	0.18	0.32	0.31	0.13	0.07	0.05	0.13	0.07	0.05	0.13	0.07	0.05
$\overline{RAE}(OF)$	0.12	0.18	0.18	0.14	0.09	0.06	0.14	0.08	0.06	0.14	0.08	0.06
$\Gamma_{0.05}^{AF}$	0	0	0	19	37	41	20	37	41	20	36	41
$\Gamma_{0.05}^{RQ}$	10	0	0	25	45	56	26	44	56	25	43	56
$\Gamma_{0.05}^{OF}$	23	4	0	25	35	50	25	36	51	25	35	51

True absolute values of *WTP* are AF = 0.13, RQ = 0.52, OF = 0.97. The $\bar{\Delta}_{SE}$ values are multiplied by 1,000.

Table 13.6. Summary statistics of Monte Carlo distributions of *WTP* estimates for DGP MXL-ε.

	MNL			MNL-Asc			NL			MXL-ε		
Sample size	700	1,400	2,900	700	1,400	2,900	700	1,400	2,900	700	1,400	2,900
$\bar{\Delta}_{SE}^{AF}$	8.12	6.21	6.55	0.31	0.14	0.07	0.28	0.13	0.06			
	(13.98)	(8.87)	(8.28)	(1.22)	(0.55)	(0.31)	(1.14)	(0.53)	(0.27)			
$\bar{\Delta}_{SE}^{AF}$	13.56	38.95	40.44	2.89	3.15	1.21	6.31	2.45	0.88			
	(38.55)	(60.17)	(54.96)	(14.18)	(13.86)	(5.45)	(17.9)	(9.72)	(3.96)			
$\bar{\Delta}_{SE}^{AF}$	−13.8	28.16	30.61	13.68	8.57	4.96	17.86	13.59	10.32			
	(19.95)	(65.16)	(55.3)	(53.66)	(33.41)	(19.93)	(79.52)	(48.89)	(33.35)			
$I(AF)$	2	0	0	37	31	32	35	35	36			
$I(RQ)$	22	2	0	39	38	41	31	38	39			
$I(OF)$	58	15	8	42	40	38	42	37	31			
$I(AF, RQ, OF)$	2	0	0	8	6	6	5	9	8			
$\overline{RAE}(AF)$	0.73	0.64	0.65	0.21	0.13	0.1	0.21	0.13	0.09	0.17	0.17	0.08
$\overline{RAE}(RQ)$	0.26	0.39	0.39	0.17	0.13	0.09	0.19	0.13	0.08	0.14	0.1	0.07
$\overline{RAE}(OF)$	0.14	0.2	0.19	0.19	0.12	0.09	0.19	0.13	0.11	0.16	0.09	0.07
$\Gamma_{0.05}^{AF}$	0	0	0	16	25	32	14	27	34	19	30	38
$\Gamma_{0.05}^{RQ}$	6	0	0	18	24	39	18	26	43	23	31	45
$\Gamma_{0.05}^{OF}$	19	4	0	16	25	38	18	25	30	16	33	46

True absolute values of *WTP* are AF = 0.13, RQ = 0.52, OF = 0.97. The $\bar{\Delta}_{SE}$ values are multiplied by 1,000.

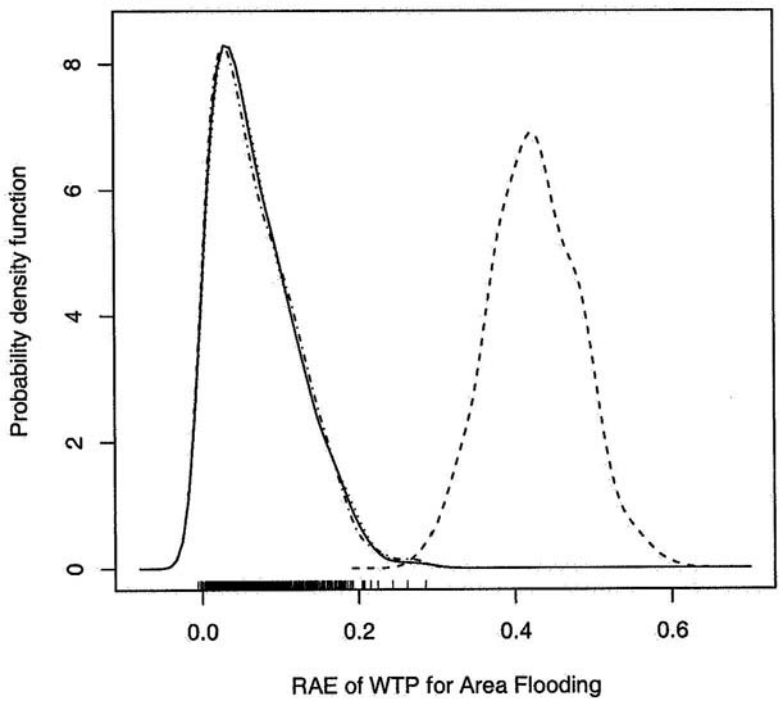


Figure 13.1. Plot of kernel-smoothed distribution of the Relative Absolute Error of *WTP* estimates for Area Flooding. True data generating process (Continuous line NL): MXL-ε dashed-dotted line; Dashed line MNL; Dotted line MNL-Asc.

SQ or the alternatives different from the SQ beyond what can be explained on the basis of the attributes values alone. We found that in our samples the conditional logit model, that ignores any source of SQ effect, produces the lowest estimates of benefits from provision of externalities. While from the societal viewpoint such a conservative estimate would guide investments in a cautious way, it would still represent a sub-optimal resource allocation, as many potentially beneficial proposals would fail the Pareto efficiency test by providing too low a benefit estimate.

Following other authors (Haaijer, 1999; Kontoleon and Yabe 2003), we have argued that there are very good reasons for investigating the existence of SQ effects in the application of choice-experiments, and that these reasons might be particularly compelling in non-market valuation of environmental goods.

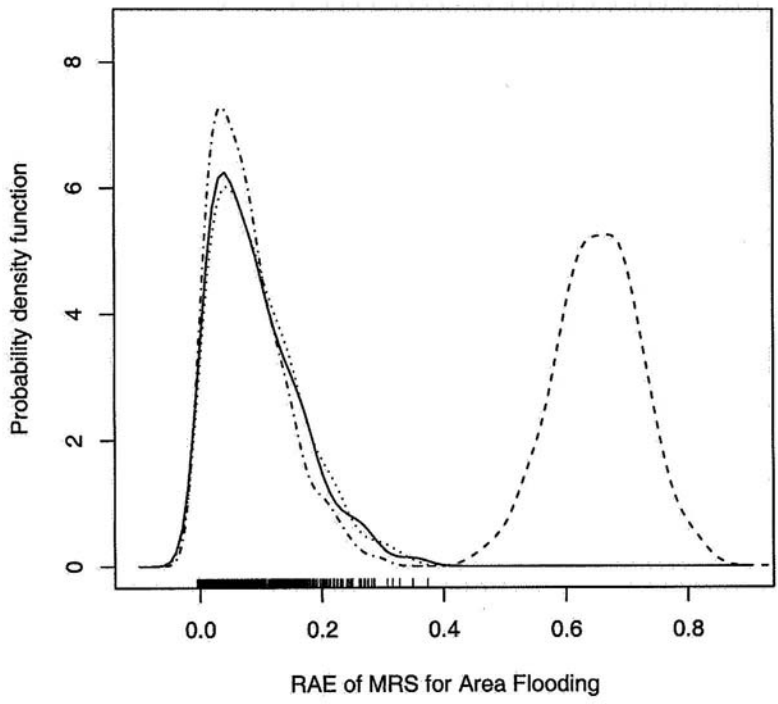


Figure 13.2. Plot of kernel-smoothed distribution of the Relative Absolute Error of *WTP* estimates for Area Flooding. True data generating process: MXL-ε (dashed-dotted line). Continuous line NL; Dashed line MNL; Dotted line MNL-Asc.

We examined three specifications that can be used to account for these effects: the conventional logit model with alternative-specific constant, the nested logit model and and the less conventional mixed logit with error components and alternative specific constant.

Secondly, we reported how we observed different forms of statistical evidence of SQ effects in two separate studies on preferences for water management attributes, which include important public goods, such as number of areas protected by flooding and number of households protected from odour and flies. While in a study we observe that all three specifications accounting for SQ afford similar statistical performance and *WTP* estimates, in the other application we observe that the mixed logit with error component and

alternative-specific constant statistically dominates the nested logit and MNL-Asc, but this dominance does not implies statistically different estimates.

Finally, we investigated the effects of mis-specification using in turn the three SQ data generating processes by means of Monte Carlo experiments over a plausible range of sample sizes. The results of the experiments suggest a number of points.

First, when SQ effects are a concern, the use of simple conditional logit specifications may produce strongly biased estimates for the taste parameters. These will also produce biased welfare measures.

Secondly, when the true DGP is mis-specified, the MXL-ε specification generally provides a good performance in our Monte Carlo experiments. Such performance is not matched neither by the NL model nor by the MNL-Asc model when the true DGP is MXL-ε.

In conclusion, our empirical results confirm the existence of a systematic effect of the status-quo alternative on choice selection. This was previously discussed and evidenced in general terms by Samuelson and Zeckhauser (1988) and Hartman *et al.* (1991). Such effect was examined more specifically in the context of choice-experiment in market research by Haaijer (1999) and Haaijer *et al.* (2001) and addressed in environmental economics by Hanley and Wright (2003), and Li *et al.* (2004) by means of nested logit models.

We find that a less usual specification, namely the MXL-ε consistently achieves better results than MNL with an alternative-specific constant for the SQ and NL specifications. The MXL-ε model is parsimonious, yet, it captures SQ effects in both the systematic component of preference via alternative-specific constant, and the unobserved heterogeneity associated with hypothetical changes described by unfamiliar attribute levels. It also breaks away from the restrictive independence of irrelevant alternatives.

Of course the usual caveats pertaining to Monte Carlo results apply here. Namely, these results might be not very general and perhaps they are due to the particular data employed in this study. Nevertheless we find quite plausible that a specification that accommodates status-quo effects simultaneously in both the stochastic and deterministic component of utility outperforms specifications that only address one at the time.

Further research should investigate how general these preliminary results are, and how status-quo effects can be related to the various features of the experimental design, investigating — for example — the relationship between choice-complexity and degree of familiarity with attributes levels defining the status-quo vis-à-vis the proposed changes.

Chapter 14

DPSIM MODELLING: DYNAMIC OPTIMIZATION IN LARGE SCALE SIMULATION MODELS

Richard T. Woodward
Department of Agricultural Economics
Texas A&M University
2124 TAMU College Station
TX 77843-2124, U.S.A.
r-woodward@tamu.edu

Wade L. Griffin
Department of Agricultural Economics
Texas A&M University
2124 TAMU College Station
TX 77843-2124, U.S.A.
wgriffin@tamu.edu

Yong-Suhk Wui
Department of Business Administration and Economics,
University of Arkansas at Pine Bluff
AR 71601, U.S.A.
yswui@yahoo.com

Abstract Although it is well established that dynamically optimal policies should be "closed loop" so that policies take into account changing conditions of a system, it is rare for such optimization to actually be carried out in large-scale simulation models. Computational limitations remain a major barrier to the study of dynamically optimal policies. Since the size of dynamic optimization problems grows approximately geometrically with the state space, this problem will continue to inhibit the identification of dynamically optimal policies for the foreseeable future. In this chapter, we explore in detail the problem of solving dynamic

optimization problems for large-scale simulation models and consider methods to work around the computational barriers. We show that a reasonable approach is to solve a small-scale problem to identify an approximate value function that can then be embedded directly in the simulation model to find approximately optimal time-paths. We present and compare two ways to specify the small-scale problem: a traditional "meta-modelling" approach, and a new "direct approach" in which the simulation model is embedded directly in the dynamic optimization algorithm. The methods are employed in a model of the Gulf of Mexico's red snapper fishery and used to identify the dynamically optimal total allowable catch for the recreational and commercial sectors of the fishery.

Keywords: Dynamic programming, Bellman's equation, curse of dimensionality, Moores law, General Bio-economic Fishery Simulation Model (GBFSM), value function, Gulf of Mexico Fishery Management Council, Red snapper.

1. Introduction

As the chapters in this book demonstrate, simulation modelling is a critical tool for applied economic analysis. This is particularly true when interdisciplinary analysis is to be carried out. Natural systems are complicated, and although economists may be comfortable with general specifications of the physical characteristics of such systems, the best available models are often quite detailed, representing the system using many variables and empirically estimated relationships. This presents an important challenge for the analyst. If a large simulation model is to be used, in which the state of the system is represented by dozens if not hundreds of variables, then it is computationally impossible for the analyst to identify dynamically optimal policy paths. In this chapter we explore means by which (approximate) dynamically optimal paths can be found, even for very large simulation models.

This paper is tied closely to Woodward, Wui and Griffin (2005, hereafter WWG) in which we spell out the problem of carrying out dynamically optimal policy analysis when working with a large scale simulation model. In that paper we propose a new way of solving such problems, which we call the *direct approach*. In the direct approach the complete simulation model is embedded in the dynamic programming algorithm. This is distinguished from the *meta-modelling approach*, in which a system of equations is estimated that, in effect, simulate the simulation model itself. We were unable to find any analysts that had applied the direct approach previously, while the meta-modelling approach is quite frequently used.

In this chapter we implement and compare the direct and meta-modelling approaches. Although we cannot draw general conclusions as to which method might be preferred, in our application the direct approach seems to be preferred, yielding a better approximation of the simulated system and, therefore, is probably preferred for dynamic policy analysis.

The chapter is organized as follows. In the next section we spell out the difficulty and importance of combining simulation analysis with dynamic optimization—what we call DPSim modelling. Section 2 summarizes two approaches that can be used to carry out DPSim analysis. The following section discusses our empirical application and discusses how the two approaches were implemented in this case. The results of our modelling are presented and the direct and meta-modelling approaches are compared. The conclusion summarizes our findings and discusses issues that require further research.

2. The need for and difficulty of dynamic optimization in large scale simulation models

Simulation modelling plays an important role in policy analysis. However, except for extremely small and simple models, it is computationally impossible to analyze all possible policy paths that might be pursued. Hence, truly optimal policy paths cannot be identified. This is an important limitation for simulation modelers since many policies have important dynamic consequences, affecting both future outcomes and future options. To the extent that simulation modelling can assist in identifying preferred policies, it is certainly desirable that dynamically optimal policies be presented.

The standard approach to solving empirical dynamic optimization problems is numerical dynamic programming (DP) (Bertsekas 1976, Kennedy 1986). In such problems, the state of the system is defined by a vector of *state variables*, x. Under fairly generalizable conditions, almost any dynamic optimization problem can be solved to yield an optimal policy rule, $z^*(x)$, and an optimal value function, $v(x)$. Assuming an infinite horizon,[1] the value function is defined implicitly by the Bellman's equation,

$$v(x_t) = \max_z E(u(x_t, z_t) + \beta v(x_{t+1})) \tag{2.1}$$

$$s.t. \quad x_{t+1} \sim g(x_t, z_t)$$

where $u(\cdot)$ defines the benefits in the current period as a function of x and choices, z; $g(\cdot)$ defines the conditional probability distribution over future states, x_{t+1}; E is the expectation operator; and β is the discount factor.[2]

If the function $v(x)$ were known then starting at a known initial state, x_0, an optimal policy path could be simulated by successively solving the problem

$$z^*(x_t) = \arg\max_z E(u(x_t, z_t) + \beta v(x_{t+1})). \tag{2.2}$$

[1] All of our analysis below is also applicable to finite horizon problems. We present the infinite horizon case because it is notationally cleaner.

[2] In our empirical analysis we use an annual discount rate of seven percent as specified in Executive Office of the President (1992).

In the traditional simulation analysis, a policy rule (frequently expressed as a constant policy) is simulated for a number of periods. Simulating an optimal path has advantages both normatively and positively. From a normative perspective, policies that follow a dynamically optimal policy path will tend to yield greater net benefits. From a positive perspective, simulating a dynamically optimal path frequently makes more sense because decision makers do adjust over time to changing conditions. Hence, equation (2.2) provides a preferred foundation for policy simulation.

As noted by WWG, if $v(x)$ were known, then the optimal paths could be simulated by repeatedly solving equation (2.2) with the complete simulation model used to simulate the optimal policy path. For simulation analysis, therefore, the goal of a DP exercise should not be to find the optimal policy function, $z^*(x)$, but to find an approximate value function, $v(\cdot)$. Once this function is found, an approximately optimal policy path can be found by including the full simulation model in the policy optimization problem, equation (2.2). In this manner, dynamically optimal analysis can be carried out without abandoning the large simulation model. We call this approach, which combines simulation analysis with dynamic optimization, DPSim modelling.

The challenge of DPSim modelling is, of course, that the value function is not known *a priori*. In principle, $v(\cdot)$ can be found using the methods of numerical DP. In practice, however, it can only be identified for relatively small systems. Regardless of the solution method employed, solving DP problems requires that a separate optimization problem be solved at a large number of points in the state space. Hence, the computational size of a dynamic optimization problem tends to grow exponentially with the number of state variables; a fact that Bellman referred to as the "curse of dimensionality".

Suppose that there are n_x variables in the state space and that each of these is represented by m different values. In this case to solve the DP problem m^{n_x} different optimization problems would need to be solved repeatedly. If n_x increases by one, the size of the problem grows m-fold.

Because of the curse of dimensionality, dynamic optimization problems can be solved only for relatively small systems: $n_x < 10$, and usually less than 3. Certainly DP problems cannot be solved for the large systems in which 100 or more variables are used to replicate complicated dynamic systems. Although enormous improvements in the computational speed have been achieved in recent years, this computational burden will continue to limit the size of DP problems for many years to come. "Moore's law" describes the regular tendency for the density of computer circuitry (and processor speed) to double every eighteen months (Schaller 1997). This "law", which has held up surprisingly well since its conception in 1965, has startling implications for simulation modelers: a simulation model could double in size every 1.5 years without slowing down. The implications for DP, however, are not nearly so promising.

For example, in a model in which each state variable takes on just 8 possible values, it would be 4.5 years before just one state variable could be added without increasing the run time of the program. The solution of DP problems with hundreds of state variables lies only in the far distant future.[3]

3. Two approaches to DPSim modelling

We describe here two ways in which to work within the computational limits to solve a dynamic optimization problem linked to a large simulation model. To formalize our discussion we need to introduce some notation. Let X_t be the (large) vector of state variables used in the simulation model and Z_t be a vector of policy options. A policy path is denoted $Z = \{Z_0, Z_1, \ldots\}$, where Z_t is the vector of policies implemented in period *t*. The simulation model can be thought of as a pair of mappings S^1 and S^2 from choices and conditions in period *t*, Z_t and X_t, into future conditions, X_{t+1}, and net benefits in *t*, u_t, respectively.

To illuminate the general discussion in this section, we will frequently refer to our case study that we examine in detail below. Our case study is a fisheries management problem in which we use the General Bioeconomic Fishery Simulation Model (GBFSM) as our model foundation. GBFSM is a highly detailed model that has been parameterized for the Gulf of Mexico's red snapper fishery. In the model, red snapper stocks fall into 360 different groups distributed across two depth zones (depths 2 and 3 in the model) for a total of 720 depth-cohort groups. Hence, the complete vector of state variables, X_t would include 720 elements. One of the policies being considered in the case study is the total allowable catch (TAC) of red snapper, so that Z_t is the TAC chosen. A run of GBFSM for a single year takes a starting stock, X_t, and a specified TAC level, Z_t, and yields a prediction of the stock in the following year, X_{t+1}, and the producer and consumer surplus generated by the fishery, u_t.

In general, if in each period there are n_z possible policies, then the total number of possible paths over *T* periods is $N_z = n_z T$. Since N_z grows exponentially in *T*, for all but extremely simple or short-run cases it is computationally infeasible to evaluate every possible policy path. Hence, in most simulation exercises only a small subset of all possible policy paths are evaluated. Hence, instead of simply optimizing over this extremely large set of policy paths, it is usually better to solve for $v(\cdot)$ in equation (2.1).

As we noted above, since DP problems with hundreds of state variables cannot be solved, the alternative way to identify approximate solutions to the DP

[3]Of course brute computational force is not the only option available to the analyst. Improvements in the efficiency of DP algorithm (e.g., Judd 1998 and Mrkaic 2002) and opportunities exist for very efficient programming including the use of parallel processors (Rust, 1997b) or Rust's randomization approach (1997a), which can avoid the geometric growth in some problems.

problem is to specify an approximating DP problem in terms of a small set of variables, x, which approximate the complete set in X. The problem of identifying an approximating value function for a simulation model is presented graphically in Figure 14.1 (taken from Woodward *et al.* 2005). For simplicity, we present the model as if it is deterministic and references to Z and S^2 are suppressed to focus on the dynamic processes of the system, S^1.

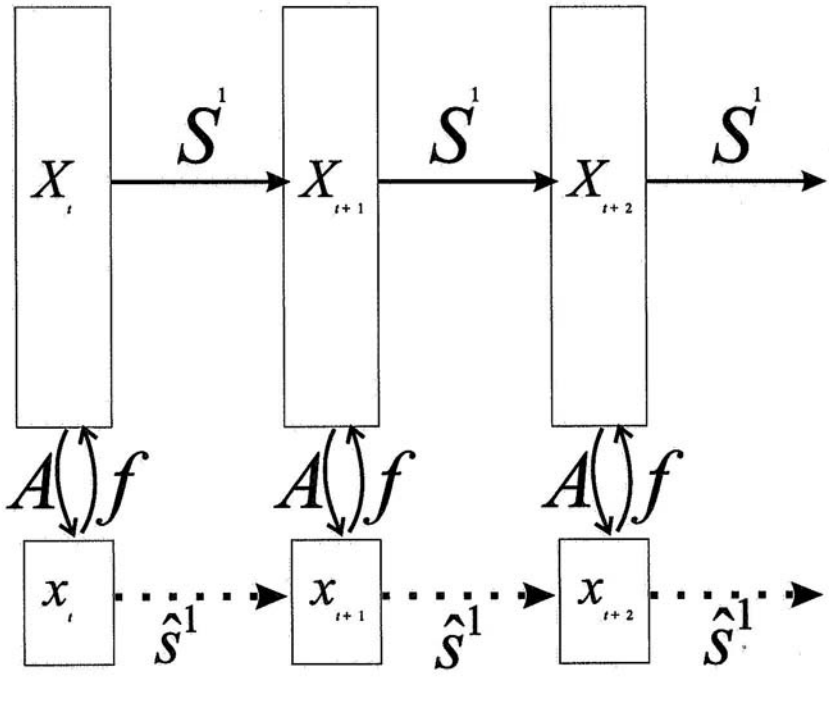


Figure 14.1. A conceptual diagram of a simulation model (Reproduced with the permission of American Agricultural Economics Association).

The top half of Figure 14.1 presents the dynamics of the state variables in the simulation model while the bottom half of the figure represents a smaller set of state variables, x, which are used in the DP specification to approximate X. Let $x \in \boldsymbol{x} \subset \mathbb{R}^n$ be a vector of state variables chosen so that the approximating problem can be solved in a "reasonable" length of time while still capturing critical features of the simulation model. In our example, the vector x consists of aggregate stocks of juvenile and adult stocks. The vector-valued function A: $\boldsymbol{X} \rightarrow \boldsymbol{x}$, maps from the full set of state variables into the small set of variables. As in our example, where A simply adds up young and adult fish in the two depths, the function A typically takes the form of a linear aggregation or average. For each vector $X \in \boldsymbol{X}$ there is a single vector $x = A(X)$. However, the

converse is not true—there might be an infinite number of vectors *X* that could give rise to a particular vector *x*. Hence, $f(x)$ defines not a single vector *X*, but a set of vectors that might have given rise to *x* and the probability density function over that set. We say that $X \in f(x)$ if $A(X) = x$.

To identify the approximate value function, $v(x)$, a DP problem in the space of the aggregate state variables, $\boldsymbol{x}$, must be solved. For our example, $v(x)$ is the fishery's value as a function of the aggregate stocks. Finding this function requires the mappings s^1, from (x_t, Z_t) to x_{t+1}, and s^2, from (x_t, Z_t) to u_t, where $u_t = U(Z_t, X_t, \varepsilon_t)$. As shown in the Figure 14.1, the true mapping s^1 involves three steps: from x_t to X_t, then from X_t and Z_t to X_{t+1} through S^1, and finally from X_{t+1} to x_{t+1}, i.e., $s^1(x_t, Z_t) = A(S^1[f(x_t), Z_t])$. As we have noted, however, the mapping f is one-to-many. Hence, even if X^1 and $X^2 \in f(x)$, the vectors can differ so that $S^1(X^1, Z) \neq S^1(X^2, Z)$. Hence, like $f(x)$, the composite mapping $s^1(\cdot)$ is also a mapping into a probability space; given any value of x_t and a choice vector Z_t, there is a distribution of values for x_{t+1} that might result.[4]

The challenge of DPSim analysis, therefore, is the development of representations of simulation models to approximate the composite mappings s^1 and s^2. Regardless of how this is done, it is clear that the DP model is stochastic, even if the simulation model itself is deterministic. In general, the functions s^1 and s^2 can be deterministic only in the unlikely case $A^{-1}(x)$ exists.

There are two basic approaches that might be used to capture s^1 and s^2. Both approaches have in common that they begin by carrying out a wide range of simulation runs to generate observations of *X* that are used as "data". The traditional approach, which we call the "meta-modelling" approach, approximates the mappings s^1 and s^2 using postulated functional forms, say $\hat{s}^1(x_t, Z_t)$ and $\hat{s}^2(x_t, Z_t)$. For example, x_{t+1} might be predicted using a set of possibly high-order polynomials of x_t and Z_t. In the meta-modelling approach the coefficients of $\hat{s}^1(\cdot)$ and $\hat{s}^2(\cdot)$ are econometrically estimated using the "data" and these functions are then used to solve the DP problem. In our case study, therefore, the meta-model allows us to predict future aggregate stocks based on current aggregate stocks and the TAC level. This approach, represented by the dotted lines in Figure 14.1, has been applied by many researchers such as Bryant, Mjelde and Lacewell (1993), Watkins, Lu and Huang (1998), and van Kooten, Young, and Krautkraemer (1997).

In a meta-modelling specification, the simulation model is bypassed entirely once $\hat{s}^1(\cdot)$ and $\hat{s}^2(\cdot)$ are found. In the alternative, which we call the *direct approach*, the large simulation model is used directly in the course of solving the DP problem. To accomplish this, it is necessary to approximate the correspon-

[4]The argument would also hold for s^2 and u_t.

dence *f*, i.e. the arrow up from x_t to X_t in Figure 14.1. This provides a means by which to approximate the full vector of state variables X_t so that the large simulation model can be used directly to predict X_{t+1} and u_t.

In the direct approach the "data" generated from the simulation model are interpreted as observations of the joint probability distribution of X_t and x_t and econometric methods are used to estimate the one-to-many mapping, *f*. Once the mapping *f* is approximated, the direct approach involves three steps. First, given a value of x_t, the disaggregate vector X_t is predicted. Then the large simulation model is run to obtain an X_{t+1} and u_t. Finally, using the aggregation function *A*, x_{t+1} is found as $A(X_{t+1})$. In our case study these three steps are: 1) the 720 cohort-depth groups are predicted based on the aggregate stocks in x_t; 2) GBFSM is run incorporating a TAC policy, yielding a value of consumer and producer surplus and the cohort-depth groups in *t*+1, X_{t+1}; and 3) the stocks are aggregated, yielding, x_{t+1}, a prediction of the next-period's aggregate young and old fish stocks. This three-step process is repeated numerous times following standard Monte Carlo methods to obtain the distribution of u_t and x_{t+1}, which are then used in the standard way to solve the DP problem.

It is clear that although either the meta-modelling approach or the direct approach might be employed to solve the DPSim problem, neither approach will exactly solve the true DP problem. Both models find an approximate value function, $v(x_t)$, which calculates the value of the system as a function of the aggregate variables x_t instead of the true value function, which would be a function of X_t. Estimation errors are introduced because the simulation model is based on *X*, while the DP model is solved based on *x*. The meta-modelling approach introduces errors in the use of functional forms to approximate the complete mappings s^1 and s^2. The direct method introduces errors in the prediction of X_t. It is not possible to *a priori* rank the two methods. We can say, however, that the meta-modelling approach will have a clear advantage in terms of computational speed since it only has to evaluate the small set of functions incorporated in $\hat{s}^1(\cdot)$ and $\hat{s}^2(\cdot)$. In the remainder of this chapter we present our case study in more detail and use it to compare the two methods.

4. An empirical application of DPSim analysis: The Gulf of Mexico Red Snapper Fishery

4.1 The policy setting

Our empirical application of the methods outlined is for the red snapper fishery of the Gulf of Mexico. This fishery is an important economic asset: commercial harvests of red snapper were valued at nearly $12.0 million from 5.08 million pounds in 2000 (U.S. Department of Commerce) and recreational

harvests in the same year were estimated at 4.15 million pounds (Gulf of Mexico Fishery Management Council).

The two policy variables we consider here are the total allowable catch (TAC) for the red snapper fishery, and the distribution of the TAC between the commercial and recreational sectors. The TAC is a fundamentally dynamic choice variable. Reducing the TAC is essentially an investment decision: benefits today are foregone to increase the value of the fishery in the future. The value of this investment depends not only on biological factors, but also on the extent to which future policies respond to changes in the stock. Dynamic optimization is, therefore, essential to the identification of the appropriate TAC policies.

To some degree, the need for a dynamically optimal policy has been recognized by the Gulf of Mexico Fishery Management Council (the council) in their plan to rebuild the red snapper fishery. The management alternatives evaluated by the Council call for re-evaluation of the policies every five years based on updated indicators of the fishery's health. Although the management alternatives considered by the Council are somewhat more rigid and limited than what might be desirable under a dynamically optimal policy, it appears that they *will* react to the changing conditions over time.

4.2 The simulation model: GBFSM

The simulation model that is at the heart of our analysis is the General Bioeconomic Fishery Simulation Model (GBFSM). GBFSM permits a high degree of flexibility, allowing for multiple species, depths, areas, cohorts, fishing vessels, and more. The model was originally developed to predict how alternative management policies would affect fisheries (Grant, Isaakson, and Griffin) and has been used extensively for analyzing the effects of management policies in the Gulf of Mexico (Blomo *et al.*, 1978; Grant and Griffin, 1979; Griffin and Stoll, 1981; Griffin and Oliver, 1991; Gillig *et al.*, 2001). GBFSM consists of two main parts: a biological submodel and an economic submodel. The biological submodel represents the recruitment, growth, movement, and natural and fishing mortality of shrimp and finfish. The economic submodel, which includes a harvesting sector and a policy sector, represents the monetary and industry impact. Entry and exit in the commercial sector follow the standard bio-economic approach based on economic rents. Recreational effort is predicted using econometrically estimated recreation-demand functions (Gillig *et al.*, 2001). The model is parameterized using coefficients from the literature,

econometrically estimated functions, and calibration, so that to the greatest extent possible, the model replicates historical trends in the fishery.[5]

Forty GBFSM simulations were run, each for fifty years. These simulations generated simulated data that were then used to parameterize the direct and meta-modelling approaches. We also used this data to determine the relevant range for the simulations, so that the only values for the aggregate state variables were those that were in the neighborhood of values that were actually observed in the simulations.

4.3 Solving the DP problem: The direct method

The direct approachrequires estimation of functions mapping a state-vector, x, into a vector of states for the full set of state variables in the simulation model. Three aggregate variables are used: the recreational catch per day of effort (CPUE) in the previous year, x_c, the stocks at the beginning of the year of young red snapper (2-years old or less), x_y, and adult red snapper (3-years old and above), x_a. The CPUE variable, x_c, coincides exactly with a variable in the simulation model so that a one-to-one mapping is possible for this variable. For the population variables, however, the size of the array X is much larger than could be incorporated into a DP model. The red snapper population in GBFSM is composed of 720 depth-cohort groups. Using the simulated data, a conditional expectation function was estimated for each depth-cohort combination, x_{dc}:

$$E(\ln X_{dc}) = \alpha_{0dc} + \alpha_{1dc} \ln x_c + \alpha_{2dc} \ln x_y + \alpha_{3dc} \ln x_a \qquad (4.1)$$

The double-log specification was used to avoid the prediction of negative stocks.

Overall, these estimated equations were able to predict the cohort populations quite accurately. The goodness of fit is the most important diagnostic variable. As can be seen in Figure 14.2, for cohorts less than 13 years old, most of the R^2s exceeded 80%. Although the R^2s drop off for older cohorts, the failure of the model to predict these older stocks is less critical since the population is dominated by younger fish due to natural and fishing mortality. On average across the simulations, only about 5% of all fish were 14 years or older. These results indicate that the distribution of X_t conditional on x_t is relatively "tight" so that it is possible to predict X_t quite accurately, giving a high degree of confidence in the direct approachfor solving the DP model. Higher-order polynomial specifications of the approximating function were evaluated

[5]There are many details that are suppressed here. A detailed description of GBFSM and its calibration are available at http://gbfsm.tamu.edu. The discount rate used in the analysis is the federally mandated rate of 7% per annum (Executive Office of the President).

and these specifications improved the overall goodness of fit. However, such specifications resulted in some unrealistic predicted values of X, particularly at points on the edge of the state space. Monte Carlo methods were used to esti-

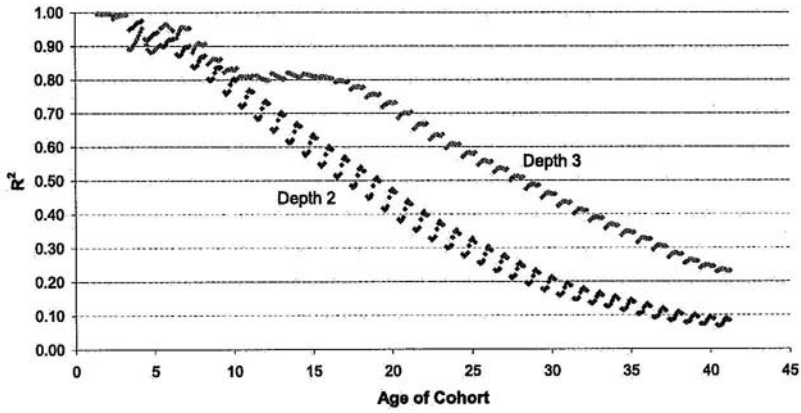


Figure 14.2. R^2 values from equations used to predict cohort populations in the direct method.

mate the expectations $Ev(x_{t+1})$ and Eu_t. The distribution $f(x_t)$ is made up of 720 jointly distributed variables. This distribution was estimated by drawing n_{mc} =50 observations of vectors of residuals from the econometric model so that the i^{th} observation of each depth-cohort group was

$$\ln \hat{X}^i_{dc} = \alpha_{0dc} + \alpha_{1dc} \ln x^i_c + \alpha_{2dc} \ln x^i_y + \alpha_{3dc} \ln x^i_a + \varepsilon^i_{dc}\,. \tag{4.2}$$

in this case since it avoids the need to make an ad hoc assumption about the distribution (e.g., joint normality), which almost certainly does not hold true.

The vectors predicted using equation (4.2) do not typically sum exactly to their associated aggregate variable, e.g., $x^i_y \neq \sum_{\forall d} \sum_{c \leq 2} \hat{X}^i_{dc}$. Hence, after obtaining this initial prediction $\hat{X}^i_{dc}$ using equation (4.2), all cohorts were proportionally scaled up or down so that for the cohort vectors used in the GBFSM it holds that $A(X) = x$.

A three-dimensional $20 \times 20 \times 20$ uniform grid was generated to approximate the state space. However, in the simulated data there was a high degree of correlation among the three aggregate state variables and many potential combinations of state variables were never observed. To take advantage of this and to avoid making prediction in portions of the state space that are never realistically observed, only vertices of the grid that were adjacent to points in the data were included in the state space. This reduced the number of points in the state space from 8,000 vertices to only 519.

4.4 Solving the DP problem: The meta-modelling approach

The meta modelling specification was parameterized with the same data used for the direct method that is discussed above. Four functions were estimated, three to predict each of the state variables, x_{t+1}, and a fourth to predict the fishery's annual surplus. The independent variables in each of these functions are the state of the fishery in year *t* as described by the aggregate state variables, x_t, and the control variables, Z_t. We experimented with a variety of specifications, all of which can be represented by the function:

$$\hat{y}_k = a_k + \sum_{i=1}^{5}\sum_{j=1}^{n_p} b_{ijk}(v_i)^j + \sum_{i=1}^{4}\sum_{j=i+1}^{5} c_{ijk}(v_i v_j),$$

where $\hat{y}_k$ is a predicted variable, an element of x_{t+1} or surplus in period *t*, n_p is the order of the polynomial, and the *v*'s are the independent variables with v_1 through v_3 equal to the elements of the state space, and v_4 and v_5 equal to values of the control variables. Double-log specifications were also used, in which case, for example, $\hat{y}_k = \ln(\hat{x}^1_{t+1})$ and $v_1 = \ln(x^1_t)$. Using second and third-order polynomials, a total of four specifications were estimated.

Table 14.1. R-squared Values for the Specifications Used in the Meta-Modelling Approach

	Equation			
	u_t	x_{t+1}		
Specification	*(benefits)*	x_c	x_y	x_a
Second order polynomial	0.889	0.990	0.992	0.998
Double-log second order polynomial	0.965	0.978	0.990	0.996
Third order polynomial	0.944	0.991	0.993	0.999
Double-log third order polynomial	0.966	0.981	0.990	0.996

The estimated functions for the meta-modelling approach appear to offer a strong statistical foundation for simulating the simulation model. The R^2 values for each of the 16 equations are presented in Table 14.1, and clearly all the specifications are able to explain most of the variation found in the data. As we will discuss below, the second-order non-log specification was used to solve the meta-modelling application of the DP problem. The coefficients for this specification are presented in Table 14.2. In this specification, approximately two-thirds of the parameters are significantly different from zero at the 5% level and similar levels of significance held for the other specifications as

well. Interestingly, the vast majority (85%) of the coefficients on the third-order polynomials were significantly different from zero at the 5% level.

Table 14.2. Coefficients for Second-Order Polynomial Specification of Meta-Modelling Parameters (Standard Errors in Parentheses)

	Benefits		x_c		x_y		x_a	
	Est.	*St. Err.*	*Est.*	*St. Err.*	*Est.*	*St. Err.*	*Est.*	*St. Err.*
Intercept	17.147	1.660	–1.227	0.214	16.542	1.364	–2.26	0.456
x_c	1.074	1.209	–0.499	0.156	–4.476	0.993	0.231	0.332
x_c^2	–4.093	0.358	–0.102	0.046	1.898	0.294	–0.403	0.098
x_y	0.047	0.086	0.106	0.011	0.101	0.070	0.322	0.024
x_y^2	0.002	0.001	–0.002	0.000	0.011	0.001	–0.004	0.000
x_a	0.047	0.300	0.203	0.039	2.578	0.246	0.233	0.082
x_a^2	–0.116	0.017	0.001	0.002	0.077	0.014	–0.012	0.005
Z_T	0.320	0.069	–0.031	0.009	0.033	0.001	–0.142	0.019
Z_T^2	–0.003	0.001	0.000	0.000	0.001	0.001	0.001	0.000
Z_D	0.549	0.101	0.017	0.013	–0.038	0.083	–0.036	0.028
Z_D^2	–0.017	0.003	0.000	0.000	0.005	0.025	0.004	0.028
$x_c \cdot x_y$	0.097	0.03	0.033	0.004	0.007	0.025	0.039	0.008
$x_c \cdot x_a$	1.457	0.154	–0.011	0.020	–0.683	0.127	0.087	0.042
$x_c \cdot Z_T$	–0.079	0.038	–0.015	0.005	0.039	0.031	–0.035	0.010
$x_c \cdot Z_D$	0.289	0.054	0.006	0.007	–0.114	0.044	–0.022	0.015
$x_y \cdot x_a$	–0.035	0.007	–0.001	0.001	–0.033	0.006	0.009	0.002
$x_y \cdot Z_T$	0.003	0.002	0.001	0.000	–0.003	0.001	0.002	0.000
$x_y \cdot Z_D$	–0.003	0.003	0.000	0.000	0.001	0.002	–0.004	0.001
$x_a \cdot Z_T$	0.013	0.009	0.002	0.001	–0.003	0.007	0.003	0.002
$x_a \cdot Z_D$	–0.028	0.012	–0.005	0.002	0.020	0.010	0.002	0.003
$Z_T \cdot Z_D$	–0.016	0.004	0.001	0.001	–0.003	0.003	0.005	0.001
R^2	0.889		0.99		0.992		0.998	

Monte Carlo simulation was carried out taking into account both the variation in the parameter estimates and in the residuals accounting for cross-equation correlation assuming that both parameter estimates and the residuals are jointly distributed normally.

5. Results

Figures 14.3 and 14.4 depict the value functions obtained using the direct and the meta-modelling approaches. Of the meta-modelling specifications used, only the second-order polynomial non-log specification is presented. As we note above, higher order polynomials can be unstable in the DP algorithm as they give highly variable predictions at the boundaries of the state space

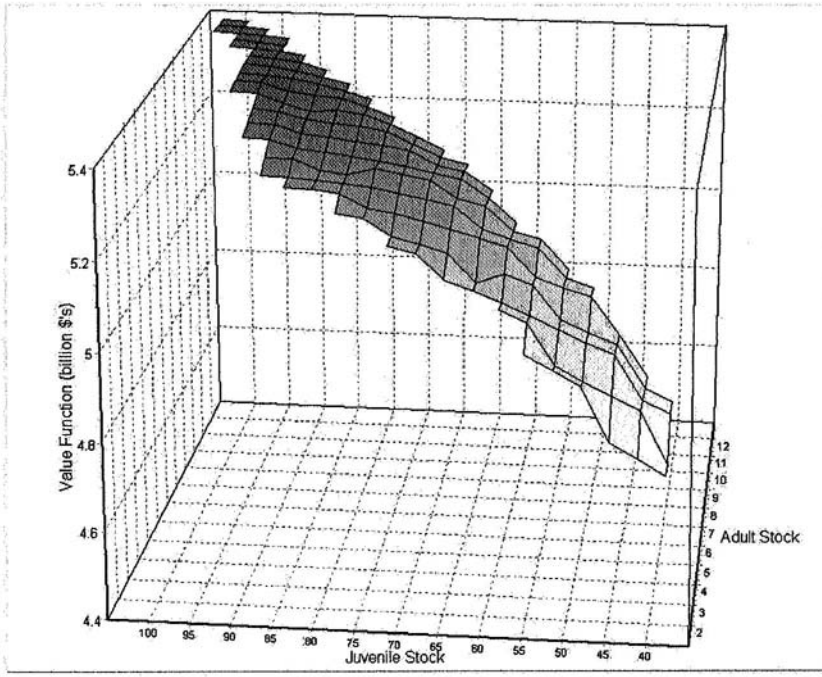


Figure 14.3. Value function obtained using the direct method.

and this is what occurred in the model specified here; the log and third-order polynomial specifications did not converge so those results are not presented. As seen in Figure 14.4, even the second-order specification gives rather unintuitive results with the value function falling as higher stocks are reached. The failure to achieve convergence in some of the meta-modelling specifications may surprise some readers who have been assured of the convergence of the successive approximation approach to solving DP problems. However, the contraction mapping properties of the algorithm do not necessarily hold when the value function is predicted over a continuous state space. An alternative solution approach that might have been more successful would be to implement the collocation method advocated by Judd (1998). However, this method requires the specification of a functional form for the value function. The functional form preferred by Judd is a Chebyshev polynomial. However, since Chebyshev polynomials must be defined over a rectangular grid, they are unsuitable for the non-uniform grid that we favor in this application to avoid extrapolation to unrealistic state-space combination.

Figure 14.5 presents the DPSim results, the simulated optimal policy paths for both the direct and meta-modelling approaches. At each point along these

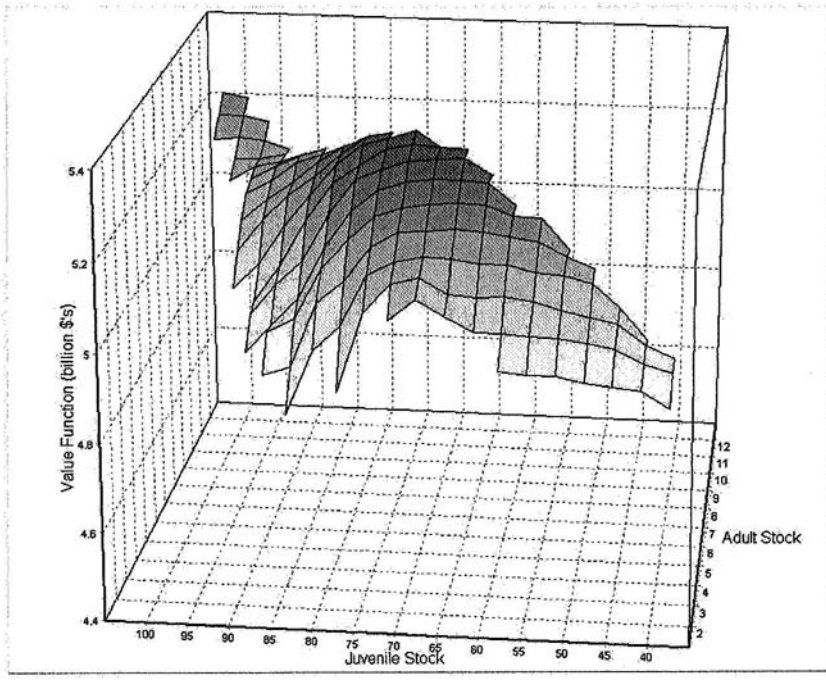


Figure 14.4. Value function obtained using the meta-modelling approach, using coefficients from Table 14.2.

paths an optimal policy is found by solving the static optimization problem, equation (2.1), with u_t and x_{t+1} found by GBFSM and the value function, $v(x_{t+1})$, obtained from the solution of the optimization problem as presented in Figures 14.3 and 14.4. Once an optimal policy is identified, that policy is introduced into GBFSM to determine the state of the fishery in the next year. Both simulations start with the same stock at t=0, but because the policies differ the predicted stocks underlying the simulated policy paths differ. As seen in Figure 14.6, the red snapper stock recovers significantly when policies from the direct-method are followed, while the stock stays essentially stable if policies from the meta-modelling approach are used. The reason for the differences in the policies chosen must be attributable to the value functions presented in Figures 14.3 and 14.4. Because of the upward slope in the value function identified using the direct method, there is an identified benefit to increasing the stock. In contrast, the value function found using meta-modelling approach was not monotonically increasing so that stock enhancements were not treated favorably when equation (2.1) was evaluated to identify the optimal policies at each point in time.

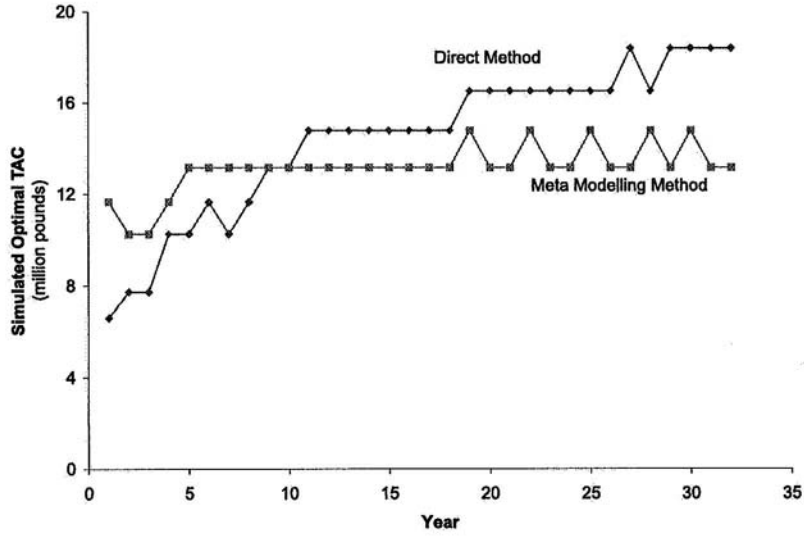


Figure 14.5. Simulated optimal policy paths with direct and meta-modelling specifications.

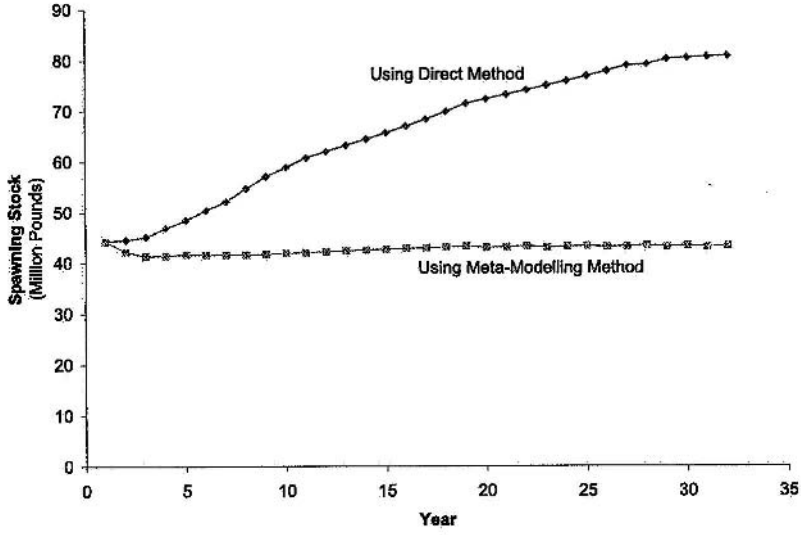


Figure 14.6. Simulated spawning stock along paths using direct and meta-modelling approaches.

The above figures give reason to believe that in this application the direct method is more reliable than the meta-modelling method. What is the reason for this? Recall that either method introduces potential errors. The direct method introduces errors through the incorrect prediction of the disaggregated stocks in period *t*, which then leads to errors in the prediction of u_t and x_{t+1}. The meta method introduces errors more directly through the inaccurate prediction of u_t and x_{t+1}. One cannot say *a priori* which errors will be more important. To compare the relative accuracy of the two methods, we compared

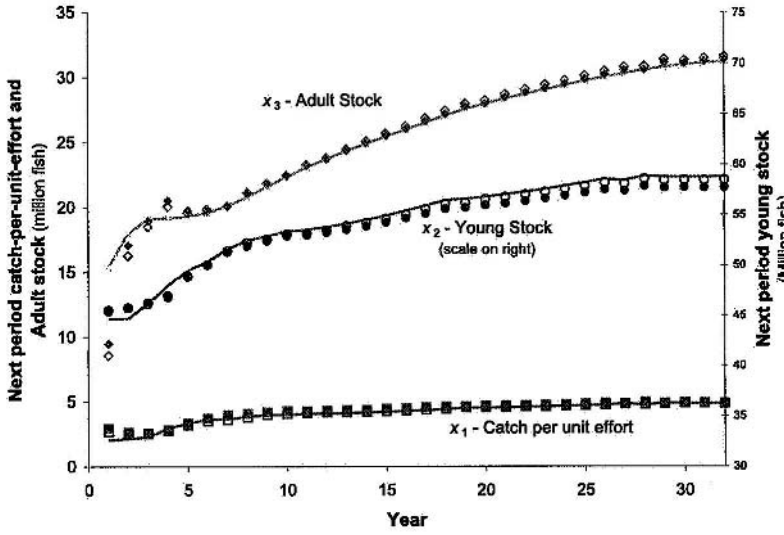


Figure 14.7. Comparison of actual and predicted next-period state variables. Actual values from GBFSM (lines), Direct-method predictions (Empty symbols) and Meta-modelling predictions (solid symbols).

the mean predicted values for surplus and each of the state variables with the values that actually were found in the simulation model along the direct method policy path in Figure 14.5.[6] The results of these comparisons are presented in Figures 14.7 and 14.8. As seen in Figure 14.7, either method is reasonably precise in the prediction of the next period state variables and neither method is preferred. Both methods err significantly in the prediction of the adult stock in the first few periods, but after four periods none of the prediction of the next period's stock differed by more than five percent and the largest error in the prediction of the catch per unit effort was just over eight percent. The re-

[6]The results for the meta-method path are qualitatively similar to those from the direct-method path, but are less interesting because of the relatively constant optimal TAC policies used along that path.

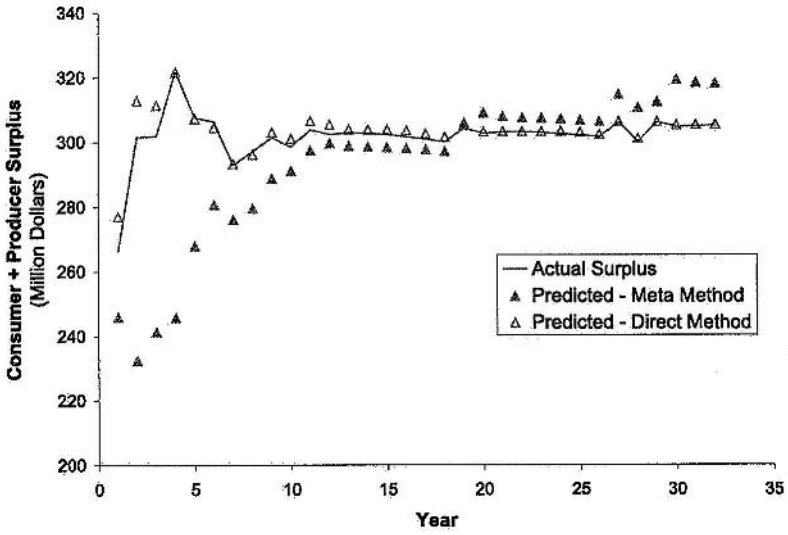


Figure 14.8. Comparison of actual and predicted surplus. Actual values from GBFSM (line), Direct-method predictions (Empty symbols) and Meta-modelling predictions (solid symbols).

sults for surplus in each period, however, show important differences between the two methods. The direct method is quite accurate in its prediction of u_t, never deviating by more than four percent from the value actually predicted by GBFSM. This indicates either that the model is quite good at predicting the disaggregated stocks (as is suggested by Figure 14.2 above) or that the distribution of the predicted returns in a year is not very sensitive to the distribution of the cohorts, or both. In contrast, the meta-modelling approach is quite inaccurate in its prediction of returns in a year, particularly in the first several years when the optimal TACs are low. In the first several years, the prediction of u_t are as much as 23% below the value predicted by GBFSM. This poor prediction of the surplus when the TAC and stocks are low, is probably an important part of the reason that low TACs are avoided in the predicted optimal path found using the meta-modelling approach as shown in Figure 14.5.

6. Conclusions

This chapter has presented an approach to DPSim modelling, conducting dynamic optimization with a large simulation model. When analysts have sought to unify a simulation model to dynamic optimization, they have typically taken the meta-modelling approach. We offer an alternative, which we call the direct method. The meta-modelling approach has an important advan-

tage in that its computational burden is significantly less than for the direct method. But we find that this benefit may have costs - in our application the direct method provided better results. First, the value function found using the direct method was much more plausible and did not violate monotonicity as does the meta-modelling approach. Secondly, the policy path seemed intuitively more reasonable, with low TACs initially, and higher TACs once the stock has recovered. Finally, and most importantly, we found that the direct method's prediction of annual surplus was much better than the prediction from the meta-modelling approach.

As was found in WWG, we believe that optimal TAC management for the Gulf of Mexico's red snapper fishery will involve reductions in the TAC in the short term, followed by expansion in the TAC in the long-term. This policy recommendation follows from the results of the direct method. If we instead used the meta-modelling approach, the policy recommendation would be quite different: greater harvests in the short term and lower harvests in the long term. The solution method makes a difference.

We wish to emphasize two important points. First, in DPSim analysis the optimal simulation runs are carried out with the full simulation model; regardless of the approach taken, the DP models' results are used only to calculate the value of future stocks so that optimal policies at each point in time can be identified. Secondly, although the direct method was preferred here, either approach is a plausible way to work around the curse of dimensionality. Unless general results can be found on which of the two methods is preferred under what conditions, we recommend that analysts use and compare both approaches when attempting to carry out dynamic optimization linked to large simulation models.

Chapter 15

AN EXPOSITION OF STRUCTURAL ESTIMATION OF DISCRETE DYNAMIC DECISION PROCESSES

Bill Provencher
University of Wisconsin-Madison
provencher@aae.wisc.edu

Kenneth A. Baerenklau
University of California-Riverside
ken.baerenklau@ucr.edu

Abstract In the analysis of dynamic decision problems, the vast majority of the literature has focused on normative aspects: what should a resource manager do to maximize a particular objective function? Rarely have resource economists attempted to answer the positive question: what decision problem does a resource manager actually solve? There are several candidate explanations for this emphasis on normative modeling, but in this chapter we take the perspective that although structural estimation of discrete dynamic decision problems is not especially difficult, it is difficult enough that most analysts require some explanation of why they should bother with it at all. We develop a general model of a discrete dynamic decision problem, distill it to a tractable form, and present the estimation methodology. We then provide an empirical example and investigate the implications of using a reduced-form static model of behavior when the underlying data-generating process is dynamic.

Keywords: Dynamic optimization, structural estimation, multinomial logit.

1. Introduction

In the analysis of dynamic decision problems (DDP's), the vast majority of the literature has focused on the normative aspects of such problems: what resource managers should do to maximize a particular objective function. Ex-

amples are ubiquitous in the literature, covering a range of issues from the land development decision, to the optimal rate of livestock grazing on arid land, to the well-known Faustmann model (and its many derivatives) of when to harvest a timber stand. Rarely have resource economists directly employed their structural dynamic models in the pursuit of the positive question, What decision problem do resource managers actually solve? The typical empirical study instead presents dynamic structural models in a theoretical discussion to motivate the choice of regressors in a reduced-form statistical analysis.[1] The primary explanation for this weakness in the literature is that such estimation is computationally very difficult. It usually requires a large amount of computer processing time, and perhaps more importantly, it requires a substantial amount of difficult programming by the analyst. Another explanation is that in the estimation of a parametric structural model the analyst is required to declare specific functional forms—state variable transition functions and benefit/utility functions, for instance—for which there is very little supporting evidence. This alone may give the analyst pause. A third explanation provides the justification for this chapter: by and large, resource economists remain fundamentally unfamiliar with the methods used to estimate the structure of DDP's.

This chapter takes the perspective that structural estimation of discrete DDP's is not especially difficult, and yet is difficult enough that most analysts require some explanation of why they should bother with it at all. In the next section we present the basic logic of structural estimation of discrete dynamic decision problems by developing the general case and distilling it to the tractable form most often found in the literature. In section 3 we illustrate the method using as a case study a modified version of the timber harvesting problem examined by Brazee and Mendelsohn (1988) (hereafter, the modified BM model). In section 4 we discuss, in the context of the modified BM model, the estimation of reduced-form and static models in lieu of structural estimation of the "true" dynamic model. We show that reduced-form estimation recovers the behavior engendered by the modified BM model with surprising ease, but that such estimation lacks policy-relevant empirical content. Moreover, insofar as one can relabel reduced-form estimation as structural estimation of a static decision model, it is not possible to distinguish empirically whether microeconomic data is generated by static or dynamic behavior, without maintained structural assumptions about the form of intra-period benefit (utility, profit) functions, and even then it may not be possible. In the last section we briefly discuss future research directions and opportunities.

[1]Some exceptions in resource economics are Provencher (1995a,b), Provencher and Bishop (1997), and Howitt *et al.* (2002).

2. Structural Estimation of a Discrete Dynamic Decision Problem

Suppose an agent can make one of $S+1$ choices in each period t. We index these choices by $s = 0, \ldots, S$. The amount of the payoff in a period depends on the state variables $\mathbf{x}$, which evolve over time according to the probability density function $f(\mathbf{x}_{t+1}|\mathbf{x}_t, s)$; the decision-specific random shock ε^s; and perhaps other variables, denoted by $\mathbf{y}$, that are invariant over time. The random shocks are contemporaneously observed by the agent but never observed by the analyst. For simplicity we assume these shocks are additive, and identically and independently distributed over time. Letting $R^s(\mathbf{x}_t, \mathbf{y}) + \varepsilon_t^s$ denote the decision-specific payoff at time t, the relevant decision problem can be stated in Bellman's form,

$$v(\mathbf{x}_t, \boldsymbol{\varepsilon}_t, \mathbf{y}) = \max_s \left[R^s(\mathbf{x}_t, \mathbf{y}) + \varepsilon_t^s + \beta E_{\mathbf{x},\boldsymbol{\varepsilon}|\mathbf{x}_t,s} \left\{ v(\mathbf{x}_{t+1}, \boldsymbol{\varepsilon}_{t+1}, \mathbf{y}) \right\} \right], \quad (2.1)$$

where the expectation of future value is taken over both the observed state variables $\mathbf{x}$ and unobserved state variables $\boldsymbol{\varepsilon} = \left(\varepsilon^0, ..., \varepsilon^S\right)$, conditional on the current values of observed state variables and the decision s.

The problem in (2.1) reflects Bellman's principle of optimality; in particular, it implicitly recognizes that regardless of the decision at time t, optimal decisions are made in the future. This problem can be solved via the recursive methods of dynamic programming. The solution is an optimal decision rule $s(\mathbf{x}, \boldsymbol{\varepsilon}, \mathbf{y}; \Gamma)$, where Γ is the set of structural parameters associated with the decision problem. This includes the parameters implicit in $R^s(\mathbf{x}, \mathbf{y})$, the parameters of the density function $f(\cdot)$, the parameters of the distribution of $\boldsymbol{\varepsilon}$, and the discount factor β.

Now suppose there exist observations over time $t = 1, \ldots, T$, and across agents $j = 1, \ldots, J$, of decisions S_{jt} and variables $\mathbf{x}_{jt}$ and $\mathbf{y}_j$. Then defining,

$$v_t^s = R^s(\mathbf{x}_t, \mathbf{y}) + E_{\mathbf{x},\varepsilon|\mathbf{x}_t,s} \left\{ v(\mathbf{x}_{t+1}, \boldsymbol{\varepsilon}_{t+1}, \mathbf{y}) \right\}$$

the probability that agent j makes choice 0 at time t is given by,

$$\begin{aligned} \Pr(s_{jt} = 0|\mathbf{x}_{jt}, \mathbf{y}_j; \Gamma) &= \Pr\left(v_t^0 + \varepsilon_t^0 \geqslant v_t^1 + \varepsilon_t^1; ...; v_t^0 + \varepsilon_t^0 \geqslant v_t^S + \varepsilon_t^S\right) \\ &= \int_{\varepsilon_t^0 \geqslant \Delta v_t^{1,0} + \varepsilon_t^1} \cdots \int_{\varepsilon_t^0 \geqslant \Delta v_t^{S,0} + \varepsilon_t^S} g\left(\varepsilon^1\right) \cdots g\left(\varepsilon^S\right) d\varepsilon^1 ... d\varepsilon^S, \quad (2.2) \end{aligned}$$

where $\Delta v_t^{s,0} = v_t^s - v_t^0$, and $g(\cdot)$ is the common distribution of the unobserved state variables ε_t^s. The probability that in period t agent j makes a different choice $s_{jt} = 1, \ldots, S$ can be constructed in similar fashion.

Denoting by s_{jt}^O the observed decision by agent j in period t, and assuming agent decisions are independent of one another, the likelihood function is

$$L(\Gamma) = \prod_{j=1}^{J} \prod_{t=1}^{T} \Pr\left(s_{jt}^O | \mathbf{x}_{jt}, \mathbf{y}_j; \Gamma\right). \tag{2.3}$$

Estimation of the parameters Γ is a complicated and computationally intensive exercise, much more so than is found with typical static models. The search for the parameter vector that maximizes the likelihood function involves solving the dynamic decision problem (2.1) each time new parameter values are evaluated in the search; this is apparent by the presence of v_t^s in the limit of integration in (2.2). Solving the decision problem (2.1) requires dynamic programming, and so a DP simulation must be run *each time* new parameter values are evaluated in the search for the maximum likelihood value. Maximum likelihood estimation thus involves nesting an "inner" dynamic programming algorithm within an "outer" hill-climbing algorithm.

As a general matter, the inclusion of the unobserved state variables seriously impacts the tractability of the dynamic programming problem (2.1), as the expectation of the value function must be taken over the random components of the observed state vector $\mathbf{x}$ *and* the S-dimensional vector $\boldsymbol{\varepsilon}$. For all but the smallest problems this is not computationally feasible given that (2.1) must be solved many times during the course of the search for Γ^*, the likelihood-maxmizing value of Γ. Making matters even more difficult is the multi-dimensional integration in (2.1) associated with each observation. The reader familiar with the literature on random utility models will recognize that the properties of the Gumbel distribution can be used to resolve the difficulty of the integration in (2.2). In a seminal paper Rust (1989) shows that these same properties of the Gumbel distribution can be used to resolve the difficulty of the integration implicit in the expectation in (2.1), as follows.

We assume that $\boldsymbol{\varepsilon}$ is iid Gumbel-distributed with location parameters $\boldsymbol{\theta}_\varepsilon = \left(\theta^0, ..., \theta^S\right)$ and common scale parameter η_ε. Moreover, for expositional reasons we define

$$V(\mathbf{x}_t, s_t; \Gamma) = E_{\mathbf{x},\boldsymbol{\varepsilon}|\mathbf{x}_t,s_t}\left\{v(\mathbf{x}_{t+1}, \boldsymbol{\varepsilon}_{t+1}, \mathbf{y})\right\}, \tag{2.4}$$

in which case the decision problem (2.1) can be restated,

$$v(\mathbf{x}_t, \boldsymbol{\varepsilon}_t, \mathbf{y}) = \max_s \left[R^s(\mathbf{x}_t, \mathbf{y}) + \varepsilon_t^s + \beta V(\mathbf{x}_t, s; \Gamma)\right]. \tag{2.5}$$

Then from the standard properties of the Gumbel distribution (see Ben-Akiva and Lerman (1985)), integrating both sides of (2.5) with respect to $\boldsymbol{\varepsilon}$ on day

$t+1$ yields,

$$E_{\varepsilon} v\left(\mathbf{x}_{t+1}, \varepsilon_{t+1}, \mathbf{y}\right) = \frac{1}{\eta_{\varepsilon}} \ln\left(\sum_{s=0}^{S} \mathrm{e}^{\eta_{\varepsilon}\left(R^{s}\left(\mathbf{x}_{t+1}, \mathbf{y}\right)+\theta^{s}+\beta V\left(\mathbf{x}_{t+1}, s; \Gamma\right)\right)}\right) + \frac{\gamma}{\eta_{\varepsilon}}, \tag{2.6}$$

where γ is Euler's constant ($\approx$0.577). Substitution of (2.6) into (2.4) at time $t+1$ yields,

$$V\left(\mathbf{x}_{t}, s_{t}; \Gamma\right) = E_{\mathbf{x}|\mathbf{x}_{t}, s_{t}} \frac{1}{\eta_{\varepsilon}} \ln\left(\sum_{s=0}^{S} \mathrm{e}^{\eta_{\varepsilon}\left(R^{s}\left(\mathbf{x}_{t+1}, \mathbf{y}\right)+\theta^{s}+\beta V\left(\mathbf{x}_{t+1}, s; \Gamma\right)\right)}\right) + \frac{\gamma}{\eta_{\varepsilon}}. \tag{2.7}$$

If $v(\cdot)$ is known, the determination of $V(\cdot)$—necessary to solve the decision problem (2.5)—is now a relatively simple affair involving integration over the random elements in the observable state vector $\mathbf{x}_{t+1}$. In practice, though, $v(\cdot)$ is not known; but $V(\cdot)$ can be determined via successive approximation (backwards recursion) due to the contraction mapping properties of Bellman's form. Initially $V(\cdot)$ is set identically equal to zero, and the value function $v(\cdot)$ is approximated from (2.5) using standard techniques for approximating a function, such as linear or Chebychev polynomial interpolation. Then the expectation of $v(\cdot)$ with respect to $\boldsymbol{\varepsilon}$ is calculated from (2.6), and an approximation of $V(\cdot)$ is obtained from (2.7) using standard techniques of function interpolation and numerical integration. The approximation of $V(\cdot)$ is then used in (2.5) to obtain a new estimate of $v(\cdot)$, and so on. This iterative mechanism terminates under conditions for convergence, such as sufficiently small changes across iterations in the optimal decision rule, $s\left(\mathbf{x}, \boldsymbol{\varepsilon}, \mathbf{y}; \Gamma\right)$.

It deserves emphasis that the algorithm described above identifies the optimal decision rule – and, more to the point, the associated expected value function $V(\cdot|\Gamma)$ – *conditional* on a particular set of parameters Γ. With the expected value function in hand, the assumption that $\boldsymbol{\varepsilon}$ is iid Gumbel-distributed allows a restatement of (2.2) in the analytical form,

$$\begin{aligned}\Pr\left(s_{jt}=0|\mathbf{x}_{jt}, \mathbf{y}_{j}; \Gamma\right) &= \Pr\left(v_{t}^{0}+\varepsilon_{t}^{0} \geqslant v_{t}^{1}+\varepsilon_{t}^{1}; \ldots; v_{t}^{0}+\varepsilon_{t}^{0} \geqslant v_{t}^{S}+\varepsilon_{t}^{S}\right) \\ &= \frac{e^{\eta_{\varepsilon}\left(R^{0}\left(\mathbf{x}_{t}, \mathbf{y}\right)+\theta^{0}+\beta V\left(\mathbf{x}_{t}, 0; \Gamma\right)\right)}}{\sum_{s=0}^{S} e^{\eta_{\varepsilon}\left(R^{s}\left(\mathbf{x}_{t}, \mathbf{y}\right)+\theta^{s}+\beta V\left(\mathbf{x}_{t}, s; \Gamma\right)\right)}}.\end{aligned} \tag{2.8}$$

The upshot is that although it remains the case that a DP algorithm must be nested within the estimation algorithm used to find Γ^{*}, assuming the unobserved state variables are iid Gumbel-distributed greatly simplifies the algorithm.

Examination of (2.8) also shows that a well-documented weakness of static multinomial logit models—the property of *independence of irrelevant alternatives* (IIA)—does not hold in a dynamic model employing Gumbel-distributed random variables. The IIA property derives from the fact that, from the perspective of the analyst, the odds that one alternative is chosen over another in a static model depends only on the attributes of the two alternatives. So, for instance, in the example presented by Bockstael, the IIA property implies the unlikely result that the odds of visiting a saltwater beach instead of a freshwater lake does not depend on whether a third beach is itself a saltwater or freshwater site. However, the log odds ratio for the dynamic model can be stated,

$$\log\left(\frac{\Pr\left(s_{jt}=i\right)}{\Pr\left(s_{jt}=k\right)}\right)=\eta_{\varepsilon}\left(R^{i}\left(\mathbf{x}_{t},\mathbf{y}\right)+\theta^{i}+\beta V\left(\mathbf{x}_{t+1},i;\Gamma\right)\right)- \\ \eta_{\varepsilon}\left(R^{k}\left(\mathbf{x}_{t},\mathbf{y}\right)+\theta^{k}+\beta V\left(\mathbf{x}_{t+1},k;\Gamma\right)\right) \tag{2.9}$$

And so, by virtue of the presence of $\mathbf{x}_t$ and $\mathbf{y}$ in $V(\cdot)$, the odds of choosing alternative i over alternative k depends on the attributes (state of nature) of all the alternatives.

3. An Illustration: The Brazee-Mendelsohn Timber Harvest Problem

In this section we attempt to clarify the discussion above by examining a modification of the Brazee-Mendelsohn (BM) model of optimal harvesting when timber prices are stochastic (Brazee and Mendelsohn 1988). This is a simple but illuminating example. In the original BM model the forest owner faces, in each period, the binary decision to either harvest a timber stand or to postpone harvest. This decision depends on two state variables: the age of the timber stand a, and the price of timber p. Timber volume at stand age a is given by

$$W\left(a\right)=e^{\phi_{1}-\frac{\phi_{2}}{a}}, \tag{3.1}$$

where ϕ_1 and ϕ_2 are growth parameters.

Timber prices are independent and identically normally-distributed with mean price μ_p and standard deviation σ_p; we denote the probability density function by $g\left(\mu_p,\sigma_p\right)$. The forest owner solves the the problem,

$$v(a_t,p_t)=\max\left[\beta E_p v(a_t+1,p_{t+1}),p_t W(a_t)-c+\beta E_p v(1,p_{t+1})\right] \tag{3.2}$$

where c is the cost to harvest and replant. The problem is easily solved by ordinary dynamic programming techniques. The optimal harvest policy $s(p,a)$ is a reservation price policy in which, for a given stand age, the forest owner

harvests if and only if the observed timber price is above a reservation price. Graphically this is represented by the harvest isocline shown in Figure 15.1; harvest occurs for all combinations of price and stand age above the isocline. When applying this model to actual data the analyst must account for the pos-

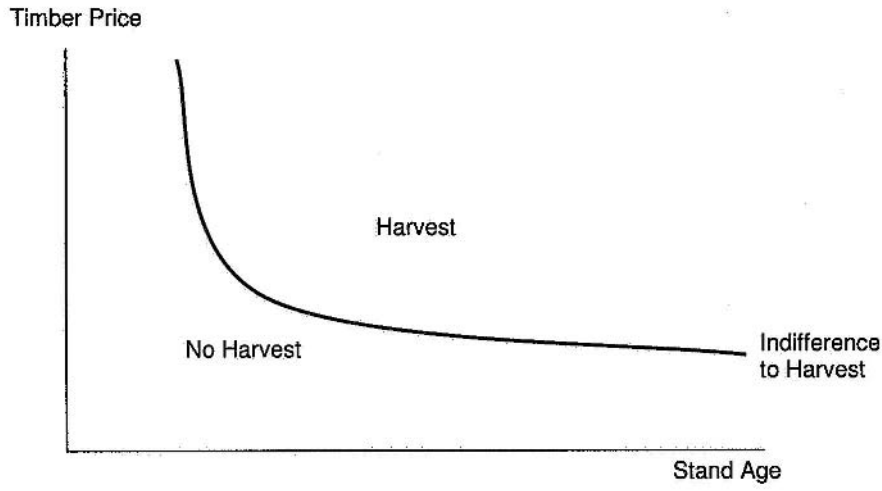


Figure 15.1. Brazee-Mendelsohn Optimal Harvest Policy

sibility that observed harvest decisions deviate from the normative decision rule. This is accomplished by introducing the decision-specific state variables $\boldsymbol{\varepsilon} = (\varepsilon^0, \varepsilon^1)$. This paradox—that the normative model does not fully describe the decision problem faced by "real world" dynamic optimizers, and so the normative model is, as a practical matter, not normative at all—is almost invariably the case for dynamic problems of resource allocation, because normative models distill a rich decision environment to a world fully described by just several state variables. The addition of the unobserved state variables can be viewed as a somewhat crude attempt to account for the richness of the real world.

The modified decision problem is,

$$\begin{aligned} &V\left(a_t, p_t, \boldsymbol{\varepsilon}_t\right) \\ &= \max\left[\varepsilon_t^0 + \beta E_p v\left(a_t + 1, p_{t+1}\right),\ p_t W\left(a_t\right) - c + \varepsilon_t^1 + \beta E_p v\left(1, p_{t+1}\right)\right] \\ &= \max[\varepsilon_t^0 + \beta V\left(a_t, 0; \Gamma\right),\ p_t W\left(a_t\right) - c + \varepsilon_t^1 + \beta V\left(a_t, 1; \Gamma\right)] \end{aligned} \tag{3.3}$$

where the control variable takes a value of 1 if the stand is harvested and 0 otherwise. Note that (3.3) is a special case of (2.1) in which only one of the observed state variables is stochastic (p), and the distribution of the state variable

is not conditional on the current value. In this framework, one possible interpretation of ε is that it is the utility received from the standing forest (though in this simple case, the utility received is not conditional on stand age – generally an unrealistic specification).

Assuming that ε is iid Gumbel-distributed with choice-specific location parameters $\theta^i, i = 0, 1$, and common scale parameter η_ε, the expected value function can be stated (using a modified version of (2.7):

$$\begin{aligned} V(a_t, 0; \Gamma) &= \int_{-\infty}^{\infty} \frac{1}{\eta_\varepsilon} \ln\left(M^0\right) g(p; \mu_p, \sigma_p)\, dp + \frac{\gamma}{\eta_\varepsilon} \\ V(a_t, 1; \Gamma) &= \int_{-\infty}^{\infty} \frac{1}{\eta_\varepsilon} \ln\left(M^1\right) g(p; \mu_p, \sigma_p)\, dp + \frac{\gamma}{\eta_\varepsilon}, \end{aligned} \tag{3.4}$$

where:

$$M^0 = e^{\eta_\varepsilon\left(\theta^0 + \beta V(a_t+1,0;\Gamma)\right)} + e^{\eta_\varepsilon\left(p_{t+1} W(a_t+1) - c + \theta^1 + \beta V(a_t+1,1;\Gamma)\right)}$$

and:

$$M^1 = e^{\eta_\varepsilon\left(\theta^0 + \beta V(1,0;\Gamma)\right)} + e^{\eta_\varepsilon\left(p_{t+1} W(1) - c + \theta^1 + \beta V(1,1;\Gamma)\right)},$$

Given parameters $\Gamma = \left\{\beta, \phi_1, \phi_2, \mu_p, \sigma_p, \eta_\varepsilon, \theta^0, \theta^1\right\}$, $V(\cdot)$ can be approximated by a simple iterative recursion. In the first iteration, $V(\cdot)$ in M^0 and M^1 is set to an arbitrary value, and an update of $V(\cdot)$ is found by solving (3.4) for each stand age a (this requires numerical approximation of the integral taken over timber prices). The update is then used on the right-hand side of (3.4) in the second iteration, and so on, until a convergence criterion is met.

For the analyst with observations on the state variables p and a, as well as the actual harvest decision $s_{jt} \in \{0, 1\}$, the probability of the observed harvest decision by forest owner j at time t is:

$$\begin{aligned} &\Pr(s_{jt} | p_t, a_{jt}; \Gamma) = \\ &\frac{(1 - s_{jt})\, e^{\eta_\varepsilon\left(\theta^0 + \beta V(a_{jt},0;\Gamma)\right)} + s_{jt} e^{\eta_\varepsilon\left(p_t W(a_{jt}) - c + \theta^1 + \beta V(a_{jt},1;\Gamma)\right)}}{e^{\eta_\varepsilon(\theta^0 + \beta V(a_{jt},0;\Gamma))} + e^{\eta_\varepsilon(p_t W(a_{jt}) - c + \theta^1 + \beta V(a_{jt},1;\Gamma))}}, \end{aligned} \tag{3.5}$$

and the likelihood function takes the form,

$$L(\Gamma) = \prod_{j=1}^{J} \prod_{t=1}^{T} \Pr\left(s_{jt} | p_t, a_{jt}; \Gamma\right). \tag{3.6}$$

It is important to understand that even though the harvest problem involves four state variables—the observed state variables a and p and the unobserved state

variables ε^0 and ε^1—equation (3.4) is the basis of the "inner" DP algorithm used in estimation, and so the only state variable affecting the dimensionality of the algorithm is stand age (a). For a given set of parameters Γ, $V(\cdot)$ is found by iteration over (3.4), as described above. With $V(\cdot)$ in hand, the likelihood value is calculated from (3.5) and (3.6). The outer search algorithm then chooses an alternative set of parameters, and so on, until convergence.

The most useful measure of forest value is the value conditional on stand age and observed price. From (3.3) and the properties of the Gumbel distribution, the expected value of the forest land for a given stand age and timber price is,

$$\begin{aligned} &E_{\boldsymbol{\varepsilon}} v(a_t, p_t, \boldsymbol{\varepsilon}_t) = \\ &E_{\varepsilon} \max\left[\varepsilon_t^0 + \beta V(a_t, 0; \Gamma),\ p_t W(a_t) - c + \varepsilon_t^1 + \beta V(a_t, 1; \Gamma)\right] = \\ &\frac{1}{\eta_\varepsilon} \ln\left(e^{\eta_\varepsilon\left(\theta^0 + \beta V(a_t, 0; \Gamma)\right)} + e^{\eta_\varepsilon\left(p_t W(a_t) - c + \theta^1 + \beta V(a_t, 1; \Gamma)\right)}\right) + \frac{\gamma}{\eta_\varepsilon}. \end{aligned} \tag{3.7}$$

3.1 Some Numerical Results Illustrating the Model

For a given timber price and stand age, harvest is probabilistic because it depends on the values of $\boldsymbol{\varepsilon}$ generated for the period. Figures 15.2 to 15.4 presents probabilities of harvest for three variations of the modified BM model (that is, the BM model with $\boldsymbol{\varepsilon}$ included). All models consider the case of Loblolly Pine on a low quality site, using parameter values found in Brazee and Mendelsohn (1988). Timber growth is implied by volume parameters ϕ_1 =12.09 and ϕ_2 =52.9 (see (3.1)). Timber prices are normally distributed with mean \$167.4 per thousand board feet, and standard deviation \$40.41. The discount rate is 3% (discount factor = .97). Harvest and replanting costs are \$147, and $\boldsymbol{\theta} = 0$.

The models underlying the panels of Figures 15.2 to 15.4 differ in the value of the scale parameter, η_ε, with this value falling across the three panels. Importantly, the scale parameter of the Gumbel distribution is inversely proportional to the variance (variance= $\pi^2/6\eta_\varepsilon^2$), and so the variance in the distribution of $\boldsymbol{\varepsilon}$ is *rising* across the panels. In the first panel the scale parameter is especially large in the context of the model (η_ε= 20.0). Consequently, values of $\boldsymbol{\varepsilon}$ are invariably close to zero, and so the model is, for all intents and purposes, the same as the original BM model; that is, for a given stand age, the timber is harvested if and only if the timber price is above the reservation price for the stand age. Graphically this is represented in Figure 15.2 by a distinct "probability ledge" tracing the harvest isocline of the original BM model: as the stand ages the reservation price falls (see Figure 15.1). Henceforth we refer to this model as the "virtually no variance" (VNV) model. In Figure 15.3, η_ε =2.0 (the "low variance" (LV) model), and so the reservation price policy no longer applies, as indicated by the transformation of the probability "ledge" of Figure 15.2 to a steep probability "hill" in Figure 15.3. An unexpected result apparent

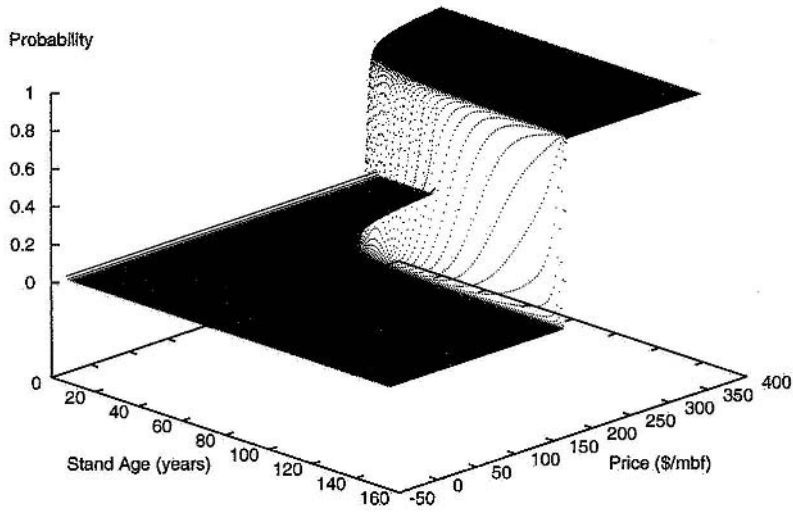


Figure 15.2. Probability of Harvest with $\eta_\varepsilon = 20$.

in the figure is that harvest often occurs at very young stand ages. This is the case because growth at these ages is relatively low, and so if, at these young stand ages, the difference $\varepsilon^1 - \varepsilon^0$ is sufficiently high—a distinct possibility because of the relatively low value of the scale parameter—it is advantageous to harvest the stand and start over. In Figure 15.4, $\eta_\varepsilon = 0.2$ (the "high variance" (HV) model), and the probability surface is noticeably smoother than in Figure 15.3. Essentially the harvest decision is now heavily driven by the observed values of ε^0 and ε^1.

Figures 15.5-15.6 presents simulated data from the VNV and LV models. The data involve 500 timber stands of varying ages observed over an 80-year sequence. Initial stand ages for the sequences are pseudo-random draws from a uniform distribution in the range [1,150]. Prices for the sequences were pseudo-random draws from the price distribution (a single 80-year price sequence applies to all timber stands in the figures). The figures show that the distribution of timber harvests is very different for the two models. The VNV model has relatively few harvests, and these are concentrated in several years. The HV model has many more harvests spread fairly evenly over time, reflecting the very young rotation ages engendered by the relatively high variation

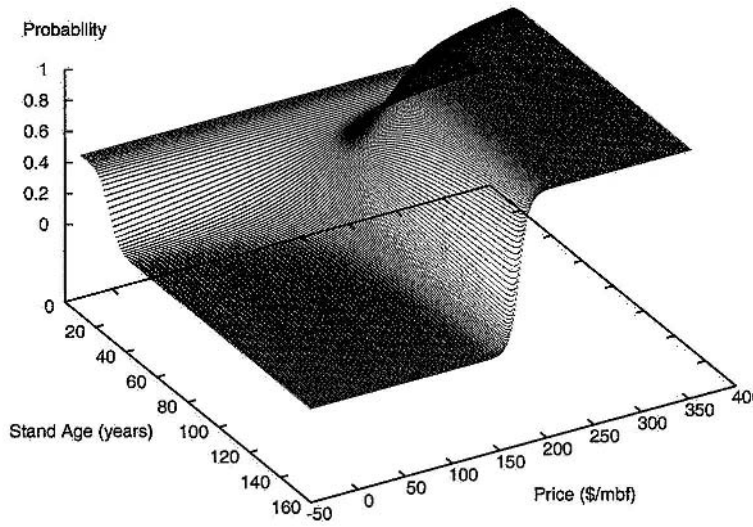


Figure 15.3. Probability of Harvest with $\eta_\varepsilon = 2$.

in ε. Although this is not realistic (generally southern pine must be at least 30 years old to be harvested for sawtimber), the sharp contrast with results from the VNV is useful because it allows us to explore whether qualitative differences in data affect the structural estimation of discrete dynamic decision processes.

Table 15.1 presents estimation results for the simulated data. We present results for both the full simulated samples of the VNV and LV models, and for partial samples in which estimation is based on only the last 40 and 20 years of the sequences.[2] For all seven estimations, the exogenous parameters – the price parameters μ_p and σ_p and the volume parameters ϕ_1 and ϕ_2 – are fixed at their actual values.[3] This leaves four potential parameters to be estimated:

[2]Reducing the sample in the time dimension reflects the judgment that it is easier for researchers to obtain a large cross section of timber harvest data than a large time series.

[3]In actual estimation, the corresponding approach is to estimate the growth function and timber price process exogenously (that is, outside the main estimation algorithm), and to use the estimated parameter values in the main estimation algorithm, thereby significantly reducing the size of the estimation problem. This approach of estimating exogenous processes outside of the main estimation algorithm is common in the literature. Alternatively, the analyst could decide, for instance, that timber owners may be using an incorrect price

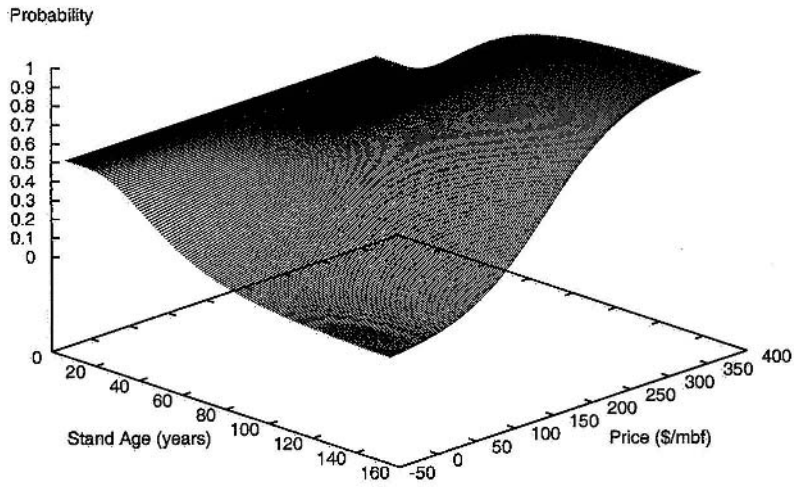


Figure 15.4. Probability of Harvest with $\eta_\varepsilon = 0.2$.

$\beta, \theta^0, \theta^1, \eta_\varepsilon$. But because the values of not harvesting and harvesting are linear in ε^0 and ε^1, respectively, only the *difference* in the location parameters θ^1 and θ^0 can be identified. To see this, observe from (3.3) that the timber owner harvests if

$$p_t W(a_t) - c + \beta\left(V(a_t, 1; \Gamma) - V(a_t, 0; \Gamma)\right) + \varepsilon_t^1 - \varepsilon_t^0 > 0. \tag{3.8}$$

Defining $\Delta\theta = \theta^1 - \theta^0$, it follows from the properties of the Gumbel distribution that the harvest decision can be cast as logistic with location parameter,

$$p_t W(a_t) - c + \beta\left(V(a_t, 1; \Gamma) - V(a_t, 0; \Gamma)\right) + \Delta\theta,$$

and scale parameter η_ε; significantly, only the difference $\Delta\theta$ can be identified. This is reflected in Table 15.1, where we fix θ^0 at zero and estimate θ^1.

When the full sample is used, estimation results are generally excellent for both the VNV and the LV models. For all models except the LV model with

process, in which case the parameters of the price process would be included in the set of parameters to be estimated in the main estimation algorithm. In this case an iterative estimation algorithm suggested by Rust (1994b) would be useful.

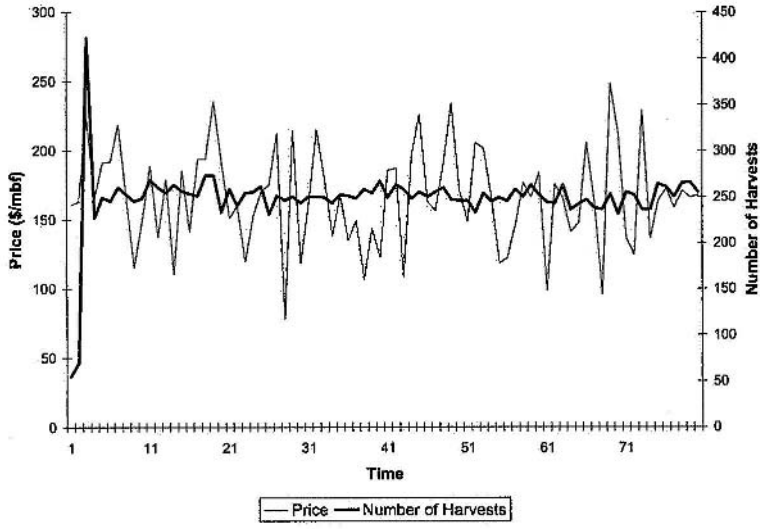


Figure 15.5. Simulated Harvest Data for the VNV model ($\eta_\varepsilon = 20$).

half the sample (Model 5 in Table 15.1), the estimate of the discount rate is within 0.002 of the actual value, and even for Model 5 the true value is within two standard deviations of the estimated value. This no doubt reflects the tremendous influence of the discount rate on the harvest decision. By comparison, in estimations using partial samples the estimates of η_ε and θ^1 tend to be *statistically* different than the true values of these parameters. Are these differences significant as a *practical* matter? Yes and no. In an investigation of timber harvest behavior, the analyst is primarily interested in two questions: the effect of stand age and timber price on the likelihood of harvest, and the expected value of bare forestland (In particular, the nontimber value of forestland). On both counts, Models 2, 3, and 6 generate results very similar to the true model. The exception to these generally favorable estimation results is Model 5, for which the estimated discount factor is considerably lower than the true discount factor. This lower discount factor (higher discount rate) has two significant effects. First, the estimated value of bare land is much higher for Model 5 than the actual value: $9,282 per acre versus an actual value of

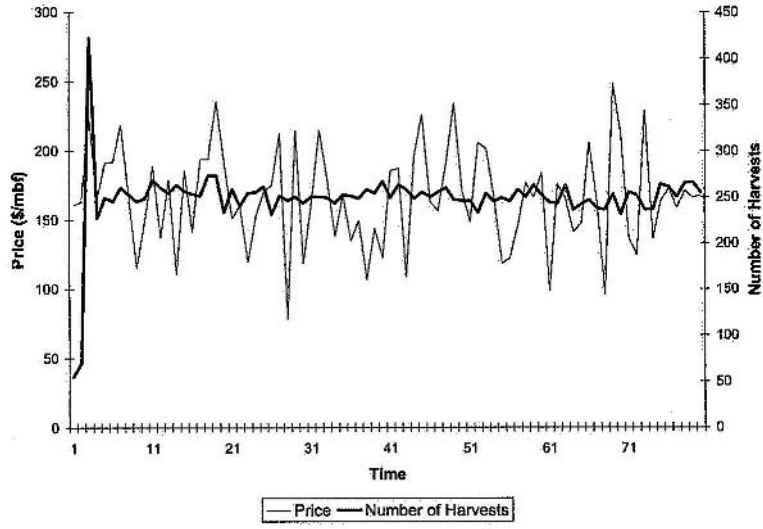


Figure 15.6. Simulated Harvest Data for the LV Model ($\eta_\varepsilon = 2$).

$1,456 per acre. Of course, the large standard error on the estimate of the discount factor in Model 5 signals the analyst that the estimate of bare land value is imprecise. Second, for timber stands that by chance mature past age 20 or so, the predicted harvest age is lower for Model 5 than for the true model.[4] Figure 15.7 tells the story on this. It presents the *difference* in harvest probabilities between the true model and the estimated Model 5. The deep trough in Figure 15.7 beginning at about stand age 25 in the price range of roughly $200-$270, indicates that for older stands Model 5 overpredicts the probability of harvest, and so underpredicts the expected harvest age. This result is entirely consistent with the usual literature on optimal timber harvesting that indicates the

[4]It is worth emphasizing that in both the true model and Model 5, the odds of a timber stand reaching age 10 is extremely low. This is apparent from Figure 15.3, which shows that for η_ε=2.0, the probability of harvest in each of the first ten years is roughly 0.4, and Figure 15.7, which shows that in the first ten years the probability of harvest in Model 5 is virtually the same as in the true model.

Table 15.1. Estimation Results from Simulated Data (see text)

Parameter	*Model 1* VNV	*Model 2* VNV	*Model 3* VNV	*Model 4* LV	*Model 5* LV	*Model 6* LV
[true value]	$\{1...80\}^{\dagger}$	$\{40...80\}$	$\{60...80\}$	$\{1...80\}$	$\{40...80\}$	$\{60...80\}$
β	0.9708	0.9703	0.9686	0.9697	0.9313	0.9702
[0.97]	(0.003)	(0.002)	(0.002)	(0.004)	(0.037)	(0.050)
η_ϵ	18.43	17.58	61.22	-	-	-
[20.0]	(1.12)	(0.51)	(2.46)			
η_ϵ	-	-	-	2.07	5.25	4.36
[2.0]				(0.14)	(0.21)	(0.4)
θ^1	–0.014	–0.051	0.124	0.016	0.094	0.081
[0.0]	(0.019)	(0.083)	(0.013)	(0.010)	(0.002)	(0.007)
Log Likelihood	–357	–226	–21	–26701	–14007	–7164

VNV="Virtually No Variance" model, with $\eta_\epsilon = 20.0$; LV="Low Variance" model, with $\eta_\epsilon = 2.0$.
† $\{x, y\}$ indicates that only sample observations between and including years x and y in the 80-year sequence of simulated prices are included in the estimation. Standard errors are given in parentheses. Additional explanation about this Table is found in the text. Starting values for all estimations: $\beta = 0.95$, $\eta_\epsilon = 10.0, \theta^1 = 0.20$.

expected harvest age falls with an increase in the discount rate (decrease in the discount factor).

It should be emphasized that the low estimated discount rate in Model 5 is not a fluke associated with the particular starting values used in estimation. In a search across a range of starting values, including the parameters of the true model, we could not find estimates generating a higher likelihood value for the sample than those presented in Model 5. It should also be emphasized that the deep probability trough observed in Figure 15.7 occurs *outside* the range of the simulated data. Of the 8,621 harvests observed in the data, 8,561 (99.3%) occur at stand age 10 or less, and all harvests occur at a stand age less than 20. In other words, for very few observations is the stand age greater than 10, and for no observations is it greater than 20. Perusal of Figure 15.7 shows, then, that *in the range of the data* the harvest probability of Model 5 is virtually the same as for the true model. This teaches an old and familiar lesson, one certainly not unique to the estimation of DDP's: in the absence of a good range in the data, identification of a model can be difficult.

4. Comments on Reduced Form (Static) Estimation of Discrete Dynamic Decision Problems

Given the difficulty of structural estimation of a discrete DDP, it is reasonable to question whether the effort is worth the gain. In particular, why not

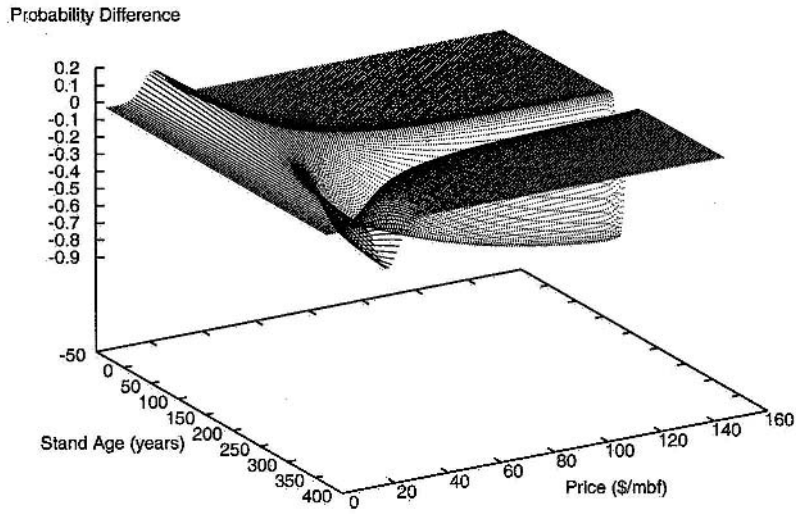


Figure 15.7. Harvest Probability Difference between the True Model and Model 5 ($\Pr_{True} - \Pr_{Model5}$).

simply estimate a reduced-form version of the problem? Inspection of the timber harvesting decision (3.8) makes clear that one can specify the decision problem as one in which the forest owner harvests trees if:

$$F\left(p_t, a_t\right) + \Delta\varepsilon_t > 0, \tag{4.1}$$

where $\Delta\varepsilon_t$ is distributed logistically. This leads to a straightforward application of logistic maximum likelihood estimation. A similar reduced form can be used to approximate the optimal decision rule of *any* discrete DDP. Of course, the presence of the value functions in (3.8) argues for a flexible form in the approximation, and even then the quality of the approximation may be poor. Figures 15.8-15.10 presents harvest probabilities from first-, second-, and third-order estimation of (4.1) for the case where the true model is the LV model ($\sigma_\varepsilon = 2.0$), and so the figure against which the panels of Figures 15.8-15.10 are to be judged is Figure 15.3. Estimates are based on simulated data of the same size as used in Table 15.1 – namely, pooled time-series, cross-sectional data of length 80 years and width 500 forest stands. The structural counterpart in estimation is Model 4 (see Table 15.1), which generated harvest

probabilities virtually identical to the true harvest probabilities (that is, it generates a probability surface that looks exactly like the probability surface of the true model presented in Figure 15.3, so we do not bother presenting the surface here). A comparison of Figure 15.8 and Figure 15.3 indicates that when $F(\cdot)$ takes the simple linear form,

$$F = \alpha_0 + \alpha_1 p + \alpha_2 a,$$

the approximation to the harvest probability surface is very poor, and the probability of harvest *decreases* as the stand age increases. Figure 15.9 indicates that when $F(\cdot)$ takes the quadratic form,

$$F = \alpha_0 + \alpha_1 p + \alpha_2 a + \alpha_3 p^2 + \alpha_4 a^2 + \alpha_5 p \cdot a$$

the approximation is considerably improved, and Figure 15.10 indicates that when $F(\cdot)$ takes the cubic "form,"

$$\begin{aligned} F = \quad & \alpha_0 + \alpha_1 p + \alpha_2 a + \alpha_3 p^2 + \alpha_4 a^2 + \alpha_5 p \cdot a \\ & + \alpha_6 p^3 + \alpha_7 a^3 + \alpha_8 p^2 \cdot a + \alpha_9 p \cdot a^2, \end{aligned}$$

the approximation is excellent. Seemingly and not surprisingly, even a fairly low-order polynomial will do a good job of approximating fairly complex decision rules.

The conceptual weakness of reduced form estimation is the same for DDPs as it is for any economic model; it does a reasonably good job of describing the effect of various state variables on decision variables, but the estimation is otherwise devoid of economic content. It identifies the variables affecting a dynamic decision, but it provides no insight about the mechanism of the relationship. In the illustrative example presented here, reduced form estimation tells the analyst that timber price and stand age do affect the harvest decision, but it is silent about the discount rate used in the harvest decision, an important issue in the long debate about whether forest owners harvest timber too soon (and therefore require incentives to take the long view); it says nothing about the nontimber benefits of forestland (as embodied in ε), a major issue in the allocation of land across various uses; it says nothing about the overall value of forestland; and it says nothing about the price expectations of forest owners.[5]

A related issue is whether it is safe to assume that a decision problem is static. For our timber example, one might simply posit that in each period the forest owner harvests if the intra-period utility from harvesting is greater than

[5]In the structural models estimated in this chapter, price distribution parameters are fixed at their true values. In other estimations conducted in preparation of the chapter but not presented here, price distribution parameters were included in the set of parameters to be estimated. Generally results were excellent, though a bit sensitive to starting values. Provencher (1995a,1995b) estimates models of pulpwood harvesting in which the price process is autoregressive.

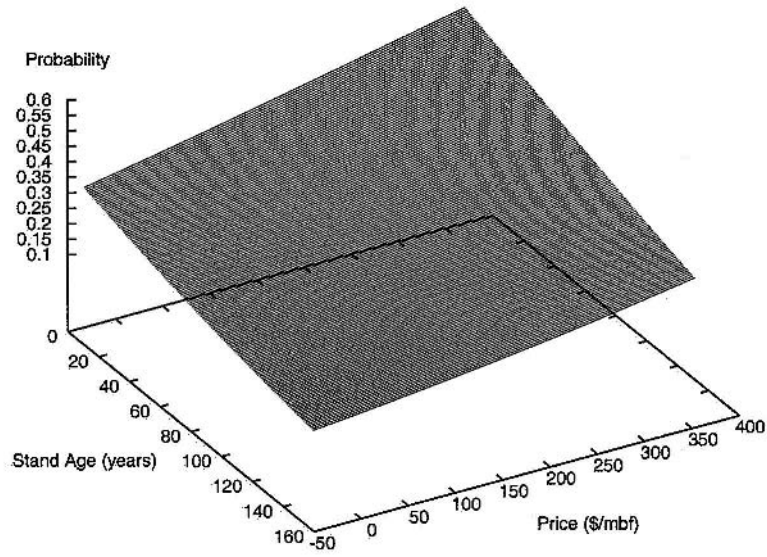


Figure 15.8. Harvest Probability Surface for Linear Reduced-Form (Logit) Model.

the utility of not harvesting. For instance, letting $H(a_t)$ denote the money-metric utility derived from a standing forest at age a_t, and assuming that harvest occurs at the beginning of the period and $H(0) = 0$, harvest occurs if

$$\begin{aligned} & p_t W(a_t) - c + \varepsilon_t^1 > H(a_t) + \varepsilon_t^0 \\ & \Rightarrow p_t W(a_t) - c - H(a_t) > \Delta\varepsilon_t, \end{aligned} \tag{4.2}$$

which, under the same assumptions about the distribution of ε used above, is the basis of a simple logistic regression. However, this approach can be problematic. For many problems—timber harvesting surely among them—it is not reasonable to assume away dynamic behavior. For problems where dynamic behavior is an open question, things are a bit complicated. Baerenklau and Provencher (2004) examine the issue of whether recreational anglers allocate a "fishing budget" over the course of a season. We estimate both a structural dynamic model and a reduced form static model of trip taking behavior and find significantly different welfare estimates across models. More troubling, we also demonstrate theoretical inconsistencies and identification problems with the static model when behavior is truly dynamic.

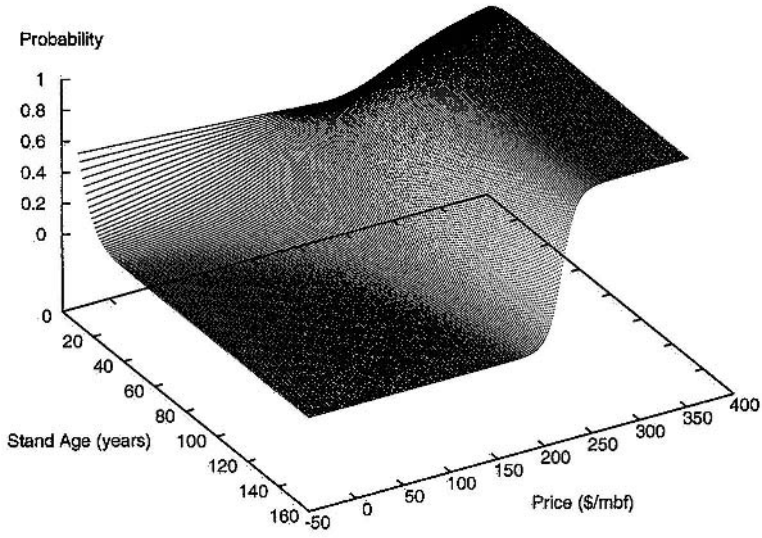


Figure 15.9. Harvest Probability Surface for Quadratic Reduced-Form (Logit) Model.

Furthermore, although dynamic models encompass static ones, actually testing for static behavior is problematical for the simple reason that if one specifies a sufficiently flexible form for the static utility function, the static model will provide an excellent fit to the data. Put another way, a reduced form model can be relabeled as a structural, albeit static, model, and as already demonstrated, reduced-form models can provide an excellent fit to the data. In our timber example, a sufficiently flexible static form requires making $H(\cdot)$ a function of p as well as a, and this is difficult to justify. Such structural restrictions are ultimately necessary to test static versus dynamic models.

For many problems there is little *a priori* knowledge about the intraperiod benefit (utility, profit) function, and so it is difficult to distinguish static from dynamic behavior. For such problems, out-of-sample forecasting may shed light on whether behavior is static or dynamic, but this is a quite expensive diagnostic. Baerenklau and Provencher (2004) conduct such a test and find that in general a dynamic model does a better job of predicting out-of-sample trip-taking behavior than does a static model.

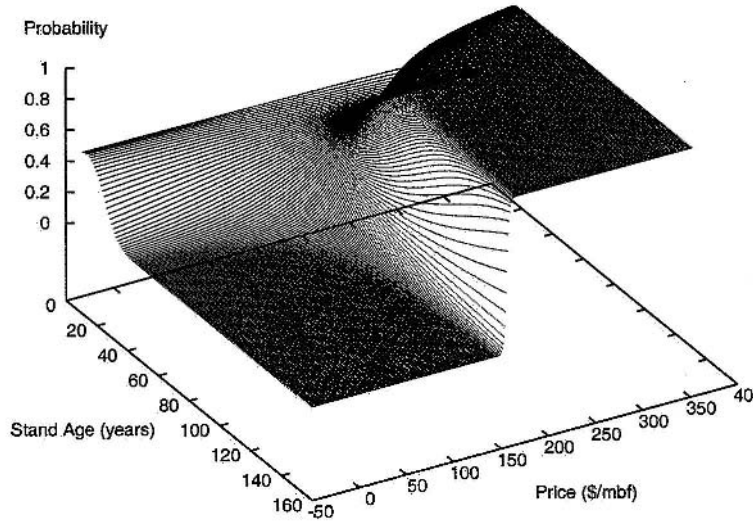


Figure 15.10. Harvest Probability Surface for Cubic Reduced-Form (Logit) Model.

5. The Future of Structural Estimation of Dynamic Decision Processes

For years, structural estimation of DDP's simply was not practical because of the substantial barrier posed by the computational requirements of estimation. Rapid advances in computational speed in the mid-1990s reduced this barrier considerably, yet the literature estimating DDP's is decidedly modest, and there is no evidence that it growing. This is partly because the econometric modeling is difficult to understand and implement; this chapter attempts to clarify the estimation method. Perhaps it also reflects two related objections.

First, estimation requires strong parametric assumptions about the decision problem generating the data. In the example considered in this chapter, we maintained the strong assumption that the forest owner knows the growth function of trees and, perhaps even more unlikely, that the forest owner knows the stochastic process generating timber prices. In the real world, such assumptions are complicated by the requirement that the analyst *correctly* specify the parametric form of various dynamic processes, where by the "correct" process we mean the process actually used by the decision-makers.

Second, the estimation maintains that agents are dynamic optimizers. In a working paper entitled, "Do People Behave According to Bellman's Principle of Optimality?", Rust (1994a) makes the important point that *any* data set of state variables and decision variables can be rationalized as the outcome of dynamically optimizing behavior. The issue becomes, then, whether the data generated can be rationalized by dynamically optimal behavior circumscribed by plausible parametric specifications of the intra-period benefit function and dynamic processes. The issue returns, in other words, to a variation of the first objection: Is the parametric specification of the dynamic model plausible in some theoretical or practical sense?

Given that a flexible static model can fit a data set as well as a dynamic one, and that dynamic optimization *per se* does not imply testable restrictions (and thus is not refutable), how might an analyst who suspects that behavior is dynamic proceed? In our view, estimation of DDP's should be informed by agent self-reporting about expectations of future states and the relative importance of the future in current decisions. So, for instance, if agents respond in surveys that future states are unimportant to their current decisions, it would seem difficult to justify a dynamic model. If agents report that they believe the best guess of prices tomorrow is the current price, a random walk model of price expectations (or at least, a specification of prices that allows a test for a random walk) is warranted.

Yet even as economists are uneasy about the strong parametric assumptions to be made in the estimation of DDP's, there exists a longstanding uneasiness in the profession about using agent self-reports to aid in the estimation of dynamic behavior. Manski (2004) argues that with regards to agent expectations, this is not much justified. The author states,

> Economists have long been hostile to subjective data. Caution is prudent, but hostility is not warranted. The empirical evidence cited in this article shows that, by and large, persons respond informatively to questions eliciting probabilistic expectations for personally significant events. We have learned enough for me to recommend, with some confidence, that economists should abandon their antipathy to measurement of expectations. The unattractive alternative to measurement is to make unsubstantiated assumptions (pg. 42).

Although the role of self-reports in an analysis can be informal, giving the analyst a rough basis on which to choose a static vs. a dynamic model, to choose one parametric specification of dynamic processes over another, and so forth, a useful direction of future research is to incorporate survey data in the estimation problem itself, similar in principle to the way revealed- and stated-choice data are integrated in the estimation of random utility models. We anticipate that in the next ten years insights gained from experimental psychology, experimental economics, and survey research will provide the basis for richer and more accurate models of dynamic economic behavior.

Chapter 16

MONTE CARLO METHODS IN ENVIRONMENTAL ECONOMICS

Giovanni Baiocchi
Department of Economics and Finance
University of Durham, United Kingdom
giovanni.baiocchi@durham.ac.uk

Abstract The role of Monte Carlo methods in environmental economics and its subdisciplines, has increased in importance during the past several years. Due to the increasing power of computers and the development of sophisticated software, Monte Carlo and other computer-based simulation methods have emerged and established themselves as a third approach for advancing environmental and resource economics along side with traditional theory and empirical research. In this paper we review the contribution in environmental and resource economics of Monte Carlo method applications, illustrate guidelines for Monte Carlo results to be effectively and accurately communicated to and independently reproduced by other researchers, and survey the main methods and software options for executing Monte Carlo experiments.

Keywords: Environmental and Resource Economics, Monte Carlo, pseudo-random numbers.

1. Introduction to Monte Carlo Methods

At the turn of the last century, Mr. Student (pseudonym of W. S. Gossett), after ascertaining the exact sampling distribution of the ratio between the estimated sampling error of the mean and its standard error, conducted an experiment to test his theory. The laborious procedure is described in Student (1908):

> The measurements were written out on 3,000 pieces of cardboard, which were then very thoroughly shuffled and drawn at random. As each card was drawn its numbers were written down in a book which thus contains the mea-

> surements of 3,000 criminals in a random order. Finally each consecutive set of 4 was taken as a sample —750 in all— and the mean, standard deviation, and correlation of each sample determined. The difference between the mean of each sample and the mean of the population was then divided by the standard deviation of the sample, ...

Today, using a computer, an analogous experiment can be repeated with considerable less effort from the part of the experimenter. The following, almost *self-explanatory*, "one liner" written in R code[1] can be used to perform the experiment on most computers in merely a fraction of a second.[2]

```
ts <- {}
for (i in 1:750) { x<-rnorm(4); ts[i]<-sqrt(4)*mean(x)/sd(x) }
```

Figure 16.1 uses a histogram to graph the computed results, which represents the sampling distribution of the statistic. This is an example of early use of what we today call a "Monte Carlo" simulation.[3] A Monte Carlo simulation is basically a statistical experiment performed using a computer. The name Monte Carlo for this kind of experiments appeared first in Metropolis and Ulam (1949). Ulam (1977) recounts that the method was named in honor of his uncle, who was an assiduous frequenter of the Monte Carlo world-famous gambling casino, on the French Riviera. Treating Monte Carlo methods clearly as experiments, highlights the fact that Monte Carlo investigations should be treated as seriously as any other scientific experiment. This has two important implications:

1 Monte Carlo simulations should be designed carefully, and

2 enough details should be provided for the simulations to be independently reproduced.

Good design is a desirable feature of the simulations as it is the key to reducing the cost of a project. Instances of Monte Carlo applications in Environmental Economics conceive to help the conduction of empirical studies abound.

---

[1] R is an open-source implementation of the S language available from the WWW's Comprehensive R Archive Network (CRAN) located at *http://cran.r-project.org/* where source code, additional libraries, documentation, and links to binaries distributions of R are available for various platforms, including Win32, Mac, and Unix/Linux. (see, e.g., Ihaka and Gentleman, 1996).

[2] The hardware used in this Chapter was a Dual Intel Pentium IV (Prestonia) Xeon Processors 3.06 GHz with HT Technology with 4 GB of RAM running on Microsoft Windows XP/2002 Professional (Win32 x86) 5.01.2600 (Service Pack 2). We used R release 2.0.0, the standard Win32 release available at the time of writing the present paper. The default normal random number generator in R implements the inverse method. The underlying default uniform random number generator is the *Mersenne Twister* developed by Matsumoto and Nishimura (1998), a modification Matsumoto's TGFSR generator of Matsumoto and Kurita (1992, 1994). This generator has a huge Mersenne prime period length of $2^{19937} - 1 \approx 10^{6000}$ and its output is "twisted" to free it of long-term correlations when considered from a viewpoint of 623 dimensions.

[3] Some authors (see, e.g., Ripley, 1987) prefer to use the term stochastic simulation.

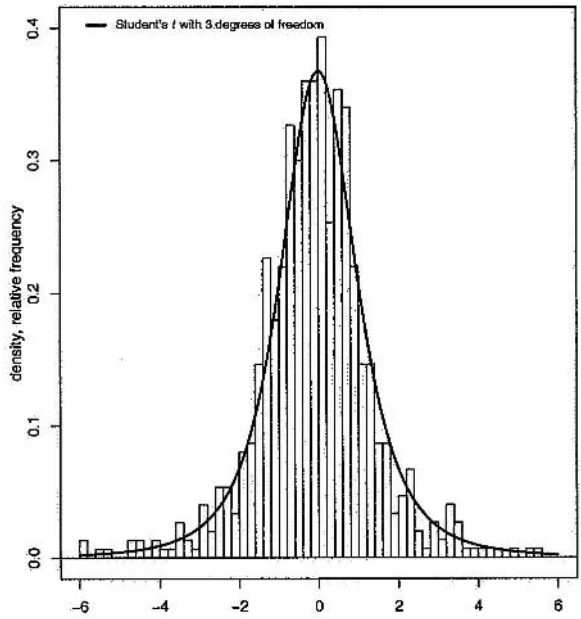


Figure 16.1. Histogram of 750 statistics with Student's t distribution.

Reproducibility is an important tenet of experimental sciences. The reproducibility of Monte Carlo experiments is regarded as particularly crucial. In fact, it can be argued that Monte Carlo experiment should be *strictly* reproducible. Reproducibility implies that, ideally, identical results should be obtainable in a short amount of time, without requiring expensive computational resources, proprietary data, licensed software, and any application-specific knowledge. Moreover, for reproducibility to be of practical use, code and data should be carefully organized and documented.

In this chapter we briefly review some of the contribution to environmental and resource economics of Monte Carlo applications, illustrate guidelines for Monte Carlo results to be effectively and accurately communicated to and independently reproduced by other researchers and students, and survey the main methods and software options for executing Monte Carlo experiments. The chapter is organized as follows. In Section 2 we review a few areas in which Monte Carlo methods have contributed to the advancement of environmental economics. Section 3 looks at some issues in designing Monte Carlo studies. In Section 4 we highlight the importance of reproducible computational results in environmental economics. Section 5 illustrates guidelines for Monte Carlo results to be effectively and accurately communicated to and in-

dependently reproduced by other researchers and students. Section 6 looks at the issue of random number generation. Section 7 discusses useful criteria for the selection of a suitable random number generator for a Monte Carlo study. In Section 9, random number generators provided by commonly used software application are reviewed and tested. Section 10 concludes.

2. Monte Carlo Methods in Environmental Economics

The role of Monte Carlo methods in environmental economics and its subdisciplines, has increased in importance during the past several years. A quick search through environmental economic journal databases reveals that, though most applications of Monte Carlo methods are recent, they are growing in number. The choice between theoretical and simulation models is rapidly tilting toward simulation methods as computation becomes ever cheaper. This trend should become more evident in an applied discipline such as environmental economics as simulations are particularly well suited at answering "what if" questions whereas the standard analysis approach is more suited to further understanding of the model. Monte Carlo methods comprise a wide set of techniques and have a wide range of possible applications. They are commonly employed to:[4]

- solve problems that are inherently complex (the so called "NP complete" problems),
- optimize (simulated annealing, genetic algorithms),
- simulate complex systems,
- establish finite sample properties of estimators,
- estimate complex models (Simulation-based methods, Bayesian methods),
- compare competitive estimators, algorithms, etc.,
- test hypotheses (non-parametric tests, bootstrap, permutation tests),

A typical econometric-oriented application of Monte Carlo methods in Environmental Economics that has generated a sizable strand of literature, focusses on various aspect of the distributions of estimators. For instance, in the CV literature, to construct confidence intervals and standard deviations for welfare measures, Monte Carlo methods have often been preferred over first order approximation methods (delta method). For instance, Park, Loomis,

[4]This classification is not meant to be exhaustive or mutually exclusive. In fact, many applications of Monte Carlo could fit in most categories simultaneously.

and Creel (1991) employed the Monte Carlo method proposed by Krinsky and Robb (1986) to construct confidence intervals for benefits from CV dichotomous choice studies; Duffield and Patterson (1991) used the Boostrap method in an analogous study. Kling (1991) compares the Monte Carlo, the bootstrap, and the first order approximation methods to estimate confidence intervals and standard deviations of welfare measures using two Monte Carlo experiments

Bias resulting from incorrect model specification has been investigated using Monte Carlo simulations. For instance, Herriges and Kling (1997) supplement analytical tools with Monte Carlo results to examine bias resulting from incorrect model specification of welfare estimates derived from nested logit model estimates (see also, e.g., Kling and Thompson, 1996, Huang and Smith, 2002, and reference therein). Alberini (1995) investigates the robustness of WTP estimators against deviations from underlying model assumptions. Among others, Cooper and Hanemann (1995) and Scarpa and Bateman (2000) compare welfare estimate efficiency of various CV bid designs. Kling (1997) looks at efficiency gains of combining revealed preference and stated preference methods.

Another important area of application of Monte Carlo method is the estimation of more complex and realistic models. Allowing for more plausible behavioral assumptions, such as non-independence between alternatives, comes at the price of additional computational complexity. Simulation methods that enabled researchers to estimate model such as the multinomial probit and the mixed logit were introduce and developed by McFadden (1989), Boersch-Supan and Hajvassiliou (1990), McFadden and Ruud (1994) to name just a few. Among the earliest applications of simulation-based mixed logit models to environmental issues we can include Revelt and Train (1998) and McFadden and Train (1996, 2000).

Another approach to the estimation of more complex models, is based on the Markov Chain Monte Carlo (MCMC) method. Though the history of MCMC is as long as the history of Monte Carlo method itself, the method was in fact developed during the second world war and was published later by Metropolis *et al.* (1995), application in environmental economics are still rare.

An early application of the MCMC method to the valuation of natural resources is given by Hausman, Leonard, and McFadden (1995). In the paper the author develop an algorithm for computing welfare impacts with random utility models using a Monte Carlo Markov chain simulator for generalized extreme-value variates. As Layton and Levine in chapter 10 point out, Bayesian econometric approaches to modeling non-market valuation data have not often been applied. Among others, León and León (2003) modeled double bounded contingent valuation data, León *et al.* (2002) considered benefits transfer from a Bayesian perspective, Layton and Levine (2003) modeled stated preference data with complex response patterns, and Train (2003) and Train and Sonnier

in chapter 7 illustrate a Bayesian approach in the estimation of mixed-logit models.

As an application of MCMC to an inherently complex problem, Fukami (2004) used a Monte Carlo approach to compute the degree of linearity of a input-output table in physical units. This measure, which provides an information on how materials are recycled in an economy, is computationally hard to obtain. Optimization problems are also routinely solved in environmental and in resource economics using MCMC methods. For instance, simulated annealing has been used in forest economics, fisheries, etc. (see, e.g., Boston and Bettinger, 1999, and references therein). Meilby, Strange, and Thorsen (1999) proposed an application to other renewable resources.

Based on this, necessarily short, review on Monte Carlo areas of application in environmental economics, we conclude that these methods are having and will continue to have a considerable impact on the discipline. The potential for applications is enormous. Due to the increasing power of computers and the development of sophisticated software, Monte Carlo and other computer-based simulation methods have emerged and established themselves as a third approach for advancing environmental economics along side with traditional theory and empirical applications. It is worth noting that the greatest development appear to be made in the development and application of Monte Carlo Markov chains and quasi-random methods: the first method generates samples that are neither independently nor identically distributed, the latter appear clearly non-random, as they attempt to span systematically the outcome space. Moreover, traditional econometric based Monte Carlo methods tend to use a minimum amount of data and a large number of computer (floating-point) operations. Monte Carlo methods used to fit more realistic models are both computation and data intensive.

3. Designing A Monte Carlo Study

With the possible exception of very small "quick-and-dirty" experiments, a Monte Carlo experiment should be carefully designed. Monte Carlo studies are ordinarily straightforward to implement. A successful Monte Carlo application in environmental and resource economics requires knowledge of:

1 the basic principles of simulation, and

2 the structure of the problem at hand.

Knowledge of the problem at hand should be used to make sure that the study is relevant to answer the researcher's original question. The quality and usefulness of the Monte Carlo study critically depends on a good understanding of the specific theoretical and empirical issues underpinning the environmental issues under scrutiny. A review and critical evaluation of the exist-

ing literature on the topic being researched is probably the best way to start a Monte Carlo study.

Also, knowledge of simulation principles can help to economize the use of the researcher's time. As resources are limited, it is usually feasible to perform only a relatively small number of experiments. On the one hand, we need to consider practical issues such as the experimenter's own time and the available computational resources, and on the other, it is imperative to obtain results that are sufficiently accurate for the purpose at hand. An important problem is to determine the number of random draws needed. Most methods to derive the necessary number of random draws to achieve some predetermined level of accuracy are based on two fundamental results in mathematical statistics from the theory of convergence of sequences of random variables:

1 the (weak) **law of large numbers**, which states that given a sequence $X_1, \ldots, X_n$ of i.i.d. random variables, with mean $E(X_i) = \mu$ and variance $\sigma^2 < \infty$, then the sample average $\overline{X} = \frac{1}{n}\sum_1^n X_i$ converges (in probability) to μ, and

2 the (classical) **central limit theorem**, which states that, under the same conditions, $\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}$ converges (in distribution) to $N(0,1)$ (for more details and extensions, see Feller, 1968, and references therein).

The law of large numbers ensures that if the sample is large enough, the sample average will be near to the population mean. The central limit theorem allows us to quantify the probability of the discrepancy, and therefore the sample size needed to achieve a certain accuracy. As an application of the above mentioned results, consider the problem of determining the size of a test statistics (i.e., the probability of committing a type I error). Following Davidson and MacKinnon (1993, p. 739), let's denote with p the quantity to be estimated. The sample of replications can be represented by n random variables, $X_1, \ldots, X_n$, where $X_i = 1$ if in replication i the test statistics exceeds the nominal critical value, and $X_i = 0$ if it does not exceed it. Each replication can be thought as an independent Bernoulli trial with probability of "success" fixed to p across trials. The number of rejections $S_n = X_1 + X_2 + \cdots + X_n$ is therefore approximately binomially distributed with parameters n and p. An estimator for p is provided by $\bar{p} = S_n/n$ which is the proportion of rejections for n replications. The mean and the variance of the estimator are p and $n^{-1}p\,(1-p)$, respectively.

Assume that it is desired that the estimator, $\bar{p}$ should differ from the population mean p, by a small quantity no larger than ε, with confidence level $1-\alpha$, $0 < \alpha < 1$, or greater, i.e,

$$P\left(|\bar{p} - p| \leq \varepsilon\right) \geq 1 - \alpha.$$

Standardizing $\bar{p}$ the problem becomes,

$$P\left(\left|\frac{(\bar{p}-p)}{\sqrt{p\,(1-p)/n}}\right| \leq \varepsilon\frac{\sqrt{n}}{\sqrt{p\,(1-p)}}\right) \geq 1-\alpha.$$

Since the distribution of the standardized $\bar{p}$, by the central limit theorem, can be approximated by the standard normal,

$$P\left(|Z| \leq \varepsilon\frac{\sqrt{n}}{\sqrt{p\,(1-p)}}\right) = 2P\left(Z \leq \varepsilon\frac{\sqrt{n}}{\sqrt{p\,(1-p)}}\right) - 1 \geq 1-\alpha.$$

or

$$P\left(Z \leq \varepsilon\frac{\sqrt{n}}{\sqrt{p\,(1-p)}}\right) \geq 1-\alpha/2.$$

so that the required sample size is

$$n \geq \left(\frac{\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{p(1-p)}}{\varepsilon}\right)^2.$$

Assuming $\alpha = .01$, the .995-quantile is $z_{.995} \approx 2.576$. For $\varepsilon = \frac{1}{100}$, and using the fact that $p\,(1-p) \leq 1/4$ for every p, a very safe number of replications would be $n \approx 16,590$. For a more plausible $p = .05$, the sample size reduces to $n \approx 3,152$. It is worth bearing in mind that these convergence results are based on large numbers ($n \to \infty$) and i.i.d. sequences. Both conditions do not hold for computer generated random sequences.

More robust methods are available but with a higher computational burden. The quality of the approximation will depend on the rate at which the distribution of the standardized estimator converges to the normal. In general, as the rate of convergence depends on the application, the normal approximation has to be regarded as an additional source of error. A worse-case scenario sample size can be derived using Chebyshev's inequality which states that for a random variable X assumed to have a distribution with finite variance, $\sigma^2 < \infty$, and mean, $\mu = 0$, for every positive k,

$$P\left(\frac{|X-\mu|}{\sigma} \geq \varepsilon\right) \leq \frac{1}{k^2}.$$

In our example, $X - \mu = \bar{p} - p$ and $\sigma = n^{-1}p\,(1-p)$, so that

$$P\left(|\bar{p}-p| < \varepsilon\sigma\right) \geq 1 - \frac{p\,(1-p)}{n\varepsilon^2}.$$

Given a confidence level $1-\alpha$, the sample size

$$n \geq \frac{p\,(1-p)}{\alpha\varepsilon^2},$$

will satisfy the required error specification. Again, since $p\,(1-p) \leq 1/4$ the worst-case scenario sample size becomes

$$n \geq \frac{1}{4\alpha\varepsilon^2},$$

for all $p \in [0,1]$. For the choice of parameters in our example $\varepsilon = 1/100$, and $\alpha = .01$, the worst-case scenario sample size is $n \geq 40,000$.

Sometimes, with particularly computer-intensive applications, instead of asking how many observations are needed to achieve a desired level of accuracy, the researcher will be constrained to ask the, methodologically less satisfactory, "inverse" question, i.e., how accurate are the results given the number of observations used in the Monte Carlo study. The problem of feasibility becomes central in this case.

4. Reproducibility of Monte Carlo Results

Claerbout (see, e.g., Buckheit and Donoho, 1995), has recently championed the issue of reproducibility in the computational sciences. Reproducing computation results from published work often proofs to be a difficult and daunting task. Reproducibility relies on a plethora of implementation details that are difficult to communicate through conventional printed publications. Buckheit and Donoho (1995) point out that in the field of computational experiments:

- researchers often cannot reproduce their own work, even a few months after the study has been completed,
- research students have difficulties in presenting their problems to their academic advisers, and
- researchers cannot reproduce computational results of other researchers and other published work.

Reproducibility implies that, ideally, identical results should be obtainable in a short amount of time, without requiring expensive computational resources, proprietary data, licensed software, and any application-specific knowledge. Moreover, for reproducibility to be of practical use, code and data should be carefully organized and documented.

Schwab *et al.* (2003) classify their computational problems according to their degree of reproducibility in:

- **Easily reproducible** result files can be regenerated within ten minutes on a standard workstation.
- **Non-reproducible** result files, such as hand-drawn illustrations or scanned figures, cannot be recalculated by the reader.

- **Conditionally reproducible** result files require proprietary data, licensed software, or more than 10 minutes for their re-computation. The author nevertheless supplies a complete set of source files and rules to ensure that readers can reproduce the results if they possess the necessary resources.

Based on this stringent requirements, most computational results in economics and in environmental economics would be classified under the headings of "conditionally reproducible" at best. In a recent investigation, Vinod (2001) found that approximately 70 per cent of articles from prestigious economic journals were not reproducible. He attributed this problem to sloppy record keeping, inaccurate software, and the lack of maintenance of software and data, in particular, after publication.

Environmental Economics does not fare much better in this respect. In my experience, obtaining a datasets and the software code needed to reproduce published work proofs at best to be a difficult task. Recently two attempts from my part of obtaining data and code for the purpose of reproducing published computational results, have failed. The authors blamed a computer virus and a computer crash to explain the loss of data and code. In one instance, I was able to obtain the data used for a paper on a leading journal in environmental economics, but was unable to exactly reproduce the computational results. Similar instances were experienced by the editors of this volume (personal communication), and indeed seem to be frequent in the profession.

Of course, insisting on exact and easily reproducible results is not always practical. In applied work, it is quite frequent that a particular commercial software, dataset, or expensive equipment makes research results difficult to reproduce.

A particularly frequent problem is that in many environmental economic applications data used in published work is considered confidential and not made available. Researchers might, in fact, collect data themselves at a considerable cost, pay other institutions to analyze the data, etc. Other times the data is provided, but it is in a format that makes it difficult to use or is insufficiently documented.

Environmental economics journal do not maintain databases of data and codes of published papers. Typically, a much milder policy is implemented. For instance, consider the Journal of Environmental Economics and Management's policy for replication as stated in their "Guide for Authors." According to the current policy, all data must be clearly documented and computational methods must be explained with sufficient details to enable replication by other researchers. The only requirement concerning the dataset is that it must be made available on request. The findings of Dewald, Thursby, and Anderson (1986) suggests that this type of policy is not adequate to guarantee reproducibility of computer-based results. In their Journal of Money, Credit and

Banking project, they attempted to replicate computation results published or submitted to the journal. Of the 92 authors asked to supply data according to the journal policy, 75 responded, and 68 submitted something. The first 35 datasets were examined and only 7 were judged to be free of problems. The authors attempted to replicate the results of 9 papers for which they had obtained data and software code; only four computational results could be reproduced closely. Based on their findings, Dewald *et al.* (1986) recommended that journals require the submission of data and programs from authors at the time empirical papers are submitted.

5. Reporting Monte Carlo Results

Results based on Monte Carlo experiments should be reported as carefully as any other scientific experiment. Hoaglin and Andrews (1975) provided a slightly outdated list of items that should accompany any Monte Carlo based result. In principle, any information useful to assess the accuracy of the results and to facilitate their reproduction, should be supplied. As a minimum, taking into account recent development, the study should provide:

- information on the simulation, including the uniform random number generator used and method used to generate non-uniform variates, which should be fully adequate for the needs of the study,
- details on any measure employed to reduce variance,
- a justification for the sample size chosen possibly in terms of standard deviation of the estimates obtained in the study,
- detailed information of programming languages or software applications used, vendor, version, serial number, alternative platforms on which it runs, etc., and
- information on the computer used, including details on the CPU, and operating system.[5]

Geweke (1996) suggests also that any published result should be checked for robustness to the choice of generator. All the items listed above provide information to help assess the accuracy of the Monte Carlo computer-based results. It is assumed that computations follow the current state of the art. Preference should be given to well-known, good algorithms and software available in the

[5]It is worth remembering that in the fall of 1994, a serious design flaw was discovered in the Intel Pentium processor, commonly referred to as the "Pentium floating-point-division bug" or "Pentium bug" in short. As a consequence, certain floating-point division operations performed by the Pentium processor produced incorrect results.

public domain. Random number generators not in the public domain that have not being tested before, should be assessed both theoretically and empirically before use (see Section 7).

Typically Monte Carlo results are presented in a tabular form. However, sometimes other forms can convey the results from the Monte Carlo experiments more effectively. For instance, when the distributional characteristics of the sampling distribution of a test statistic are of interest, graphical methods, such as histograms and density estimates, can be used. When a large number of Monte Carlo experiments are performed, other methods, such as estimating a response surface, have been effectively used in the past to summarize the results (see, Davidson and MacKinnon, 1993, for more details on the use of response surfaces in relation to Monte Carlo experiments).

6. Random Number Generation

As we noted, a Monte Carlo method is a controlled statistical experiment executed on a computer using algorithms that produce deterministic, repeating, sequences of computer numbers, referred to as *pseudo random* numbers, that "appear" as random samples drawn from a known distribution, typically, samples of independent and identically distributed $U(0,1)$ random variables. An algorithm that generates such sequences of pseudo-random numbers is commonly known as a *random number generator* (RNG).

Many programming languages, adopt the so called *linear congruential generator* (LCG) introduced by Lehmer (1949). It is obvious that the pseudo-random number sequences produced by such a generator can be considered "random" only in some limited sense. Nonetheless, their imitation of "truly" random behavior is often good enough for our purposes. The LCG is defined by the difference equation:

$$X_{n+1} = (aX_n + c) \bmod m, \quad X(0) = X_0, \quad n \geq 0, \tag{6.1}$$

for a *multiplier* a, $0 \leq a < m$, *shift* (or *increment*) c, $0 \leq c < m$, and a *modulus* m, $0 \leq m$, all integers. The sequence of pseudo-random numbers U_n is determined by equation 6.1 and by the normalization

$$U_n = X_n/m,$$

once the *seed*, X_0 is given. See Section 7.1 for more information about these parameters. For a quick "back-of-the-envelope" Monte Carlo experiment, to generate non-uniform variates, a direct application of standard theorems from mathematical statistics (see, e.g., as Hogg, McKean, and Craig, 2005), summarized in Figure 16.2, can be used. For instance, to generate pseudo-random normal numbers is the Box-Muller method. It exploits the fact that given two independent, uniformly distributed random variables, U_1 and U_2, the random

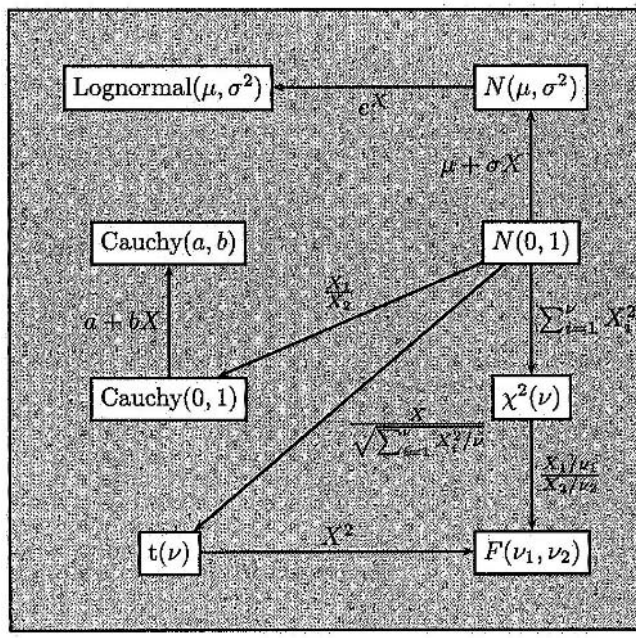


Figure 16.2. Relationships between the Standard Normal and related distributions.

variables, N_1 and N_2, obtained from the transformation:

$$\begin{aligned} N_1 &= \sqrt{-2\log(U_1)}\cos(2\pi U_2) \\ N_2 &= \sqrt{-2\log(U_1)}\sin(2\pi U_2) \end{aligned}$$

are independent standard normal random variables (see, e.g., as Hogg, McKean, and Craig, 2005, pages 290–291). To generate non-uniform variates related to the standard normal the "composition" theorems from mathematical statistics as summarized in Figure 16.2, can be used.

These approaches are simple to implement when feasible, however they are generally very inefficient and should not be used for serious research.

In general, it is preferable to avoid "reinventing the wheel" by re-writing the code implementing a random number generator when well-known good code is already available. Many useful generators are coded in languages such as FORTRAN and C (see, e.g., Gentle, 2003).

Modern statistical and econometric software application provide many useful functions for random number generation. R, for instance, offers a variety of uniform and non-uniform random number generators. The function *RNGkind* can be used to select among various uniform and normal generators. It also allows user-defined function to be used. For instance, the command

```
RNGkind( kind = "Knuth-TAOCP", normal.kind = "Box-Muller" )
```

Table 16.1. R Functions for Random Number Generation.

Name	*Distribution*	*Parameters*	*Defaults*
rbeta	beta	shape1, shape2	-, -
rbinom	Binomial	size, prob	-, -
rcauchy	Cauchy	loc, scale	0, 1
rchisq	chi-square	df	-
rexp	exponential	-	-
rf	F	df1, df2	-, -
rgamma	Gamma	shape	-
rgeom	Geometric	prob	-
rhyper	Hypergeometric	m, n, k	-,-,-
rlnorm	log-normal	mean, sd (of log)	0, 1
rlogis	logistic	loc, scale	0, 1
rnbinom	Negative Binomial	size, prob	-, -
rnorm	normal	mean, sd	0, 1
rpois	Poisson	lambda	-
rstab	stable	index, skew	-, 0
rt	Student's t	df	-
rtukey	Studentized Range	nmeans, df, nranges	-, -, 1
runif	uniform	min, max	0, 1
rweibull	Weibull	shape, scale	-, 1
rwilcox	Wilcoxon Rank Sum Statistic	m, n	-, -
rsignrank	Wilcoxon Signed Rank Statistic	n	-

selects the Knuth-TAOCP RNG (see, Knuth, 1998), as the default uniform random number generator, and the Box-Muller algorithm as the default normal RNG. The selected uniform random number generator is then used to obtain all other non-uniform variates. Table 16.1 shows the names and the parameters of some the uniform R's various non-uniform RNG functions useful for parametric and nonparametric Monte Carlo methods. For instance, to generate a sample of size 20 from a binomial distribution with 10 independent Bernoulli trials, with probability of success fixed to 0.4 across trial, we can use the following command.

```
>  rbinom(20,10,.4)
 [1] 5 2 5 5 5 3 3 5 3 4 5 4 6 5 6 3 3 1 2 4
```

Modern languages offer many other functions and specialized libraries, that can facilitate Monte Carlo studies. Consider the case of bootstrapping. For instance, to obtain samples of size 12, with replacement, from a vector (in this case the sequence of integers between 1 and 12), in R we can type

```
> sample(1:12, size=12, replace=T)
 [1]  8  5  9 11  4  4 10 11 10  9  4 10
```

It is important to always carefully look for the available options in a software instruction reference before implementing a Monte Carlo study.

7. Quality of Random Number Generators

Three types of considerations should guide the practitioner in the choice of the appropriate RNG: theoretical, empirical, and practical. These three aspects should always be jointly considered when selecting a suitable random number generator for a Monte Carlo study.

7.1 Theoretical Arguments: Period Length

From a theoretical point of view one of the most critical issues concerning a random number generator is its period length. The finite period of the generators implies that there is a limit on how many random numbers can be actually generated with a particular RNG. Arguments based on the "drawing without replacement" analogy and on more refined spatial considerations, have been used to justify the need to choose a generator with a period much larger than the number of random numbers actually needed for a Monte Carlo study. The recommended minimum length of the period, p, is taken to be a function of the number of random draws, n, used in the Monte Carlo experiment. Knuth, for instance, (1998, p. 185) recommended $p > 1000 \cdot n$. Other authors have made more conservative suggestions. MacLaren (1992) recommended $p > n^2$, while Hellekalak (1998) and Ripley (1987, p. 26) recommended a much more conservative $p \gg 200 \cdot n^2$. For instance, if 8 million numbers are required, p should be at least $8 \cdot 10^9 \approx 2^{33}$ following Knuth's recommendation, and $12,800 \cdot 10^{12} \approx 2^{54}$, following Ripley's suggestion. Empirical evidence supports these recommendations (see, e.g., L'Ecuyer, 1998, and references therein). It is worth highlighting that the above mentioned arguments relegate the use of a single LCG only to exploratory or small Monte Carlo projects. Any serious or intensive application would require generators with much larger periods.

Another approach, suggested by Knuth (1998), is to set the parameters of a LCG (see, Equation 6.1) to different values every time a few million numbers have been generated. For instance, in GAUSS, the functions *rndcon* and *rndmult* can be used to reset the constant shift parameter, c, and the multiplier, a, respectively of the uniform RNG. All parameters must be integers in the range $[0, 2^{31} - 1]$. The default values are $c = 0$, $a = 397204094$, and $m = 2^{32}$. We can immediately see, that some choices will produce numbers that do not appear random at all. For instance, setting $seed = 312479559$, $a = 1$, and

$c = 0$ produces a sequence with constant value $seed/m = 312479559/2^{32}$. In R,[6]

```
seed <- 312479559
a    <- 1
c    <- 0
m    <- 2^32
x    <- ( a * seed + c ) %% m
u    <- x / m
u
```

returns 0.07275482 (by default R displays only 7 digits[7]). Other choices are much harder to assess. For instance, in GAUSS, if we change the default multiplier, from $a = 1$ to $a = 65539$, using the following set of commands:

```
seed = 312479559; a = 2^16 + 3; rndmult a; rndseed seed;
```

we obtain a well-known "bad" RNG. As Figure 16.3 clearly illustrates, though the generated uniform random sequence, when plotted in pairs in the unit square does not arouse e any suspicion, it appears extremely correlated when plotted in triplets in the unit cube. In fact, all the points lie on 15 planes in the three-dimensional space. This should highlight the fact that setting the parameter of a LCG requires considerable care.

For a theorem on how to choose a and c in order to guarantee that the maximum period is achieved and to obtain parameter values that are known to produce reasonable generators, consult Knuth (1998).

7.2 Empirical Support

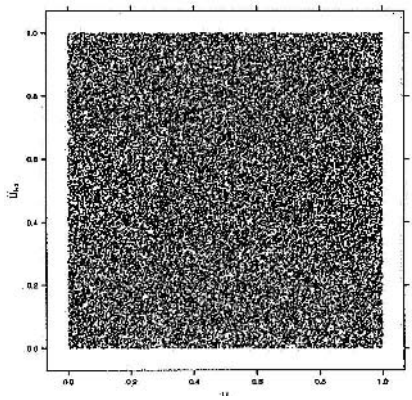
Theoretical arguments alone are not enough to help in the selection of the appropriate RNG method. In fact, even if supported by sound theoretical arguments, a particular implementation of a RNG in a software application still requires testing, as they might be implemented incorrectly.

Many random number generators implemented in various software application are known to fail even simple test of randomness (see, e.g., Sawitzki, 1985; Park and Miller, 1988; and McCullough, 1999, 2000). It is advisable to perform many tests before concluding that it is safe to use a particular generator. The quality of non-uniform random variates depends closely on the quality of the uniform random numbers on which they are based. For instance, the Box-Muller method is not recommended by the literature on random number generation because of its slowness and sensitivity to the underlying uniform random number generator (see, e.g., Ripley, 1987).

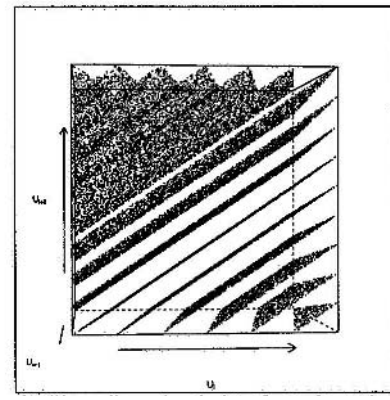
[6]Note that R uses double-precision arithmetic and not integer arithmetic as required to implement the LCG. Even using the function *as.integer* the RNG cannot be correctly reproduced since 2^{32} overflows in R whereas it is ignored in computer system implementing two's complement arithmetic.

[7]The version of GAUSS used for this chapter actually produces 0.072754817 with those settings.

Figure 16.3. Plots of 100,000 draws from GAUSS's LCG with $a = 2^{15} + 3$, $c = 0$, and $m = 2^{32}$.



(a) Two-dimensional plot of pseudo-random pairs (U_i, U_{i+1}).



(b) Three-dimensional plot of pseudo-random triplets (U_i, U_{i+1}, U_{i+2}).

There are several collection of tests for uniform RNG in wide use today (see, e.g., Gentle, 2003). A popular collection is the DIEHARD battery of randomness tests introduced by Marsaglia (1996). This battery of tests provides a wide range of statistical tests for evaluating the stream of output of RNGs. The DIEHARD program, provided as an MSDOS executable or as C source code, is freely available from *http://stat.fsu.edu/∼geo*. The DIEHARD battery of tests[8] requires as input a specially formatted binary file of 10 to 11 megabytes size. The RNG should produce 32-bit positive integers that need to be saved in a text file in hexadecimal form, 8 hex 'digits' per integer, 10 integers per line, and with no intervening spaces. Consider testing a random number generator in GAUSS for Windows. The *rndKMi* function in GAUSS returns a matrix of random integers, between 0 and 2^{32}, and the state (seed) of the random number generator. A file containing 3 millions uniform random numbers can be created by typing the following commands in the GAUSS for Windows command prompt:

```
seed            = 799;
output file     = rndKMi.dat reset;
{ x, new_seed } = rndKMi( 3e6, 1, seed);
screen off; print  x; output off;
```

[8]The version available at the moment of writing was: DOS, Jan 7, 1997.

We can create an *hex* file satisfying the stated conditions using the following Perl[9] script:

```
open( IN,  "rndKMi.dat" );
open( OUT, ">rndKMi.hex" );
while ( $line = <IN> ) {
    chomp $line;
    printf OUT "%08x", $line;    # prints random 32-bit integer
    if ( $. % 10 == 0 ) { printf OUT "\n" }
    ;                            # new line every 10 no.
}
```

which saves the output in a file named *rndKMi.hex*. The MSDOS auxiliary program, *asc2bin.exe*, provided in the test suit, can then be used to convert the hexadecimal file into a binary file, that can be directly fed to the main DIEHARD program. The following script runs the auxiliary `asc2bin` utility and the main `diehard` programs providing them with the necessary input arguments. The results from the 18 DIEHARD tests is saved in the file *rndKMi.out*.

```
# converts hex file to binary file using utility program asc2bin
open( INPUT, "| asc2bin" );
select(INPUT);
print "\n";
print "\n";
print "\n";
print "rndKMi.hex\n";    # hex input file name
print "rndKMi.bin\n";    # binary output file name
close(INPUT);

# performs DIEHARD test
open( INPUT, "| diehard" );
select(INPUT);
print "rndKMi.bin\n";          # binary input file name
print "rndKMi.out\n";          # test output file name
print "111111111111111\n";     # asks for all 18 tests
close(INPUT);
```

These tests are general and are not able to detect all possible problems a generator might encounter in practice. With a specific application in mind it is possible to design *ad hoc* tests that are more suitable to assess the usefulness of a generator for a particular application.

[9]Perl, an acronym for *Practical Extraction and Report Language* is a cross-platform, high-level, open-source programming language with network and object-oriented programming support. Perl is derived mainly from the C programming language and to a lesser extent from sed, awk, grep, and the Unix shell. Perl's process, file, and text manipulation facilities make it particularly well-suited for a wide range of tasks such as data preparation and processing. Perl can also automate what would otherwise be tedious, repetitive, and error prone activities. Perl can be downloaded from the WWW's Comprehensive Perl Archive Network (CPAN) located at *http://www.cpan.org/*, where the source package, modules, documentation, and links to binaries distributions of Perl are available for various platforms, including Win32, Mac, and Unix/Linux. For more on Perl and its uses in econometrics and statistics, see Baiocchi (2003) and Baiocchi (2004).

For non-uniform variates a chi-squared or a Kolmogorov-Smirnov goodness-of-fit test cab be used to test the quality of the generator. For more information of generating and testing non-uniform random variates, consult Devroye (1986), Ripley (1987), and Gentle (2003).

8. Practical Issues

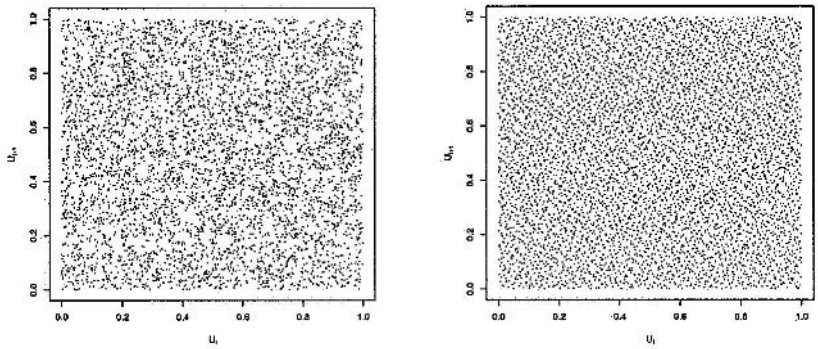
Several important practical issue can play a role in the choice of a generator. We will briefly mention only two here for the sake of time and space: execution time and available software options.

Execution time considerations are not usually thought to be relevant for economic applications (see, e.g., Geweke, 1996). The assumption is that computations using the pseudo-random numbers are generally much more time consuming than generating them. However, for some applications considerable savings can be afforded by using the appropriate generation method. For instance, Train (1999) and Bhat (2000, 2001) by introducing an alternative type of random number generator in the estimation of mixed logit models, found that the accuracy of the results is greatly improved. For four and five dimension integrals, the quasi-random approach based on Halton sequences required 125 draws to achieve the same level accuracy as 2,000 draws obtained using standard pseudo-random number sequences. They found that, in terms of execution time, using quasi-random numbers can take only 10 per cent of the time required with pseudo-random numbers to deliver the same degree of accuracy. Figure 16.4 illustrates the difference between the two random number generation approaches.[10] The figure clearly illustrates that quasi-random pairs provide a much better coverage in $[0, 1]^2$. For implementation in practice tables of parameters for good random number generators (see, e.g., Knuth, 1998, page 106) are needed. Portability is an issue of practical relevance. For reproducibility purposes a portable implementations a generator should be preferable. Portability issues are often related to hardware structure, differences in the supporting software standard, etc.

Ideally, the software and algorithms used in a Monte Carlo should be both subject to peer review and based on openly published and freely available algorithms and source code. Freely available and open-source software would also guarantee the highest degree of quality and reproducibility. Also, though most university have a portfolio of software applications useful to economists, there will always be researchers with no access to all these applications. Finally, vendors of proprietary software rarely describe the algorithms used to implement econometric and statistical procedures, nor do they provide sufficiently detailed information about their reliability. This is a serious omission

[10]The Halton sequences were obtained using R's *rnorm.halton* function provided by the `fOptions` library.

Figure 16.4. 5,000 draws in two dimensions from uniform distribution.



(a) Plot of pseudo-random pairs (U_i, U_{i+1}). (b) Plot of quasi-random pairs (U_i, U_{i+1}).

that makes the use of "black box" packages less attractive for academic research.

9. Testing the Random Number Generators

The results of Marsaglia's DIEHARD batteries of tests of the uniform random number generators in GAUSS for Windows version 6.0, LIMDEP version 7.0,[11] and R version 2.0.0, are shown in Table 16.2. Care is required when analyzing the output of the tests. The "p-values" of the DIEHARD test are actually values of the CDF of the test statistics. For instance, consider the output line taken from a DIEHARD test output file.

```
chisquare for 99 degrees of freedom=100.946; p-value= .573288
```

The "p-value" is the value at 100.946 of the cumulative chi-squared distribution function with 3 degrees of freedom, as the following calculation in R shows.

```
> pchisq(100.946,99)
[1] 0.5732773
```

When testing the random numbers we want to reject not only when the observed frequencies are too different from the expected ones, but also when they agree too closely.[12] The standard practice of doubling the one-sided p-value to

[11] The results for LIMDEP are taken from McCullough (1999, p. 199).

[12] Fisher (1936) suggested that the null in a goodness-of-fit test could also be rejected if the p-value is too large. Sometimes the agreement between observed and expected frequencies is "too good" to have arisen by chance only. According to Fisher (1936):

obtain a two-sided p-value is inappropriate for an asymmetric distribution. The CDF scale can be conveniently used to assess the results of the tests. A large CDF-value corresponds to a small p-value and indicates a significant result. A small CDF-value indicates a close match between observed and expected frequencies. In theory, if the CDF-value is less that 0.025 or greater than 0.975 we reject the null at the 5 per cent significance level, however, in practice, because of the large number of tests, unless stated differently, a test is considered failed if all CDF-values are either zero or one. Table 16.2 summarizes all the results of the DIEHARD tests.

LIMDEP uses L'Ecuyers (1999) multiple part random number generator. This generator has a period of roughly 2^{191}.

GAUSS implements two generators. The functions *rndi* (and *rndus*) is based on the multiplicative ($c = 0$) congruential generator with period 2^{32}. GAUSS' *rndKMi* (and *rndKMu*) function is reportedly (see, Ford and Ford, 2001) based on the "KISS" and on the "Monster" uniform generators introduced by Marsaglia (2000) and has period of greater then 10^{8888} and dimension greater than 920. The KISS generator is used to initialize the 920 seeds that the Monster generator number generator requires. The initialization of this generator requires 920 random numbers. Marsaglia's KISS generator has a period of about 2^{124} and itself requires 6 seeds to be initialized. In order to simplify the use of *rndkmi* all but one of these seeds are chosen according to Marsaglia's suggestions (Marsaglia, 2000). In practice, for the the KISS generator, only one seed requires setting.

Results in table 16.2 show that R passes all the tests[13] and, given its long period, appears to be suitable for large projects. According to McCullough (1998) results, LIMDEP passes almost all tests, and because of its shorter period is useful for small projects. The results for the GAUSS *rndi* (and *rndus*) function agrees with Vinod's (2000) findings. The *rndKMi* (and *rndKMu*), which was not tested before, fails 5 tests. We agree with Vinod's conclusions that GAUSS fails too many tests to be trusted for serious Monte Carlo simulations.

These test results should be interpreted with extreme caution. The DIEHARD tests were designed with generators with a 2^{32} period in mind. The observed inverse relationship between period length and number of failed tests suggests that further testing is required, before a more definite conclusion on the safety of a generator can be reached.

> Fictitious data can seldom survive a careful scrutiny, and, since most men underestimate the frequency of large deviations arising by chance, ...

[13] For details on the generator used by R see footnote 2.

Table 16.2. Marsaglia's DIEHARD tests.

Test	*GAUSS rndi*	*GAUSS rndKMi*	*LimDep*	*R*
Birthday Spacings	fail[a]	fail[b]	pass	pass
Overlapping 5-Permutation	pass	pass	pass	pass
Binary Rank (31×31)	pass[c]	pass[d]	pass	pass
Binary Rank (32×32)	fail	fail	pass	pass
Binary Rank (6×8)	fail[e]	fail[f]	pass	pass
Bitstream (p-values)	fail[g]	fail[g]	pass	pass
OPSO	fail[h]	pass	fail	pass
OQSO	fail[h]	pass	pass	pass
DNA	fail[h]	pass	pass	pass
Count the 1's (stream of bytes)	fail	fail	pass	pass
Count the 1's (specific byte)	pass[i]	pass	pass	pass
Parking Lot	pass	pass	pass	pass
Minimum Distance	pass	pass	pass	pass
3-D Spheres	pass	pass[j]	pass	pass
Squeeze	pass	pass	pass	pass
Overlapping Sums	pass	pass	pass	pass
Runs	pass	pass	pass	pass
Craps	pass	pass	pass	pass

[a]The CDF-value of KS test of uniformity for the 9 CDF-values is 1.
[b]The CDF-value of KS test of uniformity for the 9 CDF-values is 1.
[c]The CDF-value of the χ^2 goodness-of-fit test is 0.987400.
[d]The CDF-value of the χ^2 goodness-of-fit test is 0.999997.
[e]The CDF-value of KS test of uniformity for the 25 CDF-values is 1.
[f]The CDF-value of KS test of uniformity for the 25 CDF-values is 0.999940.
[g]All 20 CDF-values are equal to 1.
[h]Most CDF-values are either 1 or 0.
[i]8 CDF-values out of 25 are equal to 1. Two are below 0.005.
[j]The CDF-value of KS test of uniformity for the 20 CDF-values is 0.005550.

10. Conclusion

In this chapter we have briefly reviewed some areas of application of the Monte Carlo method in environmental and resource economics and discussed several important issues related to the implementation of Monte Carlo studies.

First, we have focussed on design issues that the practitioner should bear in mind when setting up a Monte Carlo study. Though the quality and usefulness of the Monte Carlo study critically depends on a good understanding of the specific theoretical and empirical issues underpinning the environmental issues under scrutiny, knowledge of simulation principles can help to economize the use of the researcher's time, and ensure that the results are correct and of high quality.

Second, we have considered the important issues of reproducibility of Monte Carlo computer-based simulation and illustrated the guidelines for Monte Carlo results to be effectively and accurately communicated.

Third, we have briefly introduced random number generation on computers and discussed the theoretical, empirical, and practical issues involved in the selection of the appropriate random number generator for Monte Carlo applications. We have highlighted the importance of selecting a random number generator with good theoretical properties, that has been thoroughly tested using current best available methods, and that is reasonably easy to implement in practice.

Finally, we have demonstrated how to test a random number generator in practice and compared several popular software application used by environmental and resource economists. We have highlighted the potential dangers of using existing software without careful consideration.

Monte Carlo methods are now routinely used in environmental economics to solve problems that were considered intractable only a few years ago. With more computationally intensive tools available to the practitioner for the development, estimation, and testing of more realistic environmental and resource economics models, I reckon that the challenges for the future will be:

- the further development and exploration of quantitative methodologies for environmental policy analysis, including more realistic and useful behavioral model, understanding the condition for their applicability and their limitations,
- integration with other discipline, such as geography (for instance, use of spatial data) and other social sciences,
- the development and dissemination of good practice in the obtainment, reporting, and reproduction of computer-based simulation results, and
- as our capacity to deal with more complex models is increased, the ability to identify more complex models will hinge upon the improvement

of experimental designs and collection of more detailed, high quality, comprehensive, and integrated data.

Chapter 17

GAUSSIAN QUADRATURE VERSUS SIMULATION FOR THE ESTIMATION OF RANDOM PARAMETERS

Some Evidence from Stated-Preference Choice Data

William S. Breffle
Stratus Consulting Inc.
bbreffle@stratusconsulting.com

Edward Morey
Department of Economics, University of Colorado
Edward.Morey@colorado.edu

Donald Waldman
Department of Economics, University of Colorado
Donald.Waldman@colorado.edu

Abstract In environmental economics, numerical simulation using random draws is the method most commonly used to estimate joint probabilities of individual choices in discrete-choice, random-parameters models. This paper compares simulation to another method of estimation, Gaussian quadrature, on the basis of speed and accuracy. The comparison is done using stated preference data consisting of the answers to choice questions for fishing in Green Bay, a large bay on Lake Michigan. Each sampled individual chose between a pair of Green Bay scenarios with different fishing conditions. Quadrature is found to be as accurate as simulation based on random draws, but Gaussian quadrature attains stability in estimated parameters considerably faster.

Keywords: Gaussian Hermite quadrature, random parameters, discrete-choice model, recreational fishing, fish consumption advisories.

1. Introduction

Simulation is the most common numerical method used to estimate random parameters models. Using pseudo-random draws to simulate moments of a distribution dates back to McFadden, 1989,[1] and has been used to model demand for an array of environmental commodities. Its use is likely to increase as a result of this volume and the recent book by Train, 2003.[2]

While simulation studies are numerous, little research has been conducted to examine the speed and accuracy of the simulation method, or the importance of addressing simulation noise (variance due to simulation). Brownstone and Train (1999) examine the sensitivity of average probabilities, the log-likelihood function, and parameter gradients to different numbers of draws and different sets of pseudo-random numbers (that is, different values for the random number generator seed), holding constant the values of the parameters that generate the data as they conduct the tests (that is, a new model is not estimated for every value of the seed). Breffle and Morey (2000) determine the minimum number of draws needed to stabilize the parameter estimates, and to minimize the simulation noise in estimated expected consumer surplus.

The purpose of this chapter is to compare simulation to another method of estimation, Gaussian quadrature (see Butler and Moffit, 1982; Waldman, 1985; Geweke, 1996). The comparison is done using stated preference data consisting of the answers to choice questions for fishing in Green Bay. Each sampled individual chose between a pair of Green Bay scenarios with different fishing conditions. Choices were made over multiple pairs with varied characteristics.

2. A choice model with random parameters

Assume individual i answers J pair-wise choice questions. Assume the utility from choosing alternative K in pair j is:

$$U_{ij}^{K} = \beta_i' \mathbf{x}_{ij}^{K} + \varepsilon_{ij}^{K}, \tag{2.1}$$

where the $L \times 1$ vector $\mathbf{x}_{ij}^{K}$ contains the observed characteristics of alternative K. The stochastic component ε_{ij}^{K} is assumed to be i.i.d. and normally distributed, generating a probit model. Under the assumption of random heterogeneous preferences,

$$\beta_i = \beta + \mathbf{u}_i, \text{ where } \mathbf{u} \sim N(\mathbf{0}, \boldsymbol{\Sigma}), \tag{2.2}$$

[1] See Hajivassiliou et al., 1992, McFadden and Ruud, 1994, and Stern, 1992.

[2] See, for example, Train, 1998, Brownstone and Train, 1999, Revelt and Train, 1998, Layton and Brown, 2000, and Breffle and Morey, 2000. Train, 2003 includes additional examples. This volume contains many more.

where $\mathbf{u}_i$ is a random $L \times 1$ vector representing the differences between the mean marginal utilities of the characteristics and individual i's marginal utilities, and $\mathbf{\Sigma}$ is the variance-covariance matrix of $\mathbf{u}$. Assuming that $\mathbf{u}_i$ is constant across choice occasions causes an individual's different choices to be correlated. Under the assumption that $\mathbf{\Sigma}$ is diagonal (that is, the random parameters are uncorrelated with one another), the joint probability of observing the J pair-wise choices of individual i, P_i, is:

$$\begin{aligned} P_i &= P(U_{i1}^k > U_{i1}^{3-k}, ..., U_{iJ}^k > U_{iJ}^{3-k}) \\ &= P(\varepsilon_{i1}^{3-k} - \varepsilon_{i1}^k < \beta_i'(\mathbf{x}_{i1}^{3-k} - \mathbf{x}_{i1}^k), ..., \varepsilon_{iJ}^{3-k} - \varepsilon_{iJ}^k < \beta_i'(\mathbf{x}_{iJ}^{3-k} - \mathbf{x}^k) \\ &= \int_{-\infty}^{+\infty} \cdots \int \prod_{j=1}^{J} \Phi\left[-(\beta + \mathbf{u})'(\mathbf{x}_{ij}^{3-k} - \mathbf{x}_{ij}^k)\right] f(\mathbf{u}) du_1 ... du_L \qquad (2.3) \end{aligned}$$

where $k = 1$ if alternative 1 was chosen and 2 if alternative 2 was chosen; $\Phi(\cdot)$ is the univariate, standard-normal CDF and $f(\cdot)$ is the L-variate, normal density function. That is, $\Phi(\cdot)$ is one dimensional, $f(\mathbf{u})$ is L-dimensional, $\Phi(\cdot)$ is standard normal, and $f(\mathbf{u})$ is normal, but not standard normal.

The order of magnitude of the integral in Equation 2.3 is the number of random parameters (the number of non-zero diagonal elements in $\mathbf{\Sigma}$). For example, if there are two random parameters and the parameters are ordered such that the two random parameters are in the first and second positions, Equation 2.3 simplifies to:

$$P_i = \int_{-\infty}^{+\infty} \int \prod_{j=1}^{J} \Phi\left[-(\beta + \begin{matrix}\mathbf{u} \\ \mathbf{0}\end{matrix})'(\mathbf{x}_{ij}^{3-k} - \mathbf{x}_{ij}^k)\right] f(\mathbf{u}) du_1 du_2 \qquad (2.4)$$

where $\mathbf{u}$ now has only two elements and $\mathbf{0}$ is a vector of dimension $(L-2) \times 1$.

The likelihood function is $\prod_{i=1}^{N} P_i$. The P_i can be approximated using either simulation or Gaussian quadrature.

3. Simulation

Using simulation, the multiple integral, Equation 2.3, is approximated in steps: first, a pseudo-random draw, $\mathbf{u}^d$, is taken from $\mathbf{u} \sim N(\mathbf{0}, \mathbf{\Sigma})$, then P_i is computed conditional on that draw, $P_i(\mathbf{u}^d) = \prod_{j=1}^{J} \Phi\left[-(\beta + \mathbf{u}^d)'(\mathbf{x}_{ij}^{3-k} - \mathbf{x}_{ij}^k)\right]$.[3] These two steps are repeated for D draws.

[3]Note that the $\mathbf{u}$ draw is conditional on $\mathbf{\Sigma}$. When one simulates probabilities in the context of maximum likelihood estimation, the process is called maximum simulated likelihood. The objective is to estimate the β and $\mathbf{\Sigma}$, but at each iteration, $P_i(\mathbf{u}^d)$ is conditional on the current iterative values of β and $\mathbf{\Sigma}$.

The simulated P_i is the average of the conditional $P_i(\mathbf{u}^d)$'s. Details may be found in Train (1998). The attractive feature of simulation is its versatility: it can be used to approximate many different types of probability integrals.

A problem with simulating probabilities is "simulation noise". That is, for a given β and $\mathbf{\Sigma}$, the simulated P_i is different each time it is re-simulated with a new set of random draws. A simulated P_i is an approximation that differs every time the approximation formula is reapplied, so P_i has a distribution. If the number of draws is "small" this distribution will have a variance sufficiently large that substantively different parameter estimates will result each time the model is estimated. Simulation noise declines, for a given sample size, as the number of draws increases, becoming trivial if the number of draws is large enough.[4] Put simply, the minimum number of draws required for estimation by simulation is the number necessary for model parameters to be stable.[5] Determining this number is time consuming. For example, to check whether $1,000$ draws is sufficient one must re-estimate the model many times with $1,000$ draws to make sure the parameter estimates are "almost" the same in each simulation. If they are not, one has to repeat this exercise with a larger number of draws.

4. Gaussian-Hermite quadrature

Alternatively, Equation 2.3 can be approximated using a form of Gaussian quadrature called Hermite polynomial quadrature, or Gaussian-Hermite quadrature. Put simply, any integral of the form $\int_{-\infty}^{+\infty} g(v)dv = \int_{-\infty}^{+\infty} e^{-v^2}h(v)dv$ can be approximated by

$$\int_{-\infty}^{+\infty} e^{-v^2}h(v)dv \approx \sum_{m=1}^{M} w_m h(v_m) \tag{4.1}$$

where M is the number of evaluation points.[6] Put simply, the function to be integrated, minus the e^{-v^2} term, is first evaluated at a number of carefully selected evaluation points, the v_m, $m = 1, 2, ..., M$; then the value of the function at each evaluation point, $h(v_m)$ is weighted by w_m.[7] The approximation of the integral is the sum of these weighted values. Implementation is simple: tables (such as table 25.10 in Abramowitz and Stegun, 1964) report the (v_m, w_m)

[4] In addition, for a given number of draws in the simulation, simulation noise decreases as the number of answered choice-pairs increases.

[5] Stability is subjective and depends on how the parameter estimates will be used.

[6] The accuracy of the approximation increases with the number of evaluation points. Typically, 10 or fewer are sufficient for parameter stability, not hundreds or thousands as with simulation.

[7] As explained below $w_m \neq e^{-v^2}$.

pairs for different ascending values of M; one just has to use them to calculate the $w_m h(v_m)$ terms. It is called Gaussian quadrature because $\int_{-\infty}^{+\infty} e^{-v^2} dv$ is the *Gaussian integral*, sometimes called the *probability integral*.[8] Gaussian quadrature requires fewer evaluation points than ordinary quadrature (approximately half).

The $v_1, v_2, ..., v_M$ are the M roots of the Hermite polynomial $H_M(v)$; this is why the technique is called Hermite polynomial quadrature.[9] The numerical weights, the w_m, are more difficult to explain, but implementation does not require one to understand why they are the number they are.[10]

To solidify how Hermite polynomial quadrature works, we use it here to determine the area under a univariate, normal density function with zero mean, $f_u(0, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-u^2/2\sigma^2}$, which we know is 1.

Begin by noting that $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-.5(u^2/\sigma^2)} du = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} e^{-v^2} dv$ where $v = \frac{u}{\sqrt{2}\sigma}$. Then note that $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} e^{-v^2} dv$ is a simple, special case of $\int_{-\infty}^{+\infty} e^{-v^2} f(v) dv$, where $f(v)$ is simply the constant $\frac{1}{\sqrt{\pi}}$.

Therefore, $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} e^{-v^2} dv = \sum_{m=1}^{M} w_m f(v_m) = \frac{1}{\sqrt{\pi}} \sum_{m=1}^{M} w_m$.

---

[8]Note that $\int_{-\infty}^{+\infty} e^{-v^2} dv = \sqrt{\pi}$.

[9]Hermite polynomials are a set of orthogonal polynomials with domain $(-\infty, +\infty)$. By way of example, $H_0(v) = 1$, $H_1(v) = 2v$, $H_2(v) = 4v^2 - 2$, $H_3(v) = 8v^2 - 12v$, $H_4(v) = 16v^4 - 48v^2 + 2$ and $H_{10}(v) = 1024v^{10} - 23040v^8 + 161280v^6 - 403200v^4 + 302400v^2 - 30240$. Visually, these polynomials become more and more wave-like as M increases. The M roots of $H_M(v)$, the v_m, are those values of v that make the Hermite polynomial zero. For example $H_2(v)$ has two roots: $\pm.5\sqrt{2}$. And, the four roots of $H_4(v)$ are $\pm\frac{1}{2}\sqrt{\sqrt{34} + 6}$and $\pm\frac{1}{2}\sqrt{6 - \sqrt{34}}$. *Mathematica*, for example, generates Hermite polynomials with the command $HermiteH[M, v]$.

[10]Here are some of the w_m, reported with their corresponding values of v_m

M	v_m	w_m
2	$\pm\frac{1}{2}\sqrt{2} = \pm 0.707107$	$\frac{1}{2}\sqrt{\pi} = 0.886227$
3	0	$\frac{2}{3}\sqrt{\pi} = 1.18164$
	$\pm\frac{1}{2}\sqrt{6} = \pm 1.22474$	$\frac{1}{6}\sqrt{\pi} = 0.295409$
4	$\pm\sqrt{\frac{3-\sqrt{6}}{2}} = \pm 0.524648$	$\frac{\sqrt{\pi}}{4(3-\sqrt{6})} = 0.804914$
	$\pm\sqrt{\frac{3+\sqrt{6}}{2}} = \pm 1.65068$	$\frac{\sqrt{\pi}}{4(3+\sqrt{6})} = 0.0813128$

For each M , the weights sum to $\sqrt{\pi}$.

If $M = 2$, this is $\frac{1}{\sqrt{\pi}}(0.886227 + 0.886227) = 1.000\,000\,1$; if $M = 4$ it is $\frac{1}{\sqrt{\pi}}(0.804914 + 0.804914 + 0.0813128 + 0.0813128) = 0.999\,999\,86$, a closer approximation, but both are close enough.

As noted above, Gaussian quadrature requires that the integral be of the form $\int_{-\infty}^{+\infty} e^{-v^2} h(v) dv$; Equation 2.3 is of this form. For the case of one random-parameter, Equation 2.3 simplifies to

$$
\begin{aligned}
P_i &= \int_{-\infty}^{+\infty} \prod_{j=1}^{J} \Phi \left[- \left(\beta + \begin{array}{c} u \\ \mathbf{0} \end{array} \right)' \Delta \right] \phi(u) du \qquad (4.2) \\
&= \int_{-\infty}^{+\infty} \prod_{j=1}^{J} \Phi \left[- \left(\beta + \begin{array}{c} u \\ \mathbf{0} \end{array} \right)' \Delta \right] \left[\frac{1}{\sqrt{2\pi}\sigma} e^{-u^2/2\sigma^2} \right] du
\end{aligned}
$$

where Δ is $(\mathbf{x}_{ij}^{3-k} - \mathbf{x}_{ij}^{k})$, u is a scalar and $\mathbf{0}$ is a vector of dimension $(L-1)\times 1$. With a change of variable $v = \frac{u}{\sqrt{2}\sigma}$, this becomes

$$
\begin{aligned}
P_i &= \int_{-\infty}^{+\infty} \prod_{j=1}^{J} \Phi \left[- \left(\beta + \begin{array}{c} v\sigma\sqrt{2} \\ \mathbf{0} \end{array} \right)' \Delta \right] \left[\frac{1}{\sqrt{2\pi}\sigma} e^{-(v\sigma\sqrt{2})^2/2\sigma^2} \right] d(v\sigma\sqrt{2}) \\
&= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} \prod_{j=1}^{J} \Phi \left[- \left(\beta + \begin{array}{c} v\sigma\sqrt{2} \\ \mathbf{0} \end{array} \right)' \Delta \right] \left[e^{-v^2} \right] dv \\
&= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} h(v) \left[e^{-v^2} \right] dv \approx \frac{1}{\sqrt{\pi}} \sum_{m=1}^{M} w_m h(v_m) \qquad (4.3)
\end{aligned}
$$

where $h(v) = \prod_{j=1}^{J} \Phi \left[- \left(\beta + \begin{array}{c} v\sigma\sqrt{2} \\ \mathbf{0} \end{array} \right)' \Delta \right]$.

If there are two random parameters, as in Equation 2.4, and one adds the assumption they are uncorrelated,

$$
\begin{aligned}
P_i &= \int_{-\infty}^{\infty}\int \prod_{j=1}^{J} \Phi\left[-\left(\beta + \begin{matrix} u_1 \\ u_2 \\ \mathbf{0} \end{matrix}\right)' \Delta\right] f_{u_1}(u_1) f_{u_2}(u_2) du_1 du_2 \\
&= \int_{-\infty}^{+\infty}\left\{\int_{-\infty}^{+\infty} \prod_{j=1}^{J} \Phi\left[-\left(\beta + \begin{matrix} u_1 \\ u_2 \\ \mathbf{0} \end{matrix}\right)' \Delta\right] f_{u_1}(u_1) du_1\right\} f_{u_2}(u_2) du_2 \\
&= \frac{1}{\pi}\int_{-\infty}^{+\infty}\left\{\int_{-\infty}^{+\infty} \prod_{j=1}^{J} \Phi\left[-\left(\beta + \begin{matrix} v_1\sigma\sqrt{2} \\ v_2\sigma\sqrt{2} \\ \mathbf{0} \end{matrix}\right)' \Delta\right] e^{-v_1^2} dv_1\right\} e^{-v_2^2} dv_2 \\
&\approx \frac{1}{\pi}\sum_{m_1=1}^{M_2} w_{m_2}\left[\sum_{m_1=1}^{M_1} w_{m_1} h(v_{1m_1}, v_{2m_2})\right] \qquad (4.4)
\end{aligned}
$$

where

$$
h(v_{1m_1}, v_{2m_2}) = \prod_{j=1}^{J} \Phi\left[-\left(\beta + \begin{matrix} v_{1m1}\sigma\sqrt{2} \\ v_{2m_2}\sigma\sqrt{2} \\ \mathbf{0} \end{matrix}\right)' \Delta\right] \qquad (4.5)
$$

In summary, Gaussian-Hermite quadrature is easy to apply if one assumes a small number of uncorrelated normally-distributed random parameters, a common assumption in many applications.

5. An application: Green Bay fishing under different conditions

The sample consists of 647 randomly-sampled anglers who purchased licenses in eight counties near the bay of Green Bay, a large bay on Lake Michigan, and who fished the bay at least once in 1998. Each angler was presented eight pairs of Green Bay alternatives. Anglers made their selections on the basis of the average time to catch a fish for four species ($PERCH$, $TROUT$, $WALLEYE$, and $BASS$), a fish consumption advisory index (FCA), and a boat launch fee (FEE). The FCA index takes one of nine discrete levels, the first being no advisory, and contains combined information on separate advisories for the four species.

The four parameters on the catch rates and the eight parameters on the FCA dummy variables are all assumed to be random and normally distributed.

Specifically, for each angler and each pair, it is assumed that the conditional indirect utility function for alternative is K is

$$\begin{aligned} V_i^K &= (\beta_c + u_{ci}) \\ &\quad \times \left[\beta_p PERCH^K + \beta_t TROUT^K + \beta_w WALLEYE^K + \beta_b BASS^K\right] \\ &\quad + (\beta_{FCA} + u_{FCAi}) \left[\beta_{FCA2} FCA^K + ... + \beta_{FCA9} FCA^K\right] \\ &\quad + \beta_y(-FEE^K) \end{aligned} \tag{5.1}$$

where

$$f(u_c, u_{FCA}) = N\left(\begin{array}{c} 0 \\ 0 \end{array}, \begin{array}{cc} \sigma_c & 0 \\ 0 & \sigma_{FCA} \end{array}\right) \tag{5.2}$$

The parameter β_y is the marginal utility of money. The parameter β_c is the mean catch-parameter; the β_p,β_t,β_w and β_b allow the means and variances of the catch parameters to vary by species; β_{FCA} is the mean FCA parameter; and the $\beta_{FCA2}, ..., \beta_{FCA9}$ allow the means and variances of the FCA parameters to differ by FCA regime. With this specification, the ratio of the mean parameter to the standard deviation is the same for each of the four catch rates, and for each of the eight FCA levels, so only two standard deviation parameters need to be estimated, σ_c and σ_{FCA}. Assuming that the standard deviation varies in proportion to the mean is a common way of dealing with heteroskedasticity and a reasonable way to limit the number of random parameters that need to be estimated. It also causes an individual's four catch parameters to be correlated with one another, and his eight FCA parameters to be correlated with one another; something one would expect, and this is accomplished in a way that does not require that one assumes separate u_{cp} and u_{ct} that are correlated. Of course, this specification for the random structure is not necessarily appropriate for other applications and is in no way required for quadrature to work. β_p is set to one to identify the catch parameters and β_{FCA2} is set to one identify the FCA parameters.[11]

This model was estimated using both Hermite quadrature and simulation with pseudo-random draws. Parameter estimates are reported in Table 17.1. Results from various model runs show that 500 draws in simulation and 9 evaluation points using quadrature are sufficient for parameter estimates to be stable. That is, for quadrature parameter-estimates, when more than 9 evaluation points are used, the individual parameter estimates are never more than 2% different from the estimates obtained with 9 evaluation points. When at least 500 draws are used in simulation, parameter estimates vary by at most 2% across runs. An important finding is that simulation took almost three times longer than quadrature to reach this level of stability.

[11] Code for both estimation techniques can be found at http://www.colorado.edu/Economics/morey/dataset.html.

Results are also reported for 100 draws using simulation, and 3 and 6 evaluation points for quadrature,but one should not make too much of these. Comparing the properties of estimates estimated with too few quadrature points to estimates estimated with too few random draws is a questionable endeavor: one would never present either as one's parameter estimates.

6. Conclusions

In conclusion, this paper provides an example of a choice question probit model with two random parameters. We demonstrate that to obtain a high level of accuracy, quadrature is faster than simulation with pseudo-random draws. What should be made of this? When it comes to estimating random-parameter models, there is an alternative to the ubiquitous method of simulation with random draws, and it can be faster and more accurate. It's best suited to applications with a small number of uncorrelated random parameters.

An interesting issue is what happens to the magnitudes of the simulation errors as one increases the number of draws and what happens to the approximation errors as one increases the number of evaluation points. One would hope that simulation noise always decreases as the number of draws increases; that is, it would be comforting to those doing simulation to know that if they re-estimate their model with more draws, there will be less simulation error in the parameter estimates. Unfortunately, this is not always the case as we demonstrate below.

Those using quadrature would hope that re-estimating their model with more evaluation points will always decrease the approximation errors in the parameter estimates. This is the case.

To investigate, we define the bias in parameter β as $Bias_\beta = \left|\hat{\beta} - \beta\right|$ where β is our stable estimate of β (obtained with either 1,000 draws or 10 evaluation points). To make this measure of bias comparable across parameters (by accounting for the relative flatness of the likelihood-function), we divide each by the standard error, $s.e.$, of the stable β, $\left|\hat{\beta} - \beta\right| / (s.e._\beta)$. Averaging these over all of the parameters provides one possible measure of aggregate bias, denoted $Bias$. By definition, this measure of aggregate bias is, in our example, zero for simulation with 1,000 draws and quadrature with 10 evaluation points.

Note that this measure of aggregate bias, $Bias$, will change every time one reruns the simulation program, even if one takes the same number of draws in each run - the draws are random. There is no comparable variation in $Bias$ when the the parameters are estimated with quadrature - $Bias$ does not vary across runs when the number of evaluation points is held constant because the parameter estimates are always the same.

Our measure of aggregate bias is plotted in Figure 1 for the parameter estimates obtained with 100, 300, 500 and 1000 draws. This example plot proves,

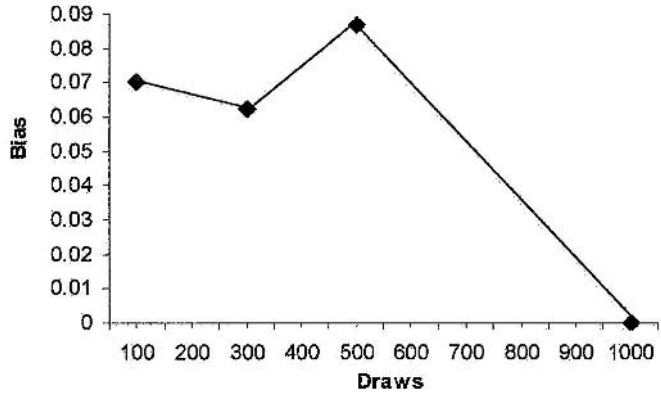


Figure 17.1. Example simulation bias with 100, 300, and 500 draws.

by example, that $Bias$ does not always decrease as the number of random draws increases (the biggest "error" for these three sets of estimates is with 500 draws). Every time one re-estimates the model with 100, 300, 500 draws Figure 17.1 will change, and while one would expect it to usually be monotonically decreasing, it does not have to be, as our example demonstrates.[12]

Figure 17.2 plots $Bias$ for 3, 6 and 9 and 10 evaluation points and the plot is invariant to reruns. Aggregate bias is monotonically decreasing in the number of evaluation points, as it must when the integral evaluated is of the form $\int_{-\infty}^{+\infty} e^{-v^2} h(v)dv$. Note the big drop from 9 to 10 evaluation points.

What should be concluded? Use enough pseudo-random draws or evaluation points to get stable parameter estimates. Determining whether one's parameter estimates are stable is more difficult with simulation than with quadrature. With quadrature, just keep increasing the number of evaluation points until the estimates with e evaluation points and $e+1$ evaluation points differ by less than some predetermined percentage. With simulation, increase the number of draws until the parameter estimates differ by less than that predeter-

[12]Consider another measure of bias, $\left|ave\hat{\beta} - \beta\right|$ where $ave\hat{\beta} = ave[\hat{\beta}^1 + ... + \hat{\beta}^R]$ and $\hat{\beta}^r$ is the estimate of β obtained the r^{th} time the parameters are simulated holding constant the number of draws (Hess et al., 2004, Sandor and Train, 2004). $\left|\hat{\beta} - \beta\right| = \left|ave\hat{\beta} - \beta\right|$ for quadrature but not for simulation; for simulation there is less randomness in $\left|ave\hat{\beta} - \beta\right|$ than in $\left|\hat{\beta} - \beta\right|$. Therefore, $\left|ave\hat{\beta} - \beta\right|$ is likely to be monotonically decreasing in the number of draws, particularly for large R. However, there is no great practical significance to this likelihood unless the researcher plans to run the simulation program multiple times with each number of random draws.

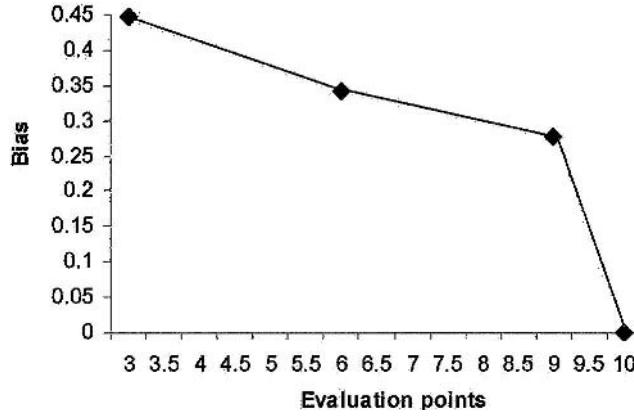


Figure 17.2. Quadrature bias with 3, 6, and 9 evaluation points.

mined percentage, then increase the number of draws some more to make sure the stability was not a random fluke.

A body of literature is now emerging on simulation with non-random draws; that is, taking systematic draws from the density of interest. Halton draws are one example, and a few applications with environmental implications are available.[13] These methods have the potential to significantly reduce simulation noise and the required number of draws. Train, 2003 devotes much of his Chapter 9 to a discussion of simulating with systematic draws. After estimating the same model twice, first with 1,000 random draws and then with 100 Halton draws, Train summarizes:

> These results show the value of Halton draws. Computer time can be reduced by a factor of ten by using Halton draws instead of random draws, without reducing, and in fact increasing, accuracy. These results need to be viewed with caution, however. The use of Halton draws and other quasi-random numbers in simulation-based estimation is fairly new and not completely understood.

Future research might compare quadrature with simulation using systematic methods of drawing parameter vectors.

In closing, consider a direct mathematical comparison of the simulation formula and quadrature formula for P_i. Consider the case of one random param-

[13] See Goett et al., 2000b, Hess et al., 2004, and Sandor and Train, 2004.

eter. With simulation

$$P_i \approx \frac{1}{D}\sum_{d=1}^{D} h(u^d) \tag{6.1}$$

and for quadrature

$$P_i \approx \frac{1}{\sqrt{\pi}}\sum_{m=1}^{M} w_m h(v_m) \tag{6.2}$$

where $h(\omega) = \prod_{j=1}^{J} \Phi\left[-(\beta + \begin{matrix}\omega \\ \mathbf{0}\end{matrix})'(\mathbf{x}_{ij}^{3-k} - \mathbf{x}_{ij}^{k})\right]$. It is evaluated at $\omega_m = v_m\sigma\sqrt{2}$ for quadrature and at $\omega_d = u^d$ with simulation, quadrature carefully selecting the evaluation points, simulation randomly selecting the evaluation points. One should not be surprised that when the draws are carefully selected (e.g., Halton draws), simulation does better than when they are randomly selected. The other big difference is the weights: simulation uses equal weights independent of of how the draws are made; quadrature uses carefully selected weights, each specific to the point at which the function is evaluated.

Table 17.1. Parameter Estimates and Computing Time.

Method	*Gaussian Hermite quadrature*			*Simulation*	
Evaluation points/random draws	3	6	9	100	500
Utility parameters					
β_c	–0.637 (–11.638)	–.649 (–11.487)	–0.645 (–11.607)	–0.639 (–11.439)	–0.648 (–11.607)
β_{PCA}	–0.288 (–3.961)	–0.348 (–4.499)	–0.327 (–4.916)	–0.329 (–4.671)	–0.324 (–4.513)
β_p	1.0 (fixed)	1.0 (fixed)	1.0 (fixed)	1.0 (fixed)	1.0 (fixed)
β_t	0.0470 (6.332)	0.0477 (6.248)	0.0480 (6.384)	0.0488 (6.341)	0.0478 (6.348)
β_w	0.0636 (7.855)	0.0643 (7.842)	0.0647 (7.985)	0.0661 (7.949)	0.0650 (7.989)
β_b	0.0534 (7.220)	0.0554 (7.285)	0.0544 (7.295)	0.0561 (7.346)	0.0544 (7.306)
β_{FCA2}	1.0 (fixed)	1.0 (fixed)	1.0 (fixed)	1.0 (fixed)	1.0 (fixed)
β_{FCA3}	1.616 (4.937)	1.610 (5.941)	1.618 (6.224)	1.621 (6.010)	1.643 (5.774)
β_{FCA4}	2.232 (5.003)	2.129 (6.094)	2.189 (6.519)	2.172 (6.179)	2.215 (5.938)
β_{FCA5}	3.080 (4.743)	2.856 (5.713)	2.963 (6.151)	2.916 (5.847)	3.000 (5.608)
β_{FCA6}	2.499 (4.602)	2.393 (5.551)	2.463 (5.944)	2.450 (5.654)	2.503 (5.437)
β_{FCA7}	3.633 (4.506)	3.421 (5.414)	3.531 (5.857)	3.494 (5.545)	3.578 (5.326)
β_{FCA8}	4.980 (4.339)	4.622 (5.160)	4.813 (5.607)	4.741 (5.296)	4.881 (5.098)
β_{FCA9}	5.471 (4.324)	5.091 (5.087)	5.300 (5.526)	5.209 (5.226)	5.384 (5.035)
β_y	0.0545 (15.322)	0.0555 (15.206)	0.0555 (15.267)	0.0556 (15.282)	0.0556 (15.282)
Standard deviations[a]					
σ_c	0.423(–5.397)	0.432(–5.131)	0.428(–5.270)	0.417(–5.334)	0.431(–5.322)
σ_{FCA}	0.247(–5.232)	0.342(–4.478)	0.302(–5.638)	0.317(–5.099)	0.296(–5.238)
Run time (hrs.) on the same computer	0.43	1.14	1.97	1.13	5.56

[a] t-statistics are for the natural logarithms of the standard deviations.
The parameters were exponentiated in estimation to restrict them to be positive.

Chapter 18

SIMULATION NOISE AND THE ESTIMATION OF LAND USE DECISIONS IN KENYA

John McPeak
Departments of Public Administration and Economics.
Syracuse University.
jomcpeak@maxwell.syr.edu

Abstract This study investigates issues surrounding the nature and importance of simulation noise when using maximum simulated likelihood methods in bivariate tobit estimation of panel data. The application presented considers land use decisions made by nomadic herders in northern Kenya. The study focuses on issues of parameter instability arising from the use of simulation methods to control for an unobserved household specific effect. It is found that parameters are more stable across estimation runs for variables for which there is a higher degree of within household variability and when the parameter is estimated with a higher degree of precision in the initial run. The study also finds that there is less variability in simulating estimation results when different draws are used to simulate results of a given estimation run than when results from different estimation runs generated by using different draws are used for simulation. It is also found that simulation noise does not have a large impact on a main policy finding of the estimation and simulation: reducing risk of accessing remote grazing areas can improve the spatial distribution of grazing pressure and thus address localized degradation and a failure to provide security can lead to environmental degradation.

Keywords: Bivariate Tobit, Maximum Simulated Likelihood, Simulation Noise, Pastoralism, Grazing Land Degradation.

1. Introduction

In a recently published study, I used full information maximum simulated likelihood to estimate a bivariate tobit model of land use decisions made by nomadic herders in northern Kenya (McPeak 2003). This earlier study developed a model to investigate how herder and rangeland heterogeneity could be used to explain localized degradation in a commons, and illustrated how policy measures based on simple common property models may have unintended consequences when degradation is localized rather than widespread. One motivation for the previous study was the observed pattern of rangeland degradation in northern Kenya. Areas around towns in northern Kenya are overused and showing signs of degradation, while vast areas distant from towns are underused and show no sign of degradation. In addition, the data that is available suggests that there was no time in recorded history when the study area's ecological carrying capacity was exceeded by aggregate stocking levels. The study set out to investigate how degradation of sub-areas within a commons can occur, and what policies can be identified to address such degradation.

To simplify analysis in the previous study, information from only one of two areas in which I gathered data was analyzed. The reasoning was that land use decisions could differ parametrically due to differences between the areas in which data were gathered, making pooling of the data set questionable.[1] In addition, the way in which the land use variable of satellite camp use was recorded differed between the two areas due to differences in the way satellite camps are used.[2] Satellite camps are temporary settlements located in remote grazing areas, and will be explained in more detail below.

This paper uses the data set from the other area in which data was gathered to investigate land use decisions, but places greater emphasis on methodological issues arising from the use of full information maximum simulated likelihood (FIMSL) and simulation of estimation results. To clarify, simulation has two meanings in the methodology of this study. First, simulation means Monte Carlo integration to control for an unobserved household specific random effect in bivariate tobit estimation of panel data. Second, simulation means taking a set of estimation results and calculating expected values of dependent variables

[1] Data was gathered in two areas: Chalbi, used in the previous study, and Dukana-Sabarei, used in this study. The areas differ in that there are many water points in the Chalbi area, but few in the Dukana Sabarei area. In addition, the Chalbi area is lower in altitude, lower in rainfall, has larger permanent settlements, and is better connected to other areas in Kenya (roughly speaking, Chalbi has one truck a day headed to the capital of the District, while in Dukana there is one truck a week headed to the District capital).

[2] In the Chalbi area, the satellite camp variable recorded the size of the household labor force sent to satellite camp. In the Dukana area, the satellite camp variable recorded the size of the household herd sent to satellite camp. In Chalbi, satellite camps tends to rely on household labor, while in Dukana there appears to be more labor sharing across households in satellite camps.

to assess the quantitative importance of changes in given variables of policy interest.

When conducting the earlier study I noted two main issues arising from use of FIMSL to generate parameters to use in simulating results of policy interest. First, it was clear that there was variation in parameter estimates across different estimation runs that relied on different pseudo-random draws to represent the random effect term, and that the magnitude of this variation differed for different parameters.[3] Second, different pseudo-random draws to represent the random effect term in simulating estimation results using a given set of parameter estimates led to different simulation results.

This study is an attempt to follow up on these issues in detail. I first investigate the degree of parameter change across estimation runs. I then illustrate there is some degree of predictability to this variation. I then turn to the issue of simulation of results. I focus on a particular variable influencing land use decisions that has clear policy implications—insecurity. One explanation for the spatial pattern of rangeland use is that underused areas are not utilized because these areas are insecure. Results are simulated by changing the value of a dummy variable recording whether an armed raid took place anywhere in the rangelands used by herders in this area in the period or not. First, results are simulated for the parameter estimates obtained from the different estimation runs. Second, results are simulated for a given parameter vector using multiple draws to account for the random effect. These results are then compared.

The overall objectives of this study are twofold. First, from a methodological standpoint, as the use of FIMSL methods becomes increasingly common, it is important to understand both the nature and importance of simulation noise. While generalizing the results of a single empirical study such as this one must be viewed with caution, as the results may be specific to this data set and application, the findings do present information which can be compared and contrasted with future applications to build towards a generalized understanding of the nature and importance of simulation noise in this estimation setting. Second, the study is intended to investigate a question of policy relevance—can we empirically identify the magnitude of the impact insecurity has on land use patterns in a common property rangeland? While it seems quite intuitive to suggest that it has an impact, how large is this impact?

[3] One option available when there is parameter instability across runs is to continue increasing the number of pseudo-random draws used in the estimation procedure to reduce overall variability. A different question that could be investigated using this data is how parameter variation decreases as the number of draws increases. However, as will be discussed below, I became intrigued by the fact that different parameters at a given number of draws behaved quite differently in terms of variability across runs, leading to the question of whether these variations were predictable. Relative differences in variability in parameters would still be an issue no matter how many draws were selected, although it is certainly true that overall variability should decrease.

The outline of the paper is as follows. Section two describes the context of the study and the data used in the estimation. Section three discusses estimation issues and presents empirical results. Section four presents simulation results, and section five concludes.

2. Empirical Analysis of Land-Use Decisions

Data for this study were gathered from 49 Gabra nomadic households who graze their animals between the towns of Dukana and Sabarei in Marsabit District, Kenya. Dukana and Sabarei are two towns close to the Ethiopian border in northern Kenya. Gabra are pastoral herders who reside in an extremely arid environment in which cultivation is impossible. Instead, they rely on their herds of camels, cattle, sheep and goats to generate livestock products and cash income.

The data set records household specific information for four time periods (two rainy seasons and two dry seasons[4]) per-year from 1993 to 1997. The longitudinal nature of the data allows empirical exploration of how land-use decisions made by users of a common rangeland change in response to changes in the state of nature as well as changes in the household's own characteristics. The data gathering methodology was retrospective, and the sampling framework was based on a transect. Enumerators walked between towns of the study area, interviewing herders at compounds they encountered along their way.

The focus of the analysis presented in this section is estimation of land use decisions by nomadic herders. The first decision considered is the herder's base camp location. The base camp is defined as the location of the main dwelling of the pastoral household, and is where women and children tend to be permanent residents, and males are resident when not at satellite camps as defined below. Base camps are mobile, with the dwelling being disassembled, loaded onto camels, and moved to a new location every few months. The base camp distance variable used in this study measures the number of hours it takes to walk from the base camp to the nearest town. As degradation in this area tends to take the form of a circle of approximately five hour's walk radius around towns (Schwartz *et al.*, 1991), this definition allows us to estimate when a household is in the degraded zone or out of the degraded zone. Also included in the estimation is a variable recording the base camp location in the previous period. Use of the lagged variable reflects the fact that base camp moves may be costly, as such moves require dismantling, loading, and reconstructing the base camp dwelling.

[4]There are two rainy seasons per-year in northern Kenya. What is called the long rains occurs in March-May while the short rains occur in mid-September- mid-December. Two dry seasons of approximately three months length separate the rainy seasons.

A herder's decision to send animals to a satellite camp is viewed as sending animals completely out of the zone of degradation.[5] Satellite camps have little infrastructure besides night enclosures for animals and windscreens behind which herders can shelter and cook. The satellite camp decision is recorded by a variable indicating the share of the household herd sent to a satellite camp. The lagged variable is again used as there are some sunk costs in terms of labor and logistics involved with establishing a satellite camp. Together, the base camp distance from town and the satellite camp labor variables capture decisions that determine the share of a herder's animals located in the zone of degradation in different time periods.

A variable that controls for the influence of raid threats on location decisions is a dummy variable that records periods when raids occurred. This information was gathered in a meeting with community elders, where they were asked to report any period in which a raid occurred in the grazing lands used by people from this area. A particularly violent raid occurred in March 1997 in which over fifty people were killed, and a smaller set of raids occurred in early 1998 in which three people were killed. There are certain areas where raids are more likely to occur than others, but the impact of a raid in any area used by herders is to increase overall anxiety and lead to increased caution in migration and location decisions, leading to an overall movement towards where others are for mutual security - that is a general fall back towards town.

The food aid variable used in the estimation records total maize deliveries recorded at Kalacha and North Horr for each time period.[6] Maize is the main component of food aid in this area. Food aid was available from 1993 until mid-1995 and again in early 1997. Food aid data were not available for the Dukana-Sabarei area, but since they are next in line on the road that goes through Kalacha and North Horr, the information from the latter two sites is used as a proxy for the former two sites.

Rainfall conditions are included as exogenous variables in the estimation procedure. Three variables are used to capture these changing conditions. The first records the average of North Horr and Kalacha rainfall in a given six-month period (these are the nearest sites to the Dukana area that had rainfall records available). The second and third are dummy variables indicating

[5] While there may be multiple objectives behind sending animals to a satellite camp (species specific grazing needs, risk diversification, labor issues), for the purpose of the study this is the most relevant aspect. The survey data suggest base camp proximity to town and satellite camp establishment are positively correlated, as the average base camp distance from town for observations when there is only a base camp and no satellite camp is just under 11 hours, compared to just under 5 hours for observations where a satellite camp and a base camp was used. Based on my field experience, satellite camps are only established in areas distant from towns.

[6] Data were obtained by the author at the Catholic mission in North Horr and the A.I.C. mission in Kalacha.

whether the three month period in question is either of the annual rainy seasons.

Household characteristics are also included as exogenous variables in the estimation procedure. A quadratic representation of the age of the household head is utilized, as is a measure recording the size of the household in adult equivalents.[7] Because ownership of baggage camels may reduce the cost of changing locations, a separate variable records the number of pack camels owned by the household. Also included is a quadratic representation of a household's animal wealth recorded in Total Livestock Units (TLUs).[8] Household specific variables are defined to reflect conditions at the start of the three-month period in question, and are thus exogenous to the land-use decisions within the period. Table 18.1 reports the descriptive statistics of the variables used in the estimation.

Table 18.1. Descriptive statistics for estimation variables.

	Average	*St. dev.*
Distance base camp to town (Hours)	8.17	8.15
Satellite camp (% of household herd)	0.24	0.31
Distance last period (hours)	8.40	8.29
Satellite camp last period (%)	0.24	0.30
Rain in past 6 months (mm)	60.94	44.28
Long rains dummy	0.21	0.41
Short rains dummy	0.26	0.44
Food aid delivered to town (tons)	62.72	87.26
Raid dummy	0.11	0.31
Number of Pack camels	0.49	0.82
Herd size in TLU	18.79	6.86
Age of oldest male (years)	53.23	12.09
Age of oldest female (Years)	36.61	10.03
Age ratio oldest male to oldest female	1.52	0.44
Adult equivalents in household	4.15	0.99

Number of observations = 931. Number of households = 49.
Number of time periods = 19.

[7]The household equivalent scale follows Martin (1985). Males and females older than 15 years old equal 1 household adult equivalent, ages 5-14 equal 0.6 household equivalent, ages 2-5 equal 0.3 household equivalent, and ages below 2 equal 0.1 household equivalent.

[8]Herd size is measured in Tropical Livestock Units (TLUs), where 0.7 camel=1head of cattle=10 smallstock (goats and sheep). This differs slightly from the weighting scheme reported in Schwartz *et al.* (1991) as this source suggests 11 goats = 10 sheep = 1 TLU. The data set records the combined smallstock herd size, so no distinction is made between sheep and goats.

3. FIMSL Estimation of a Bivariate TOBIT

In this section, we develop an econometric model for estimating the land use decisions of how far from town to locate a base camp and the share of the herd that will be sent to a satellite camp. The econometric specification is based on the behavioral model presented in McPeak (2003). Three major issues must be confronted when using the panel data on land use decisions described above. First, both dependent variables are by definition non-negative and have observations at zero (3% for the distance variable, 56% for the satellite camp variable). Second, the two decision variables are obviously related to each other as they are both decisions that place livestock in particular rangeland areas. Third, the data set is longitudinal in nature, introducing the possibility that there are household specific effects. The first two issues are addressed by use of simultaneous tobit estimation methodology following Amemiya (1974) and Maddala (1983). The third is addressed by explicitly controlling for household specific effects. Define d as the distance from town variable, f as the satellite camp variable (f stands for *fora* which is the local term for satellite camp), β as coefficients on exogenous variables, x as exogenous variable matrices, a as time-invariant household specific effects, and u as unobserved terms. The d and f notation is used as a label when on the right hand side in the model and as a dependent variable when on the left hand side of the model. Define the following bivariate tobit model, where i indexes households and t indexes time.

$$\begin{aligned} d_t^i &= \beta d' x d_t^i + a d_t^i + u d_t^i \quad \text{if RHS} > 0 \\ d_t^i &= 0 \quad \text{if RHS} \leq 0 \\ f_t^i &= \beta f' x f_t^i + a f_t^i + u f_t^i \quad \text{if RHS} > 0 \\ f_t^i &= 0 \quad \text{if RHS} \leq 0. \end{aligned} \tag{3.1}$$

The approach taken in this study to control for household specific effects follows Mundlak (1978) by defining a household specific effect for the base camp decision by $ad^i = \lambda d' \overline{x}^i + \omega d^i$ and for satellite camp decision by $af^i = \lambda f' \overline{x}^i + \omega f^i$. Let $\overline{x}_i$ record the mean of household specific variables for household i, $t = 1, \ldots, T$. In the estimation procedure, household specific means are included for the age of the household head, the age ratio of the husband or eldest son to his wife or mother, the size of the household herd, and the size of the household in adult equivalents. The random effects represented by the parameters are dealt with following the methodology outlined by Gourieroux and Monfort (1993). Assume that ωd^i is drawn from a $N(0, \sigma^2_{\omega d})$ distribution and ωf^i is drawn from a $N(0, \sigma^2_{\omega f})$ distribution. Take (n, H) pseudo-random draws from two separate $N(0, 1)$ distributions, and assign all $t = 1, \ldots, T$ observations for household $i = 1, \ldots, n$ in draw $h = 1, \ldots, H$ a unique pair of these draws. The $2 \times (n, H)$ draws are multiplied by a (2×1)

scaling parameter δ that is estimated. The parameter $(\delta_d)^2$ provides an estimate of $\sigma^2_{\omega d}$, and $(\delta_f)^2$ provides an estimate of $\omega^2_{\omega f}$. Gourieroux and Monfort (1993) state that provided n and H go to infinity in such a way that $\frac{\sqrt{n}}{H} \rightarrow 0$, the parameters resulting from this estimation are consistent and asymptotically efficient. In the results presented below, $n = 49$, $H = 500$.[9]

State the log-likelihood function for this estimation by defining four different regimes (r) as follows:

- $r1 = 1$ if $d^i_t > 0$, $f^i_t > 0$, 0 otherwise;
- $r2 = 1$ if $d^i_t > 0$,$f^i_t = 0$, 0 otherwise;
- $r3 = 1$ if $d^i_t = 0$,$f^i_t > 0$, 0 otherwise;
- and $r4 = 1$ if $d^i_t = 0$,$f^i_t = 0$, 0 otherwise.

Assume that ud^i_t and uf^i_t when normalized by their standard deviations σ_d and σ_f respectively are distributed bivariate standard normal with a correlation coefficient defined by $\rho = \frac{\sigma_{df}}{\sigma_d \cdot \sigma_f}$. Denote the probability density function for the standard normal density by $\phi(.)$, the probability density function for the standard bivariate normal by $\phi(., \rho)$ the cumulative distribution function for the standard normal distribution by $\Phi(.)$, and the cumulative distribution function for the standard bivariate normal by $\Phi(., \rho)$. Recognize that simulation methods are being used to control for the random effects ω by introducing $red^i(h)$and $ref^i(h)$ to represent the h^{th} pseudo-random draw from a standard normal for household i for the distance estimation and the satellite camp estimation respectively and rearranging equation 3.1 to the following.

$$\begin{aligned} ud^i_t(h) &= d^i_t - \beta d' x d^i_t - \lambda d' \overline{x}^i - \delta_d \cdot red^i(h), \\ uf^i_t(h) &= f^i_t - \beta f' x d^i_t - \lambda f' \overline{x}^f - \delta_f \cdot ref^i(h), \end{aligned} \tag{3.2}$$

I suppress the h notation in what follows, but keep in mind that the likelihood function is being expressed conditional upon a particular set of pseudo-random draws for each of the 49 households. Next, take the equations in 3.2 above and normalize them to allow use of the standard normal distribution functions.

$$\begin{aligned} udn^i_t &= \frac{ud^i_t}{\sigma_d} \\ ufn^i_t &= \frac{uf^i_t}{\sigma_f} \end{aligned} \tag{3.3}$$

[9]The use of 500 draws matches the number of draws in McPeak (2003). This number was chosen for three main reasons. First, it appeared in the earlier study to be a level at which parameters and the log-likelihood did not change very much across runs. Second, it was within the range of draws found in the literature which use this technique (draws tend to range between 500 and 1000). Third, since this study involved repeated estimation and each run took about two days on a desktop computer at 500 draws (and computation time increases in the number of draws used), practical considerations of run time led me to choose 500 draws.

Then make use of the properties of the standard bivariate normal distribution to express one unobserved term conditional on another to allow regimes two and three to be expressed as the product of ϕ and Φ as described by Pudney (1989) and Cornick *et al.* (1994).

$$ufd_t^i = \frac{\left(uf_t^i - u_t^i \cdot \left(\frac{\sigma_{df}}{\sigma_d^2}\right)\right)}{\sqrt{\sigma_f^2 - \frac{(\sigma_{df}^2)}{\sigma_d^2}}} \tag{3.4}$$

$$udf_t^i = \frac{\left(ud_t^i - uf_t^i \cdot \left(\frac{\sigma_{df}}{\sigma_f^2}\right)\right)}{\sqrt{\sigma_d^2 - \frac{(\sigma_{df}^2)}{\sigma_f^2}}} \tag{3.5}$$

This allows us to state the contribution of a particular observation to the likelihood function in the following form:

$$\begin{aligned} L_t^i = {} & r1_t^i \cdot \phi(udn_t^i, ufn_t^i, \rho) + r2_t^i \cdot \left(\frac{1}{\sigma_d}\right) \\ & \cdot\phi(udn_t^i) \cdot \Phi(ufd_t^i) + r3_t^i \cdot \left(\frac{1}{\sigma_f}\right) \\ & \cdot\phi(ufn_t^i) \cdot \Phi(udf_t^i) + r4_t^i \cdot \Phi(udn_t^i, udf_t^i, \rho) \end{aligned} \tag{3.6}$$

Reintroduce the h notation to make clear the role of the simulation methods, and express the likelihood function as follows:

$$L = \sum_{i=1}^{N} L_i = \sum_{i=1}^{N} \frac{1}{H} \cdot \sum_{h=1}^{H} \prod_{t=1}^{T} L_t^i(h) \tag{3.7}$$

The optimization method used in this study relies on the log of this expression, and is performed on this function using the OPTMUM library in GAUSS. The objective function can be expressed as a transformation of equation 3.7:

$$\ln(L) = \sum_{i=1}^{N} \ln(L_i) = \sum_{i=1}^{N} \ln\left(\frac{1}{H} \cdot \sum_{h=1}^{H} \prod_{t=1}^{T} L_t^i(h)\right) \tag{3.8}$$

Estimation was conducted in turn for eleven separate pseudo-random draws with H =500. The first three columns of tables 18.2 and 18.3 report the standard output from the first run; the coefficient estimates, the standard errors of

the estimate, and the corresponding t-ratio.[10] Column four reports the average coefficient estimate for the ten different runs that followed the initial run, and column five reports the coefficient of variation for the coefficient estimates across the ten following runs. The first run is chosen as the standard of comparison for no reason other than it was the first run.

Table 18.2. Results for the FIMSL estimation (Distances).

Variables	*Coeff. run 1*	*Stan. error run 1*	*t-stat. run 1*	*Coeff. average 10 runs*	*Coeff. of var. 10 runs*
DISTANCE($\times 10^{-1}$)					
Constant	1.2122	0.8484	1.4287	0.7909	0.2650
Lagged distance($\times 10^{-1}$)	0.4610	0.0333	13.8560	0.4637	0.0042
Herd size in TLU($\times 10^{-2}$)	0.7384	1.7703	0.4171	0.6477	0.1339
Herd size^2 in TLU($\times 10^{-3}$)	–0.2314	0.3390	–0.6825	–0.2049	–0.1031
Age male($\times 10^{-1}$)	–1.2404	2.0149	–0.6156	–1.1411	–0.0554
Age male^2 ($\times 10^{-3}$)	0.4517	0.3347	1.3496	0.3473	0.1892
Household size in AE($\times 10^{-1}$)	1.5705	0.8289	1.8948	1.5572	0.0069
Food aid ($\times 10^{-2}$)	0.0275	0.0679	0.4050	0.0276	0.0034
Raid dummy	–0.1534	0.1097	–1.3988	–0.1533	–0.0015
♯ of pack camels	–0.0326	0.0567	–0.5745	–0.0392	–0.3151
Rainfall in six months($\times 10^{-2}$)	–0.1987	0.1648	–1.2059	–0.1982	–0.0044
Rainfall in six months^2($\times 10^{-4}$)	0.1259	0.0964	1.3064	0.1258	0.0056
Long rains dummy	–0.0326	0.1097	–0.2970	–0.0322	–0.0081
Short rains dummy	–0.0668	0.0468	–1.4277	–0.0668	-0.0005
Time trend($\times 10^{-1}$)	0.3867	0.6882	0.5619	0.3888	0.0042
Time trend^2($\times 10^{-2}$)	–0.1243	0.1339	–0.9289	–0.1241	–0.0036
Average herd size($\times 10^{-1}$)	0.0354	0.1115	0.3171	0.0522	0.4158
Average age male($\times 10^{-1}$)	0.7541	1.7782	0.4241	0.7689	0.0084
Average age ratio	0.0062	0.1253	0.0496	0.0150	1.7091
Average household size($\times 10^{-1}$)	–1.0791	0.9931	–1.0866	–0.8737	–0.1754
Scaling parameter RE	–0.2733	0.0403	–6.7860	–0.2464	–0.0714
Sigma	0.5488	0.0132	41.5940	0.5494	0.0007

I calculated t-statistics for parameter difference for the ten different runs compared to the results of the first run. One variable (the average age ratio between the husband and wife in the satellite camp estimation) generated t-statistics that indicated significant difference at a 10% level or less in five out

[10]For the quadratic expressions, the joint test of parameter significance generated the following Wald $\chi^2_{(2)}$ statistics. Herd size in the distance equation, 0.9. Herd size in the satellite camp equation, 0.1. Age in the distance equation, 1.9. Age in the satellite equation, 0.8.

of ten runs. All other variables in all other runs generate results indicating the parameters are not significantly different from the first run results at any standard level of statistical significance. The average of the absolute value of the t-statistic for parameter equality across runs was between 0.001 and 0.01 for 34% of observations, 0.01 and 0.1 for 36% of observations, 0.1 and 1.0 for 27% of observations, and above 1.0 for 2% of observations (the 1 variable noted above, which had an average absolute value of the t-statistic of 1.45). From the standpoint of statistical significance, I can conclude that there is no significant difference in results across parameter runs (with the possible exception of the one troublesome variable).

Table 18.3. Results for the FIMSL estimation (Satellite Camp).

Variables	*Coeff. run 1*	*Stan. error run 1*	*t-stat. run 1*	*Coeff. average 10 runs*	*Coeff. of var. 10 runs*
Satellite Camp %					
Constant	0.1419	2.3575	0.0602	0.1115	2.7185
Lagged distance($\times 10^{-1}$)	0.7335	0.0710	10.3280	0.7312	0.0073
Herd size in TLU($\times 10^{-2}$)	0.5035	3.6253	0.1389	0.2329	0.6060
Herd size^2 in TLU($\times 10^{-3}$)	–0.0975	0.9145	–0.1066	–0.0289	–1.2455
Age male($\times 10^{-1}$)	–0.5476	0.6215	–0.8811	–0.4997	–0.1905
Age male^2 ($\times 10^{-3}$)	0.0607	0.8207	0.0740	0.0141	7.0879
Household size in AE($\times 10^{-1}$)	0.1128	0.5457	0.2068	0.1259	0.1563
Food aid ($\times 10^{-2}$)	0.0625	0.0373	1.6768	0.0622	0.0059
Raid dummy	–0.1123	0.0652	–1.7228	–0.1116	–0.0040
Rainfall in six months($\times 10^{-2}$)	–0.1552	0.0932	–1.6643	–0.1553	–0.0074
Rainfall in six months^2($\times 10^{-4}$)	0.0883	0.0552	1.5992	0.0878	0.0053
Long rains dummy	–0.0067	0.0878	–0.0763	–0.0064	–0.0416
Short rains dummy	–0.0606	0.0251	–2.4147	–0.0605	–0.0027
Time trend($\times 10^{-1}$)	0.4886	0.3498	1.3967	0.4876	0.0063
Time trend^2($\times 10^{-2}$)	–0.1182	0.0680	–1.7369	–0.1176	–0.0085
Average herd size($\times 10^{-1}$)	–0.1867	0.1702	–1.0968	–0.1513	–0.4067
Average age male($\times 10^{-1}$)	0.4470	0.9456	0.4728	0.4628	0.0468
Average age ratio	0.2167	0.0823	2.6333	0.1191	0.7066
Average household size($\times 10^{-1}$)	–1.0218	1.7321	–0.5900	–1.2581	–0.2623
Scaling parameter RE	–0.3714	0.1218	–3.0482	–0.3654	–0.0629
Sigma	0.2171	0.0084	25.8870	0.2170	0.0014
Sigma 12	–0.0021	0.0048	–0.4414	–0.0021	–0.0309

However, there is still variability in parameter estimates across runs. The mean coefficient of variation for the parameter estimates across runs is 0.46, and the median is 0.04. The percent of parameters that have a coefficient of

variation in the range: below 0.001, 5%; from 0.001 to 0.01, 41%; from 0.01 to 0.1, 18%; from 0.1 to 1, 27%; from 1 to 10, 9%. Clearly there are a few high values that pull the mean far above the median. What variable characteristics could be influencing this variation? One possibility is that colinearity across variables could be playing a role. In fact, a few of the variables were intentionally defined to be highly colinear to see if this could be an issue (including ages, average age, and a time trend leads to a highly colinear variables). I regressed each variable on the other variables in the regressor matrix in turn to obtain an R^2 measure and use this derived result as a measure of colinearity.

A second possibility was that the degree of variation in the variable could influence the stability of the parameters across runs. I calculated two measures of parameter variation: first, the overall coefficient of variation for the variable; second, the average of the household level coefficient of variation for this variable (absolute values were used for both). The first reflects overall variation in the data set, while the second reflects how much of this variation is within household units.

A third possibility was that scaling issues influenced parameter stability, so I used the absolute values of the variable mean and the absolute value of the parameter estimate obtained from the first run of the estimation. A final possibility was that the precision of the estimate as reflected in the t-value of the first run influenced parameter stability across runs.[11]

These six regressors were used in an estimation of the absolute value of the coefficient of variation for the parameter estimates across runs. A log transform of the dependent variable was adopted due to the pattern of the dependent variable reported above. Table 18.4 reports the results of this regression.

Overall, the adjusted R^2 is 0.53 and $n = 44$ as that is the total number of coefficients estimated in tables 18.2 and 18.3. The main findings from this analysis are as follows. First, the degree of colinearity does not seem to have a significant impact on parameter stability across runs. Second, the degree of variation for the variable as measured by the overall coefficient of variation does not significantly influence parameter stability, but the same can not be said of household level variation. The results suggest that increased within

[11]This list of characteristics to be used in this regression was derived from reflecting on the kinds of things that could matter in such a setting based both on econometric theory and on past experience with optimization in GAUSS. As such, it is admittedly a rather *ad hoc* list of "the usual suspects". I suspected colinearity, since highly colinear variables are difficult to identify with precision according to basic econometric theory. Variation both overall and within the household over time was suspected to matter since greater variation in the regressor allowed independent estimation from the intercept term in the case of overall variation or the household specific effect in the case of household level variation. Scaling issues were suspected since optimization methods in my experience can be impacted by scaling. The initial t-statistic was included since I suspected that the higher confidence with which we can view the precision of an estimate as reflected by the t-statistic in a given run will also be related to the precision with which the variable is estimated when we compare runs.

Table 18.4. Determinants of parameter instability across runs.

	Coeff.	*St. err.*	*t-stat*
Intercept	–1.1950	0.7230	–1.6529
R^2 regression on other regressors	0.3145	0.8037	0.3913
Coefficient of variation overall	0.3761	0.8861	0.4244
Average HH coefficient of variation	–2.4526	0.8090	–3.0316
Absolute value of variable	–0.1818	0.1110	–1.6377
Absolute value of parameter run 1	–1.1752	0.7398	–1.5885
Absolute value of t-stat run 1	–0.1606	0.0396	–4.0595

Note: $\ln(|cv|)$ is dependent variable.

household variation (variation over time for a given household) leads to greater parameter stability across runs. The scaling parameters are close to being significant at the 10% level, and suggest there is some possibility that larger values are more stable. Finally, it is clearly indicated that the greater the degree of precision with which a parameter is estimated in an initial run as measured by the t-ratio, the greater the stability of this parameter across runs.

4. Simulation of Estimation Results

While we may be reasonably confident that the variation across parameter runs is not significant in a statistical sense, it is reasonable to ask whether the variation we still observe makes a qualitative difference when interpreting estimation results. This is a particularly important issue to consider in the context of non-linear estimation, as simulation of estimation results relies on use of the entire parameter vector, rather than a single parameter. Does variability in parameter estimates translate into variation in marginal effects that can lead one to draw incorrect policy conclusions based on simulation noise?

One issue of interest in practical terms is the role of insecurity in land use decisions. As illustrated by the model presented in McPeak (2003), differential riskiness of rangeland sub-area use may lead to localized degradation in a commons. Simply put, if going to some areas exposes herders to the threat of armed raiders shooting them and stealing their animals, they are less likely to go to these areas. As a general rule, areas away from towns in the study area are less secure than those nearer to town. In addition, the areas around town are showing signs of overuse that indicate the long term productivity of the rangelands is declining. To what extent does the fear of raiding lead to spatial concentration of grazing pressure, and what is the prospect of improved security to reduce localized pressure on rangeland resources?

The joint significance of the raid dummy in the land use estimation has a p-value of 0.06, and the coefficients for both decisions indicate that the raid dummy has a negative sign. This is consistent with the argument that increased fear of raids leads herders to send fewer animals to extensive grazing camps and move closer to town. Quantitatively, how important is this variable on land use decisions? I use this question to explore methodological issues.

To conduct the simulation I begin with the following, with variables defined as above. To simplify, I define $\beta d' x d_t^i + a d_i \equiv \mathbf{D}_t^i$ and $\beta f' x f_t^i + a f_i \equiv \mathbf{F}_t^i$. By the law of total probability $E[d_t^i, f_t^i]$ can be decomposed into:

$$\begin{aligned} & E\left[d_t^i, f_t^i | u d_t^i > -\mathbf{D}_t^i, u f_t^i > -\mathbf{F}_t^i\right] \times \Pr\left(u d_t^i > -\mathbf{D}_t^i, u f_t^i > -\mathbf{F}_t^i\right) \quad (4.1) \\ & + E\left[d_t^i, f_t^i | u d_t^i > -\mathbf{D}_t^i, u f_t^i < -\mathbf{F}_t^i\right] \times \Pr\left(u d_t^i > -\mathbf{D}_t^i, u f_t^i < -\mathbf{F}_t^i\right) \\ & + E\left[d_t^i, f_t^i | u d_t^i < -\mathbf{D}_t^i, u f_t^i > -\mathbf{F}_t^i\right] \times \Pr\left(u d_t^i < -\mathbf{D}_t^i, u f_t^i > -\mathbf{F}_t^i\right) \\ & + E\left[d_t^i, f_t^i | u d_t^i < -\mathbf{D}_t^i, u f_t^i < -\mathbf{F}_t^i\right] \times \Pr\left(u d_t^i < -\mathbf{D}_t^i, u f_t^i < -\mathbf{F}_t^i\right). \end{aligned}$$

Focus on the distance equation for simplicity, as the satellite camp calculation will follow the same general pattern. Returning to the idea of each of the four sub-components of equation 3.8 as a regime, narrow attention to the first two regimes (as the expected value of distance in the final two regimes is 0).

$$\begin{aligned} E\left[d_t^i\right] = & \left(\mathbf{D}_t^i + E\left[u d_t^i | u d_t^i > -\mathbf{D}_t^i, u f_t^i > -\mathbf{F}_t^i\right]\right) \times \quad (4.2) \\ & \Pr\left(u d_t^i > -\mathbf{D}_t^i, u f_t^i > -\mathbf{F}_t^i\right) + \\ & \left(\mathbf{D}_t^i + E\left[u d_t^i | u d_t^i > -\mathbf{D}_t^i, u f_t^i < -\mathbf{F}_t^i\right]\right) \times \\ & \Pr\left(u d_t^i > -\mathbf{D}_t^i, u f_t^i < -\mathbf{F}_t^i\right). \end{aligned}$$

The probability weights needed to calculate 4.2 can be obtained through manipulation of the following elements:

$$\begin{aligned} \Pr\left[u d_t^i < -\mathbf{D}_t^i\right] &= \Phi\left(\frac{-\mathbf{D}_t^i}{\sigma_d}\right), \quad (4.3) \\ \Pr\left[u d_t^i < -\mathbf{F}_t^i\right] &= \Phi\left(\frac{-\mathbf{F}_t^i}{\sigma_f}\right), \text{and} \\ \Pr\left[u d_t^i < -\mathbf{D}_t^i, u f_t^i < -\mathbf{F}_t^i\right] &= \Phi\left(\frac{-\mathbf{D}_t^i}{\sigma_d}, \frac{-\mathbf{F}_t^i}{\sigma_f}, \rho\right). \end{aligned}$$

The expected value of the unobserved term can be obtained using the formulas and notation from Maddala (1983, p. 368).[12]

$$\begin{aligned} g_t^i = \left(-\mathbf{D}_t^i/\sigma_d\right) \quad , \quad k_t^i = \left(-\mathbf{F}_t^i/\sigma_f\right), \quad (4.4) \\ g_t^{*i} = (1-\rho^2)^{-0.5} \times (g_t^i - \rho k_t^i) \quad , \quad k_t^{*i} = (1-\rho^2)^{-0.5} \times (f_t^i - \rho g_t^i). \end{aligned}$$

[12]Note that I have changed Maddala's original use of h for the first term with g, to avoid confusion with the prior use of h to represent the number of pseudo-random draws.

Then the expression:

$$E[ud_t^i|ud_t^i > g_t^i, uf_t^i > k_t^i] = \left[\frac{\sigma_d\phi(g_t^i)[1-\Phi(k_t^{*i})] + \rho\sigma_f\phi(k_t^i)[1-\Phi(g_t^{*i})]}{\Pr(ud_t^i > g_t^i, uf_t^i > k_t^i)}\right],$$

provides the expectation of the unobserved term in regime one. Turning to regime two, we can use the finding that

$$\begin{aligned} E[ud_t^i|ud_t^i > g_t^i] &= \frac{\sigma_d\phi(g_t^i)}{\Phi(g_t^i)} \qquad (4.5)\\ &= E[ud_t^i|ud_t^i > g_t^i, uf_t^i > k_t^i] \times \Pr(ud_t^i > g_t^i, uf_t^i > k_t^i)\\ &+ E[ud_t^i|ud_t^i > g_t^i, uf_t^i < k_t^i] \times \Pr(ud_t^i > g_t^i, uf_t^i < k_t^i), \end{aligned}$$

to derive the expectation of the unobserved term in regime two by substituting in what is known to solve for $E[ud_t^i|ud_t^i > g_t^i, uf_t^i < k_t^i]$.

Armed with all these components, I can simulate the expected distance a herd settles from town and the expected herd size sent to a satellite camp which is by assumption in an area distant from town. I use the approximation that the area around town is represented by a circle with a radius the size of a five hour walk from town as noted above. This is based on the idea that that is about the limit for feasibility for making a round trip to town in a single day given the effort involved walking in the intense heat of this area and also provides a rough approximation to the area identified as degraded on a map of rangeland condition in the area (Schwartz *et al.* 1991). I simulate estimation results to calculate the aggregate herd size that settles in this zone as the sum of all base camp herd sizes when the base camp is five hours or less from town minus the herd sent by these base camps to distant satellite camps.

Returning to the overall theme of this study, there are two places where the simulation methods used in estimation may influence the simulation of estimation results. First, as discussed at some length in the previous section, the vector of estimated parameters varied across estimation runs conditional upon the particular H =500 pseudo-random draw used. Second, for a given vector of estimated parameters, there is still an issue of how to represent the household specific effect in the simulation, as the household specific effect is partially composed of the random effect that was controlled for by the use of the H =500 pseudo-random draw. How does each component impact the findings when I simulate estimation results?

The simulation exercise calculates stocking pressure in the zone around town. Table 18.5 presents percentage reduction in the herd size in the area around town when the raid dummy is set to zero compared to when the raid dummy is set to 1. The calculation is conducted for each of the 19 time periods covered in the study. The first two columns report on variation in simulation results that are generated by using the different parameter vectors from the

different estimation runs. The third and forth columns report on variation in simulation results resulting from using 10 different draws for the household specific effect for the parameter vector in the initial run.

Table 18.5. Simulation: reduction of stocking pressure.

Time period	*Different runs average*	*Different runs st. deviation*	*Different draws average*	*Different draws st. deviation*
2	87%	0.0199	89%	0.0006
3	86%	0.0501	87%	0.0470
4	92%	0.0418	100%	0.0000
5	67%	0.0588	71%	0.0159
6	91%	0.0481	94%	0.0001
7	78%	0.0719	90%	0.0001
8	96%	0.0281	100%	0.0000
9	56%	0.0630	75%	0.0050
10	87%	0.0833	89%	0.0035
11	88%	0.0787	96%	0.0021
12	99%	0.0244	100%	0.0000
13	70%	0.0539	80%	0.0003
14	60%	0.0497	55%	0.0003
15	71%	0.0459	65%	0.0402
16	84%	0.0782	89%	0.0004
17	68%	0.0667	64%	0.0204
18	75%	0.0511	67%	0.0001
19	43%	0.0434	38%	0.0002
20	89%	0.0499	94%	0.0127
Average	78%	0.0530	81%	0.0078

Note: raid dummy $= 0$ *vs* raid $= 1$.

From a policy standpoint, it is clear that increasing security in a way that lowers the possibility of raids could have a major impact on reducing stocking pressure on overused areas and that an increased insecurity will have a major impact in increasing stocking pressure in overused areas. From a methodological standpoint, it is clear that the variation in simulation outcomes generated by using different estimation results is greater than that that is generated by using a given set of estimation results with different pseudo-random draws. However, neither form of simulation noise is large enough to lead to an incorrect policy conclusion.

5. Discussion and conclusion

When using simulation methods in estimation, there is reason to be concerned that simulation noise could potentially lead the analyst to incorrect conclusions. While simulation noise can be made "small" and statistically insignificant, it is still worth investigating the nature and importance of simulation noise on questions of policy relevance. In this study, the issue of how simulation noise impacts variables differently was analyzed. It was found that parameter variability increased across runs the lower the coefficient of variation in household variables and the lower the t-statistic of the original estimation run. Scaling issues played a minor role, and colinearity between variables played an insignificant role. The results clarify that the type of variation in regressors that is desirable in this context is within household variation rather than overall variation in the variable. They also provide some measure of reassurance, in that simulation is more likely to be conducted on variables for which there is a relatively high t-statistic given conventional practice in the field.

The study also compared the impact of simulation noise generated by using different draws to estimate the parameter vector with the use of different draws to simulate results using a given parameter vector. The first overall conclusion that can be drawn is that in neither case is the simulation noise large enough to lead one astray from the overall policy implication—increasing security by eliminating the possibility of raids has the potential to drastically reduce stocking pressure on overused rangeland sub-areas and increased insecurity has the opposite effect. While there is some ambiguity about how large the impact of insecurity is on stocking levels due to simulation noise, the ambiguity is small relative to the overall impact of changing the value of the variable. It is also clear that the simulation noise across estimation runs is greater than the noise introduced by using different draws to simulate the results of a given run.

The findings of this current study are intended to be of help to other researchers who have concerns about the impact of simulation noise in their estimation procedure. The findings are of course specific to this data set, and other studies using different data sets may lead to different conclusions. However, both by encouraging further development of the underlying statistical theory that helps shape these results and by allowing other researchers to compare the findings of this study to their own work, it is hoped that this study contributes to the effort to identify generalizable findings about the nature and importance of simulation noise.

References

Abay, G. (1999). *Nachfrageabschaetzung Swissmetro: Eine stated-preference Analyse (in German)*. Berichte des Nationalen Forschungsprogrammes 41: Verkehr und Umwelt, Bericht F1. EDMZ, Bern.

Abbé, E. (2003). *Behavioral modeling: Generalized extreme value for choice models*. MSc Thesis. Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne.

Abramowitz, M. and Stegun, I. (1964). *Handbook of Mathematical Functions, with Formulas, Graphs and Math Tables, Applied Math Series 55*. National Bureau of Standards, , U.S. Government Printing Office, Washington DC.

Adamowicz, W., Swait, J., Boxall, P., Louviere, J., and Williams, M. (1997). Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation. *Journal of Environmental Economics and Management*, 32:65–84.

Adamowicz, W. L., Boxall, P., Williams, M., and Louviere, J. (1998). Stated preference approaches for measuring passive use values: Choice experiments and contingent valuation. *American Journal of Agricultural Economics*, 80:64–75.

Afriat, S. N. (1972). Efficiency estimation of production functions. *International Economic Review*, 13:568–98.

Aigner, D., Lovell, K., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier models. *Journal of Econometrics*, 6:21–37.

Aigner, D. J. and Chu, S. F. (1968). On estimating the industry production function. *American Economic Review*, 58:826–39.

Akaike, H. (1973). Information theory and extensions of the maximum likelihood principle. In Petrov, B. N. and Csake, F., editors, *Second international symposium on infromation theory*, pages 267–281, Budapest. Akademiai Kiado.

Alberini, A. (1995). Efficiency vs bias of willingness to pay estimates: bivariate and interval data models. *Journal of Environmental Economics and Management*, 29:169–80.

Alberini, A., Carson, R. T., and Kanninen, B. J. (1997). Modeling response incentive effects in dichotomous choice contingent valuation data. *Land Economics*, 73(3):309–24.

Albert, J.H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88:669–679.

Allenby, G. (1997). An introduction to hierarchical bayesian modeling. Tutorial notes, Advanced Research Techniques Forum, American Marketing Association.

Allenby, G. and Lenk, P. (1994). Modeling household purchase behavior with logistic normal regression. *Journal of the American Statistical Association*, 89:1218–1231.

Allenby, G. and Rossi, P. (1999). Marketing models of consumer heterogeneity. *Journal of Econometrics*, 89:57–78.

Alvarez, A. (2001). Some issues in the estimation of technical efficiency in a fishery. Efficiency paper series 2/2001, Department of Economics, University of Oviedo.

Alvarez, A. and Orea, L. (2001). Different methods of modelling multi-species fisheries using a primal approach. Efficiency paper series 4/2001, Department of Economics, University of Oviedo.

Alvarez, A. and Perez, L. (2000). Estimating technical efficiency in a fishery. paper presented at the North American productivity workshop, Union College.

Alvarez, A., Perez, L., and Schmidt, P. (2003). The relative importance of luck and technical efficiency in a fishery. Efficiency paper series 3/2003, Department of Economics, University of Oviedo.

Amacher, G. and Feather, P. (1997). Testing Producer Perceptions of Jointly Beneficial Best Management Practices for Improved water Quality. *Applied Economics*, 29:153–159.

Amemiya, T. (1974). Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. *Econometrica*, 42:999–1012.

Andersen, J.L. (2002). Using different inputs and outputs to estimate technical efficiency in fisheries - an application to danish seiners in the north sea and skagerrak. Working Paper 10/02, Fodevareekonomy Institut.

Andersen, J.L. (2003). Reasons for technical inefficiency of danish baltic sea trawlers. Working Papers 18/02, Fodevareekonomy Institut.

Anton, W.R.Q., Deltas, G., and Khanna, M. (2004). Incentives for environmental self-regulation and implications for environmental performance. *Journal of Environmental Economics and Management*, forthcoming.

Araña, J.E. and León, C.J. (2005). Flexible mixture distribution modeling of dichotomous choice contingent valuation with heterogenity. Forthcoming, *Journal of Environmental Economics and Management*.

Arora, S. and Cason, T.N. (1995). An experiment in voluntary environmental regulation: Participation in EPA's 33/50 program. *Journal of Environmental Economics and Management*, 28:271–286.

Arrow, K., Solow, R., Portney, P. R., Leamer, E. E., Radner, R., and Schuman, H. (1993). Report of the NOAA panel on contingent valuation. *Federal Register*, 58:4601–4614.

Baerenklau, K. and Provencher, B. (2004). Static Modeling of Dynamic Recreation Behavior: Implications for Prediction and Welfare Estimation. Manuscript, University of California-Riverside.

Baiocchi, G. (2003). Managing econometric projects using perl. *Journal of Applied Econometrics*, 18(3):371–378.

Baiocchi, G. (2004). Using perl for statistics: Data processing and statistical computing. *Journal of Statistical Software*, 11(1):1–75.

Basch, M., Pena-Torres, J., and Vergara, S. (2002). Catch efficiency in the chilean pelagic fishery: does size matter? Working paper I-140, University Alberto Hurtado, Department of Economics.

Bastin, F. (2004). *Trust-Region Algorithms for Nonlinear Stochastic Programming and Mixed Logit Models*. Ph.D. Thesis, University of Namur. Namur, Belgium.

Bateman, I. J., Langford, I. H., Jones, A. P., and Kerr, G. N. (2001). Bound and path effects in double bounded and triple bounded dichotomous choice contingent valuation. *Resource and Energy Economics*, 23:191–213.

Bateman, I. J., Munro, A., Rhodes, B., Starmer, C., and Sugden, R. (1997). A test of the theory of reference-dependent preferences. *The Quarterly Journal of Economics*, 112:479–505.

Batie, S. (1999). Green Payments as Foreshadowed by EQIP. Staff paper 99-45, Department of Agricultural Economics.

Battese, G. and Coelli, T. (1992). Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. *Journal of Productivity Analysis*, 3:153–169.

Battese, G. and Coelli, T. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, 20:325–332.

Beenstock, M., Goldin, E., and Haitovsky, Y. (1998). Response bias in a conjoint analysis of power outages. *Energy Economics*, 20:135–156.

Bellman, R. E. (1955). Functional Equations in the Theory of Dynamic Programming: Positivity and Quasilinearity. *Proceedings of the National Academy of Sciences*, 41:743–46.

Ben-Akiva, M. and Bierlaire, M. (2003). Discrete choice models with applications to departure time and route choice. In Hall, R., editor, *Handbook of Transportation Science, Second Edition*, pages 7–37. Kluwer Academic Publishers, Boston, MA.

Ben-Akiva, M., Bolduc, D., and Walker, J. (2001). Specification, Identification, and Estimation of the Logit Kernel (or Continuous Mixed Logit). Mimeo.

Ben-Akiva, M. and Francois, B. (1983). *Mu-homogenous generalized extreme value model*. Working Paper. Department of Civil Engineering, MIT, Cambridge, MA.

Ben-Akiva, M. and Lerman, S. (1985). *Discrete Choice Analysis : Theory and Application to Travel Demand*. MIT Press, Cambridge, Massahusetts, U.S.A.

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York, NY, U.S.A., 2nd edition.

Berger, J.O. and Perrichi, L. R. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, 91:109–22.

Bertsekas, D. P. (1976). *Dynamic Programming and Stochastic Control*. Academic Press, New York.

Bhat, C. (1998a). Accommodating flexible substitution patterns in multidimensional choice modeling: Formulation and application to travel mode and departure time choice. *Transportation Research B*, 32:455–466.

Bhat, C. (1998b). Accommodating variations in responsiveness to level of service variables in travel mode choice models. *Transportation Research A*, 32:455–507.

Bhat, C. (2000a). Incorporating observed and unobserved heterogeneity in urban work mode choice modeling. *Transportation Science*, 34:228–238.

Bhat, C. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research B*, 35:677–695.

Bhat, C. R. (2000b). Flexible model structures for discrete choice analysis. In Hensher, D.A. and Button, K. J., editors, *Handbook of Transport Modelling, Volume 1*, pages 71–90. Pergamon Press, Oxford.

Bhat, C. R. and Guo, J. (2004). A mixed spatially correlated logit model: Formulation and application to residential choice modeling. *Transportation Research B*, 38(2):147–168.

Bierlaire, M. (2003). BIOGEME: a free package for the estimation of discrete choice models. Proceedings of the 3^{rd} Swiss Transport Research Conference, Monte Verita, Ascona, Switzerland.

Bierlaire, M. (2004). A theoretical analysis of the cross-nested Logit model. *Annals of Operations Research*, forthcoming.

Bierlaire, M., Axhausen, K. W., and Abay, G. (2001). The acceptance of modal innovation : The case of Swissmetro. In *Proceedings of the* 1^{st} *Swiss Transport Research Conference*, Monte Verita, Ascona, Switzerland.

Bishop, R.C. and Heberlein, T.A. (1979). Measuring values of extramarket goods: Are indirect measures biased? *American Journal of Agricultural Economics*, 61:926–930.

Bjorndal, T., Kondouri, P., and Pascoe, S. (2002). Multi-output distance function for the north sea beam trawl fishery. Working Paper, Department of Economics, University of Reading.

Bjørner, T.B., Hansen, L.G., and Russell, C.S. (2004). Environmental labeling and consumers' choice - an empirical analysis of the nordic swan. *Journal of Environmental Economics and Management*, forthcoming.

Blomo, V., Stokes, K., Griffin, W., Grant, W., and Nichols, J. (1978). An Intraseasonal Dynamic Optimization Model to Allocate Irrigation Water between Crops. *Southern Journal of Agricultural Economics*, 10:119–125.

Boatwright, P., McCulloch, R., and Rossi, P. (1999). Account-level modeling for trade promotion: An application of a constrained parameter hierarchical model. *Journal of the American Statistical Association*, 94:1063–1073.

Bockstael, N. (1995). Travel Cost Models. In Bromley, D., editor, *The Handbook of Environmental Economics*, pages 655 – 671. Blackwell, Cambridge.

Bolduc, D. (1999). A practical technique to estimate multinomial probit models in transportation. *Transportation Research B*, 33:63–79.

Bolduc, D., Fortin, B., and Fournier, M. (1996). The impact of incentive policies on the practice location of doctors: A multinomial probit analysis. *Journal of Labor Economics*, 14:703–732.

Borenstein, S. (2001). The trouble with electricity markets (and some solutions). University of California Energy Institute, University of California, Berkeley.

Börsch-Supan, A. and Hajivassiliou, V.A. (1993). Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. *Journal of Econometrics*, 58:347–368.

Börsch-Supan, A., Hajivassiliou, V.A., Kotlikoff, L.J., and Morris, J.N. (1992). Health, children, and elderly living arrangements. A multiperiod-multinomial probit model with unobserved heterogeneity and autocorrelated errors. In Wise, D.A., editor, *Topics in the Economics of Aging*, pages 79–104. University of Chicago, Chicago, U.S.A.

Bos, C. S. (2002). A comparison of marginal likelihood computation methods. In Härdle, W. and Rönz, B., editors, *COMPSTAT 2002 - Proceedings in Computational Statistics*. Springer-Verlag, Amsterdam.

Boston, Kevin and Bettinger, P. (1999). An analysis of Monte Carlo integer programming, simulated annealing, and tabu search heuristics for solving spatial harvest scheduling problems. *Forest Science*, 45(2):292–301.

Bowman, A. W. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: the Kernel Approach with S-Plus Illustrations*. Oxford University Press, Oxford, U.K.

Boxall, P. C., Adamowicz, W. L., Swait, J., Williams, M., and Louviere, J. (1996). A comparison of stated preference methods for environmental valuation. *Ecological Economics*, 18:243–253.

Boyle, K.J., MacDonald, H.F., Cheng, H., and McCollum, D.W. (1998). Bid Design and Yea Saying in Single-Bounded, Dichotomous-Choice Questions. *Land Economics*, 74(1):49–64.

Bradley, M.A. and Daly, A.J. (1997). Estimation of logit models using mixed stated preference and revealed preference information. In Stopher, P.R. and Lee-Gosselin, M., editors, *Understanding Travel Behaviour in an Era of Change*. Pergamon Press, Oxford, U.K.

Brazee, R. and Mendelsohn, R. (1988). Timber harvesting with fluctuating prices. *Forest Science*, 34:359–372.

Breffle, W. S. and Morey, E. R. (2000). Investigating preference heterogeneity in a repeated discrete-choice recreation demand model of atlantic salmon fishing. *Marine Resource Economics*, 15:1–20.

Breffle, W. S. and Rowe, R. D. (2002). Comparing choice question formats for evaluating natural resource tradeoffs. *Land Economics*, 78:298–314.

Broeck, J. van den, Koop, G., Osiewalski, J., and Steel, M.F.J. (1994). Stochastic frontier models: A bayesian perspective. *Journal of Econometrics*, 61:273–303.

Brown, G., Layton, D., and Lazo, J. (1994). Valuing Habitat and Endangered Species. Institute for economic research discussion paper 94-01, University of Washington.

Brownstone, D. (2001). Discrete choice modeling for transportation. In Hensher, D., editor, *Travel Behavior Research: The Leading Edge*, pages 97–124. Elsevier, Oxford, UK.

Brownstone, D., Bunch, D., and Train, K. (2000). Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research B*, 34:315–338.

Brownstone, D. and Train, K. (1999). Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics*, 89:109–129.

Brunnermeier, S.B. and Cohen, M.A. (2003). Determinants of environmental innovation in US manufacturing industries. *Journal of Environmental Economics and Management*, 45:278–293.

Bryant, K.J., Mjelde, J.W., and Lacewell, R.D. (1993). An Intraseasonal Dynamic Optimization Model to Allocate Irrigation Water between Crops. *American Journal of Agricultural Economics*, 75:1021–29.

Buckheit, Jonathan B. and Donoho, David L. (1995). Wavelab and reproducible research. In Antoniadis, Anestis and Oppenheim, Georges, editors, *Wavelets and Statistics*, pages 55–81. Springer-Verlag, Berlin, New York.

Bunch, D.S. (1991). Estimability in the multinomial probit model. *Transportation Research B*, 25:1–12.

Butler, J. and Moffit, R. (1982). A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica*, 50:761–764.

Cameron, T. (1988). A new paradigm for valuing non-market goods using referendum data: Maximum likelihood estimation by censored logistic regression. *Journal of Environmental Economics and Management*, 15:355–379.

Cameron, T. and James, M. (1987). Efficient estimation methods for closed-ended contingent valuation survey data. *Review of Economics and Statistics*, 69:269–276.

Cameron, T. and Quiggin, J. (1994). Estimation using contingent valuation data from a 'dichotomous choice with follow-up' questionnaire. *Journal of Environmental Economics and Management*, 27:218–234.

Cantillo, V. and Ortúzar, J. de D. (2004). Modelling discrete choices in the presence of inertia and serial correlation. Manuscript.

Carlsson, F. and Martinsson, P. (2003). Design techniques for stated preference methods in health economics. *Health Economics*, 12:281–294.

Carson, R. T., Flores, N. E., and Meade, N. F. (2001). Contingent valuation: Controversies and evidence. *Environmental and Resource Economics*, 19:173–210.

Carson, R. T., Groves, T., and Machina, M. J. (1999). Incentive and informational properties of preference questions. Plenary Address to the European Association of Resource and Environmental Economists Conference, Oslo.

Carson, R. T., Wilks, L., and Imber, I. (1994). Valuing preservation of australia's kadaku conservation zone. *Oxford Economic Papers*, 46:727–749.

Casella, G. and George, E. (1992). Explaining the Gibbs sampler. *The American Statistician*, 46:167–174.

Caves, D. W., Herriges, J. A., and Windle, R. J. (1990). Customer demand for service reliability in the electric power industry: A synthesis of the outage cost literature. *Bulletin of Economic Research*, 42:79–119.

Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring efficiency of decision making units. *European Journal of Operational Research*, 2:429–444.

Chen, H. Z. and Cosslett, S. R. (1998). Environmental quality preference and benefit estimation in multinomial probits: A simulation approach. *American Journal of Agricultural Economics*, 80:512–520.

Chen, H. Z. and Randall, A. (1997). Semi-Parametric Estimation of Binary Response Models with an Application to Natural Resource Valuation. *Journal of Econometrics*, 76:323–40.

Chen, M-H., Shao, Q-M., and Ibrahim, J. (2000). *Monte Carlo Methods in Bayesian Computation*. Springer, New York, U.S.A.

Cherchi, E. and Ortúzar, J. de D. (2004). Random effects vs. error components in the mixed logit model: Correlation, heteroskedasticity and heterogeneity. Paper presented at 10th WCTR, Istanbul, Turkey, 4-8 July 2004.

Chib, S. (1992). Bayes Inference in the Tobit Censored Regression Model. *Journal of Econometrics*, 52:79–99.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90:1313–21.

Chib, S. and Carlin, B.P. (1999). On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing*, 9:17–26.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *American Statistician*, 49:327–335.

Chib, S. and Jeliazkov, I. (2001). Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association*, 96:270–281.

Claassen, R., Hansen, L., Peters, M., Breneman, V., Weinberg, M., Cattaneo, A., Feather, P., Gadsby, D., Hellerstein, D., Hopkins, J., Johnston, P., Morehart, M., and Smith, M. (2001). Agri-environmental policy at the crossroads: Guideposts on a changing landscape. Report No.794, United States Department of Agriculture, Economic Research Service Agricultural Economics, Washington, DC, U.S.A.

Claassen, R. and Horan, R. (2000). Environmental payments to farmers: Issues of program design. *Agricultural Outlook*, 272:15–18.

Cleff, T. and Rennings, K. (1999). Determinants of environmental product and process innovation. *European Environment*, 9(5):191–201.

Coelli, T. (1996). A guide to frontier 4.1: a computer program for stochastic frontier production and cost function estimation. Working Paper 96/07, Center for Efficiency and Productivity Analysis, University of New England, Australia.

Coelli, T., Rao, D. S. P., and Battese, G. E. (1998). *An introduction to efficiency and productivity analysis*. Kluwer Academic Publishers, Boston, Dordrecht, London.

Cooper, J. and Keim, R. (1996). Incentive payments to encourage farmer adoption of water quality protection practices. *American Journal of Agricultural Economics*, 78:54–64.

Cooper, J.C. (1993). Optimal bid selection for dichotomous contingent valuation surveys. *Journal of Environmental Economics and Management*, 24:25–40.

Cooper, J.C. (1997). Combining actual and contingent behavior data to model farmer adoption of water quality protection practices. *Journal of Agricultural and Resource Economics*, 22:30–43.

Cooper, J.C. and Hanemann, W.M. (1995). Referendum contingent valuation: How many bounds are enough? Working paper, USDA Economic Research Service, Food and Consumer Economics Division.

Cooper, J.C., Hanemann, W.M., and Signorello, G. (2002). One and one half bound dichotomous choice contingent valuation. *The Review of Economics and Statistics*, 84:742–50.

Cornick, G., Cox, T., and Gould, B. (1994). Fluid milk purchases: a multivariate tobit analysis. *American Journal of Agricultural Economics*, 76:74–82.

Creel, M. and Loomis, J. (1997). Semi-nonparametric distribution-free dichotomous choice contingent valuation. *Journal of Environmental Economics and Management*, 32:341–358.

Crooker, J.R. and Herriges, J.A. (2004). Parametric and semi nonparametric estimation of willingness to pay in the dichotomous choice contingent valuation framework. *Environmental and Resource Economics*, 27(4):451–480.

Daganzo, C. F. (1979). *Multinomial Probit: The Theory and Its Application to Demand Forecasting*. Academic Press, New York.

Daly, A. and Bierlaire, M. (2003). *A general and operational representation of GEV models*. Technical Report RO-030502. Institute of Mathematics, Operations Research Group ROSO, EPFL, Lausanne, Switzerland.

Daly, A. and Zachary, S. (1978). Improved multiple choice models. In Hensher, D. and Dalvi, M., editors, *Determinants of Travel Choice*. Saxon House, Sussex.

Dansie, B.R. (1985). Parameter estimability in the multinomial probit model. *Transportation Research B*, 19(6):526–528.

Dasgupta, S., Hettige, H., and Wheeler, D. (2000). What improves environmental compliance? Evidence from Mexican industry. *Journal of Environmental Economics and Management*, 39:39–66.

Davidson, R. and MacKinnon, J. G. (1993). *Estimation and Inference in Econometrics*. Oxford university Press, Oxford.

DeCanio, S. J. and Watkins, W. E. (1998). Investments in energy efficiency: Do the characteristics of firms matter? *The Review of Economics and Statistics*, 80:95–107.

Deely, J.J. and Lindley, D.V. (1972). Bayes empirical Bayes. *Journal of the American Statistical Association*, 76:833–841.

DeShazo, J. R. (2002). Designing transactions without framing effects in iterative question formats. *Journal of Environmental Economics and Management*, 43:360–85.

DeShazo, J. R. and Fermo, G. (2002). Designing choice sets for stated preference methods: The effects of complexity on choice consistency. *Journal of Environmental Economics and Management*, 44:123–143.

Devroye, D. (1986). *Non-Uniform Random Variate Generation*. Springer-Verlag, New York.

Devroye, L. (1996). *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, NY.

Dewald, W. G., Thursby, J. G., and Anderson, R. G. (1986). Replication in empirical economics: The Journal of Money, Credit and Banking Project. *American Economic Review*, 76(4):587–603.

Doane, M., Hartman, R., and Woo, C.-K. (1988a). Household preference for interruptible rate options and the revealed value of service reliability. *The Energy Journal*, 9:121–134.

Doane, M., Hartman, R., and Woo, C.-K. (1988b). Households' perceived value of service reliability: An analysis of contingent valuation data. *The Energy Journal*, 9:135–151.

Dorfman, J. (1996). Modeling multiple adoption decisions in a joint framework. *American Journal of Agricultural Economics*, 78:547–57.

Dorfman, J. and Koop, G. (2005). Current developments in productivity and efficiency measurement. *Journal of Econometrics*, 26(2):233–240.

Draper, D. (2000). Bayesian hierarchical modelling. Manuscript.

Duffield, J.W. and Patterson, D.A. (1991). Inference and optimal design for a welfare measure in dichotomous choice contingent valuation. *Land Economics*, 67(2):225–239.

DuMouchel, W.H. and Harris, J.E. (1981). Bayes methods for combining the results of cancer studies in humans and other species. *Journal of the American Statistical Association*, 78:293–308.

Ebling, G. and Janz, N. (1999). Export and innovation activities in the German service sector: Empirical evidence at the firm level. Discussion Paper 99-53, ZEW, Mannheim, Germany.

Eggert, H. (2001). Technical efficiency in the swedish trawl fishery for norway lobster. Working Paper in economics 53, Department of Economics, Goteborg University.

Embrechts, P., Lindskog, F., and McNeil, A. J. (2003). Modelling dependence with copulas and applications to risk management. In Rachev, S. T., editor, *Handbook of heavy tailed distributions in finance*. Elsevier/North-Holland, Amsterdam, The Netherlands.

Energy Information Administration (1998). The changing structure of the electric power industry: Selected issues. Technical report, U.S. Department of Energy, Washington, D.C.

Energy Information Administration (2002). Status of State electric industry restructuring. Technical report, U.S. Department of Energy, Washington, D.C.

Evans, M., Hastings, N., and Peacock, B. (1993). *Statistical Distributions*. John Wiley and Sons, New York, U.S.A.

Evans, M., Hastings, N., and Peacock, B. (2000). *Statistical distributions*. John Wiley and Sons, Inc., New York.

Executive Office of the President (1992). Guidelines and discounted rates for benefit-cost analysis of federal programs. Technical report, Office of Management and Budget, Washington, DC.

Farrell, M. J. (1957). The measurement of technical efficiency. *Journal of the Royal Statistical Society*, 120:253–290.

Faruqui, A., Chao, H.-P., Niemeyer, V., Platt, J., and Stahlkopf, K. (2001). Analyzing California's power crisis. *Energy Journal*, 22:29–52.

Feather, P. and Cooper, J. (1996). Strategies for curbing water pollution. *Agricultural Outlook*, 224:19–22.

Feller, W. (1968). *An Introduction to Probability Theory, Vol. 1*. Wiley, New York, third edition.

Felthoven, R. G. (2002). Effects of the american fisheries act on capacity utilization and technical efficiency. *Marine Resource Economics*, 17:181–205.

Fenton, V. and Gallant, A. (1996). Qualitative and asymptotic performance of snp density estimators. *Journal of Econometrics*, 74:77–118.

Fernández, C., Koop, G., and Steel, M.F. (2000). Alternative efficiency measures for multiple-output production. *Journal of Econometrics*, 98:47–49.

Fernández, C., Koop, G., and Steel, M.F. (2002a). Multiple-output production with undesirable outputs: An application to nitrogen surplus in agriculture. *Journal of the American Statistical Association*, 97:432–442.

Fernández, C., Koop, G., and Steel, M.F. (Forthcoming). A Bayesian analysis of multiple output production frontiers. *Journal of Econometrics*.

Fernández, C., León, C.J., Steel, M., and Vázquez-Polo, F. J. (2004). Bayesian analysis of interval data contingent valuation models and pricing policies. *Forthcoming, Journal of Business Economics and Statistics*.

Fernández, C., Ley, E., and Steel, M.F. (2002b). Bayesian modelling of catch in a North-West Atlantic fishery. *Journal of the Royal Statistical Society*, 51(3):257–280.

Fernández, C., Osiewalski, J., and Steel, M.F.J. (1997). On the use of panel data in stochastic frontier models with improper priors. *Journal of Econometrics*, 79:169–193.

Fieller, E. (1932). The distribution of the index in a normal bivariate population. *Biometrika*, 24:428–440.

Fisher, R. A. (1936). Has Mendel's work been rediscovered. *Annals of Science*, 1:115–137.

Ford, M. P. and Ford, D. J. (2001). Investigation of GAUSS' random number generators. Technical report for aptech systems, inc., FORWARD Computing and Control Pty. Ltd., NSW Australia.

Foster, V. and Mourato, S. (2002). Testing for consistency in contingent ranking experiments. *Journal of Environmental Economics and Management*, 44:309–328.

Foster, V. and Mourato, S. (2003). Elicitation format and sensitivity to scope: Do contingent valuation and choice experiments give the same results? *Environmental and Resource Economics*, 24:141–160.

Fried, H., Lovell, C.A.K., and Schmidt, S. (1993). *The Measurement of Productive Efficiency: Techniques and Applications*. Oxford University Press, New York.

Fukami, Mizuki (2004). Interindustry linkage analysis with physical input-output tables. Master's thesis, Environment Department, University of York.

Gallant, A. (1987). Identification and consistency in semi-nonparametric regression. In Bewley, T. F., editor, *Advances in Econometrics*, pages 145–169. Cambridge University Press, New York, U.S.A.

Garrod, G. D., Scarpa, R., and Willis, K. G. (2002). Estimating the benefits of traffic calming on through routes: A choice experiment approach. *Journal of Transport Economics and Policy*, 36:211–231.

Garrod, G. D. and Willis, K. G. (1997). The non-use benefits of enhancing forest biodiversity: A contingent ranking study. *Ecological Economics*, 21:45–61.

Garrod, G.D. and Willis, K.G. (1998). Using contingent ranking to estimate the loss of amenity value for inland waterways from public utility service structures. *Environmental and Resource Economics*, 12:241–247.

Garrow, L. (2004). *Comparison of models representing correlation and random taste variation: An application to airline passengers' rescheduling behavior*. Published doctoral dissertation. Northwestern University, Evanston, Illinois.

Geary, R. (1930). The frequency distribution of the quotient of two normal variates. *Journal of the Royal Statistical Society*, 93:442–446.

Gelfand, A. and Smith, A. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:972–985.

Gelfand, A.E., Hills, S., Racine-Poon, A., and Smith, A.F.M. (1990). Illustration of bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, 85:972–985.

Gelman, A.B., Carlin, J.S., Stern, H.S., and Rubin, D.B. (1995). *Bayesian Data Analysis*. Chapman and Hall/CRC, Boca Raton, Suffolk.

Geman, S. and Geman, D. (1984). Stochastic relaxation Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Genius, M. and Strazzera, E. (2004). The copula approach to sample selection modelling: an application to the recreational value of forests. Nota di Lavoro 73, FEEM.

Gentle, James E (2003). *Random Number Generation and Monte Carlo Methods*. Springer-Verlag, New York, second edition.

Gentleman, Robert (2004). Some perspectives on statistical computing. Technical report, Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA.

George, E.I. (1986). Combining minimax shrinkage estimators. *Journal of the American Statistical Association*, 81:437–445.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Bernardo, J., Berger, J., Dawid, A., and Smith, F., editors, *Bayesian Statistics*, pages 169–193. Oxford University Press, New York.

Geweke, J. (1996). Monte Carlo simulation and numerical integration. In Amman, Hans M., Kendrick, David A., and Rust, John, editors, *Handbook of Computational Economics, Volume 1*, pages 731–800. North-Holland, Amsterdam.

Geweke, J., Keane, M., and Runkle, D. (1994). Alternative computational approaches to inference in the multinomial probit model. *Review of Economics and Statistics*, 76:609–632.

Geweke, J., Keane, M., and Runkle, D. (1997). Statistical inference in the multinomial multiperiod probit model. *Journal of Econometrics*, 80:125–165.

Geweke, J. F. (1999). Using simulation methods for bayesian econometric models: inference, development, and communication (with discussion and reply). *Econometric Reviews*, 18:1–126.

Goett, A., Hudson, K., and Train, K. (2000a). Consumers' choice among retail energy suppliers: The willingnes-to-pay for service attributes. *The Energy Journal*, 21:1–28.

Goett, A., Train, K., and Hudson, K. (2000b). Customer choice among retail energy suppliers: the willingness-to-pay for service attributes. *Energy Journal*, 21:1–28.

Goett, A. A., McFadden, D. L., and Woo, C.-K. (1988). Estimating household value of electrical service reliability with market research data. *Energy Journal*, 9:105–120.

Good, I. J. (1983). *Good Thinking: The Foundations of Probability and Its Applications*. The University of Minnesota Press, Minneapolis, MN, U.S.A.

Good, I.J. (1980). Some history of the Bayesian hierarchical methodology. In Bernardo, J.M., DeGroot, M.H., Lindley, D.V., and Smith, A.F.M., editors, *Bayesian Statistics 2*. North Holland, Amsterdam, Netherlands.

Gottschalk, S. and Janz, N. (2003). Bestimmungsfaktoren der Innovationstätigkeit. In Janz, N. and Licht, G., editors, *Innovationsforschung heute*, pages 41–72. ZEW Wirtschaftsanalysen, Bd. 63, Baden-Baden, Germany.

Gouriéroux, C. and Monfort, A. (1993). Simulation-based inference: A survey with special reference to panel data models. *Journal of Econometrics*, 59:5–33.

Gouriéroux, C. and Monfort, A. (1996). *Simulation-Based Econometric Methods*. Oxford University Press, New York.

Grafton, R. Q., Squires, D., and Fox, K. J. (2000). Private property and economic efficiency: a study of a common-pool resource. *Journal of Law and Economics*, 43:679–713.

Grant, W.E. and Griffin, W.L. (1979). A bioeconomic model for Gulf of Mexico shrimp fishery. *Transactions of the American Fisheries Society*, 108:1–13.

Grant, W.E., Isaakson, K.G., and Griffin, W.L. (1981). A general bioeconomic simulation model for annual-crop marine fisheries. *Ecological Modeling*, 13:195– 219.

Green, D., Jacowitz, K. E., Kahneman, D., and McFadden, D. (1998). Referendum contingent valuation, anchoring, and willingness to pay for public goods. *Resource and Energy Economics*, 20:85–116.

Greene, W. (1997). *Econometric Analysis*. Prentice Hall, third edition.

Greene, W. H., Hensher, D. A., and Rose, J. M. (2005). Using classical simulation based estimators to estimate individual willingness to pay values. In Scarpa, R. and Alberini, A., editors, *Applications of simulation methods in Environmental and Resource Economics*. Springler-Kluwer Academic Press, Boston, MA.

Griffin, J. and Steel, M. F. J. (2004). Semiparametric Bayesian inference for stochastic frontier models. *Journal of Econometrics*, 123:121–152.

Griffin, W.L. and Oliver, C. (1991). Evaluation of the economic impacts of turtle excluder devices (teds) on the shrimp production sector in the Gulf of Mexico. Technical report, Final Report MARFIN Award NA-87-WC-H-06139, Department of Agricultural Economics, College Station, Texas A&M University.

Griffin, W.L. and Stoll, J.R. (1981). Economic issues pertaining to the Gulf of Mexico shrimp management plan. In Anderson, L.G., editor, *Economic Analysis for Fisheries Management Plans*, volume Economic Analysis for Fisheries Management Plans, pages 81 – 111. Ann Arbor Science Publishers, Inc., Ann Arbor, MI.

Griffiths, W.H., Hill, R.C., and Pope, P.J. (1987). Small sample properties of probit models estimators. *Journal of the American Statistical Association*, 82:929–937.

Gulf of Mexico Fishery Management Council (2001). Regulatory amendment to the Reef Fish Fishery management plan to set a Red Snapper rebuilding plan through 2032. Technical report, Gulf of Mexico Fishery Management Council, Tampa, Florida.

Haab, T. C. and McConnell, K. E. (1998). Referendum models and economic values: Theoretical, intuitive, and practical bounds on willingness to pay. *Land Economics*, 74:216–29.

Haab, T.C. (1999). Nonparticipation or misspecification? the impacts of nonparticipation on dichotomous choice contingent valuation. *Environmental and Resource Economics*, 14:443–461.

Haaijer, M. E. (1999). *Modeling Conjoint Choice Experiments with the Probit Model*. PhD thesis, University of Groningen, Labyrint Publications, The Netherlands.

Haaijer, M. E., Kamakura, W.A., and Wedel, M. (2001). The no-choice alternative in conjoint choice. *International Journal of Market Research*, 43(1):93–106.

Haefen, R. H. von and Adamowicz, W. L. (2004). Not playing the game: nonparticipation in repeated discrete choice models.

Hajivassiliou, V.A., McFadden, D., and Ruud, P.A. (1992). Simulation of multivariate normal rectangle probabilities and their derivatives: theoretical and computational results. discussion paper 1021, Cowles Foundation, Yale University.

Hajivassiliou, V.A., McFadden, D., and Ruud, P.A. (1996). Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results. *Journal of Econometrics*, 72:85–134.

Halton, J. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2:84–90.

Hanemann, M., Loomis, J., and Kanninen, B. (1991). Statistical efficiency of double-bounded dichotomous choice contengent valuation. *American Journal of Agricultural Economics*, 73:1255–1263.

Hanemann, W. M. (1978). *A Theoretical and Empirical Study of the Recreation Benefits from Improving Water Quality in the Boston Area*. PhD thesis, Harvard University.

Hanemann, W. M. (1984). Welfare evaluations in contingent valuation experiments with discrete response data. *American Journal of Agricultural Economics*, 66:332–341.

Hanemann, W. M. (1994). Valuing the environment through contingent valuation. *Journal of Economic Perspectives*, 8:19–43.

Hanemann, W. M. and Kanninen, B. (2001). The statistical analysis of discrete-response CV data. In Bateman, I.J. and Wills, K.G., editors, *Valuing Environmental Preferences*. Oxford University Press.

Hannesson, R. (1983). Bioeconomic production function in fisheries: theoretical and empirical analysis. *Canadian Journal of Aquatic Sciences*, 40:968–982.

Harabi, N. (1997). Determinanten des technischen Fortschritts auf Branchenebene: Ein Überblick. Discussion Paper 97-02, ZEW, Mannheim, Germany.

Harrison, G. W. and Kriström, B. (1995). On the interpretation of responses in contingent valuation surveys. In Johansson, P.-O., Kriström, B., and Mäler, K-G. (eds.), editors, *Current Issues in Environmental Economics*. Manchester University Press, Manchester, UK.

Hartman, R. S., Doane, M. J., and Woo, C. (1991). Consumer rationality and the status-quo. *The Quarterly Journal of Economics*, 106:141–162.

Hausman, J. A. and Wise, D. A. (1978). A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica*, 46:403–426.

Hellekalek, Peter (1998). Good random number generators are (not so) easy to find. *Mathematics and Computers in Simulations*, 46:485–505.

Henriques, I. and Sardorsky, P. (1996). The determinants of an environmentally responsive firm: An empirical approach. *Journal of Environmental Economics and Management*, 30:391–395.

Hensher, D. and Greene, W. H. (2003). The Mixed Logit Model: The State of Practice. *Transportation*, 30(2):133–176.

Hensher, D., Louviere, J., and Swait, J. (1999). Combining sources of preference data. *Journal of Econometrics*, 89:197–221.

Hensher, D., Shore, N., and Train, K. (2004). Households' willingness to pay for water service attributes. Working Paper, Department of Economics, University of California, Berkeley.

Herrero, I. and Pascoe, S. (2003). Value versus volume in the catch of the spanish south-atlantic trawl industry. *Journal of Agricultural Economics*, 54:325–341.

Herriges, J. A. and Kling, C. L. (1997). The performance of nested logit models when welfare estimation is the goal. *American Journal of Agricultural Economics*, 79(3):792–802.

Herriges, J. A., Kling, C. L., and Phaneuf, D. J. (2004). What's the use? welfare estimates from revealed preference models when weak complementarity doesn't hold. *Journal of Environmental Economics and Management*, 47:55–70.

Herriges, J. A. and Phaneuf, D. J. (2002). Inducing patterns of correlation and substitution in repeated logit models of recreation demand. *American Journal of Agricultural Economics*, 84:1076–1090.

Herriges, J. A. and Shogren, J. F. (1996). Starting point bias in dichotomous choice valuation with follow-up questioning. *Journal of Environmental Economics and Management*, 30:112–31.

Hess, S., Bierlaire, M., and Polak, J. W. (2005). Estimation of value of travel-time savings using mixed logit models. *Transportation Research A, forthcoming*, 39:221–236.

Hess, S. and Polak, J. W. (2004a). *Mixed Logit estimation of parking type choice*. Paper presented at the 83^{rd} Annual Meeting of the Transportation Research Board, Washington, DC.

Hess, S. and Polak, J. W. (2004b). Mixed logit modelling of airport choice in multi-airport regions. *Journal of Air Transport Management*, forthcoming.

Hess, S., Train, K. E., and Polak, J. W. (2004). On the use of a modified latin hypercube sampling (MLHS) method in the estimation of a mixed logit model for vehicle choice. *Transportation Research B, forthcoming*.

Hinkley, D. (1969). On the ratio of two normal random variables. *Biometrika*, 56:635–639.

Hoaglin, D. C. and Andrews, D. F. (1975). The Reporting of Computation-Based Results in Statistics. *The American Statistician*, 29(3):122–126.

Hoehn, J. P. and Randall, A. (1987). A satisfactory benefit-cost indicator from contingent valuation. *Journal of Environmental Economics and Management*, 14:226–247.

Hogg, Robert V., McKean, Joseph W., and Craig, Allen T. (2005). *Introduction to Mathematical Statistics*. Prentice Hall, New Jersey, sixth edition.

Horowitz, J., Sparmonn, J., and Daganzo, C. (1981). An Investigation of the Accuracy of the Clark Approximation for the Multinomial Probit. *Transportation Science*, 16:382–401.

Howitt, R. E., Reynaud, A., Msangi, S., and Knapp, K. C. (2002). Calibrated Stochastic Dynamic Models for Resource Management. Paper presented at the World Congress of Environmental and Resource Economists, Monterey, CA.

Huang, J. C. and Smith, V. K. (1998). Monte Carlo benchmarks for discrete response valuation methods. *Land Economics*, 74(2):186–202.

Huber, J. and Train, K. (2001). On the similarity of classical and bayesian estimates of individual mean partworths. *Marketing Letters*, 12:259–269.

Huhtala, A. (2000). Binary choice valuation studies with heterogeneous preferences regarding the program being valued. *Environmental and Resource Economics*, 16:263–279.

Ihaka, Ross and Gentleman, Robert (1996). R: a Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5:299–314.

Jaffe, A.B., Newell, R.G., and Stavins, R.N. (2002). Environmental policy and technological change. *Environmental and Resource Economics*, 22:41–69.

Jaffe, A.B. and Palmer, K. (1997). Environmental regulation and innovation: A panel data study. *The Review of Economics and Statistics*, 79:610–619.

Jeffreys, H. (1966). *Theory of Probability*. Clarendon, Oxford, U.K., 3rd edition.

Johnson, N. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, 36:149–176.

Johnson, N. and Kotz, S. (1970). *Continuous Multivariate Distributions I*. John Wiley, New York.

Johnson, R. M. (2000). Monotonicity constraints in choice-based conjoint with hierarchical bayes. Research paper, Sawtooth Software.

Jondrow, J., Lovell, K., Matterov, I., and Schmidt, P. (1982). On the estimation of technical ineffiency in the stochastic frontier production model. *Journal of Econometrics*, pages 233–38.

Joskow, P. L. (2001). California's energy crisis. *Oxford Review of Economic Policy*, 17:365–388.

Judd, K. L. (1998). *Numerical Methods in Economics*. The MIT Press, Cambridge MA.

Kahneman, D. and Tversky, A. (1979). Prospect theory: an analysis of decisions under risk. *Econometrica*, 47:263–91.

Kanninen, B. J. (1993). Optimal experimental design for double bounded dichotomous choice contingent valuation. *Land Economics*, 69(2):138–146.

Kanninen, B. J. (1995). Bias in discrete response contingent valuation. *Journal of Environmental Economics and Management*, 28:114–125.

Kass, R. E. and Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90:773–795.

Keane, M. (1992). A note on identification in the multinomial probit model. *Journal of Business and Economic Statistics*, 10(2):193–200.

Keane, M. (1994). A computationally practical simulation estimator for panel data. *Econometrica*, 62(1):95–116.

Kennedy, J. O. S. (1986). *Dynamic Programming: Applications to Agriculture and Natural Resources*. Elsevier, New York.

Khanna, M. (2001). Sequential adoption of site-specific technologies and its implications for nitrogen productivity: a double selectivity model. *American Journal of Agricultural Economics*, 81:35–51.

Khanna, M. and Anton, W. R. Q. (2002). Corporate environmental management: Regulatory and market-based incentives. *Land Economics*, 78:539–558.

Kim, J., Allenby, G. M., and Rossi, P. E. (2002). Modeling consumer demand for variety. *Marketing Science*, 21:229–250.

Kirkley, J. E., Squires, D., and Strand, I. E. (1995). Assessing technical efficiency in commercial fisheries: The Mid-Atlantic sea scallop fishery. *American Journal of Agricultural Economics*, 77:686–697.

Kling, C. L. (1991). Estimating the precision of welfare measures. *Journal of Environmental Economics and Management*, 21:244–259.

Kling, C. L. (1997). The gains from combining travel cost and contingent valuation data to value nonmarket goods. *Land Economics*, 73(3):428–439.

Kling, C. L. and Herriges, J. (1995). An empirical investigation of the consistency of nested logit models with utility maximization. *American Journal of Agricultural Economics*, 77:875–884.

Kling, C. L. and Thompson, J. C. (1996). The implications of model specification for welfare estimation in nested logit models. *American Journal of Agricultural Economics*, 78:103–114.

Knuth, D. E. (1998). *The Art of Computer Programming: Seminumerical Algorithms*, volume 2. Addison-Wesley, third edition.

Koenker, R. (1996a). Reproducibility in econometrics research. Technical report, Department of Econometrics, University of Illinois, Urbana-Champaign, IL.

Koenker, R. (1996b). Reproducible econometric research. Technical report, Department of Econometrics, University of Illinois, Urbana-Champaign, IL.

Kompas, T., Che, T. N., and Grafton, R. Q. (2003). Technical efficiency effects of input controls: evidence from Australias banana prawn fishery. Online Research Paper ECDV03-2, Asia Pacific School of Economics and Government.

Kontoleon, A. and Yabe, M. (2003). Assessing the impacts of alternative 'opt-out' formats in choice experiment studies: consumer preferences for genetically modified content and production information in food. *Journal of Agricultural Policy and Resources*, 5:1–43.

Koop, G., Osiewalski, J., and Steel, M. F. J. (1997). Bayesian efficiency analysis through individual effects: Hospital cost frontiers. *Journal of Econometrics*, 76:77–105.

Koop, G. and Poirier, D. (1993). Bayesian analysis of logit models using natural conjugate priors. *Journal of Econometrics*, 56:323–340.

Koop, G. and Poirier, D. (1996). Rank ordered logit models: An empirical analysis of voter preferences before the 1988 canadian federal election. *Journal of Applied Econometrics*, 9:369–388.

Koop, G. and Steel, M. F. J. (2001). A companion to theoretical econometrics. In Baltagi, B., editor, *Bayesian Analysis of Stochastic Frontier Models*, pages 520–537. Blackwell, Oxford, U.K.

Koop, G. and Tole, L. (2004). Measuring The Health Effects of Air Pollution: to What Extent Can We Really Say that People are Dying from Bad Air? *Journal of Environmental Economics and Management*, 47:30–54.

Koppelman, F. and Wen, C. (2000). The paired combinatorial logit model: Properties, estimation and application. *Transportation Research B*, 34:75–89.

Krinsky, I. and Robb, A. L. (1986). On approximating the statistical properties of elasticities. *Review of Economics and Statistics*, 68:715–719.

Krinsky, I. and Robb, A. L. (1990). On Approximating the Statistical Properties of Elasticities: A Correction. *Review of Economics and Statistics*, 72:189–190.

Kristöm, B. (1991). A non-parametric approach to the estimation of welfare measures in discrete response valuation studies. *Land Economics*, 66:135–139.

Kriström, B. (1997). Spike models in contingent valuation. *American Journal of Agriculture Economics*, 79:1013–1023.

Kuhfeld, W. F. (2004). Marketing research methods in SAS: experimental design, choice, conjoint, and graphical techniques. Technical support document 694c, SAS Institute Inc., Cary, NC, USA.

Kuhn, H. W. and Tucker, A. W. (1951). Nonlinear programming. In Neyman, J., editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 481–492. University of California Press, Berkeley, CA, U.S.A.

Kumbhakar, S.C. and Lovell, C. A. K. (2000). *Stochastic Frontier Analysis*. Cambridge University Press, Cambridge.

Layton, D. F. (1995). Measuring existence value for species and habitats.

Layton, D. F. (2000). Random coefficient models for stated preference surveys. *Journal of Environmental Economics and Management*, 40:21–36.

Layton, D. F. and Brown, G. (2000). Heterogeneous preferences regarding global climate change. *The Review of Economics and Statistics*, 82(4):616–624.

Layton, D. F. and Levine, R. (2004). How much does the far future matter? A hierarchical bayesian analysis of the public's willingness to mitigate ecological impacts of climate change. *Journal of the American Statistical Association*, 98:533–544.

Layton, D. F. and Levine, R. (2005). Bayesian approaches to modelling stated preference data. In Scarpa, R. and Alberini, A., editors, *This volume*. Springer.

L'Ecuyer, P. (1998). Random number generation. In Banks, Jerry, editor, *Handbook on Simulation, Chapter 4*, pages 93–137. Wiley, New York.

L'Ecuyer, P. (1999). Good parameters and implementations for combined multiple recursive random number generators. *Operations Research*, 7(1):159–164.

Lee, L.-F. (1982). Some approaches to the correction of selectivity bias. *Review of Economic Studies*, 49:355–372.

Lee, L.-F. (1983). Generalized econometric models with selectivity. *Econometrica*, 51:507–12.

Lee, L.-F. and Pitt, M. (1986). Microeconomic demand systems with binding non-negativity constraints: The dual approach. *Econometrica*, 54:1237–1242.

Lehmer, D.H. (1951). Mathematical methods in large-scale computing units. In *Proc. 2nd Sympos. on Large-Scale Digital Calculating Machinery, Cambridge, MA, 1949*, pages 141–146, Cambridge, MA. Harvard University Press.

Lehtonen, E., Kuuluvainen, J., Pouta, E., M. Rekola, M., and Li, C.-Z. (2003). Non-market benefits of forest conservation in southern finland. *Environmental Science and Policy*, 6:195–204.

León, C. J., Vazquez-Polo, F. J., Guerra, N., and Riera, P. (2002). A Bayesian Model for Benefits Transfer: Application to National Parks in Spain. *Applied Economics*, 34:749–757.

León, R. and León, C.J. (2003). Single or double bounded contingent valuation? a Bayesian test. *Scottish Journal of Political Economy*, 50(2):174–188.

Li, C.-Z. (1996). Semiparametric estimation of the binary choice model for contingent valuation. *Land Economics*, 72:462–73.

Lindley, D.V. and Smith, A.F.M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society*, 34:1–41.

Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. Springer-Verlag, New York, NY, U.S.A.

Louviere, J. (2003). Randon utility theory-based stated preference elicitation methods: Applications in health economics with special reference to combining sources of preference data. Working Paper, Faculty of Business, University of Technology, Sydney.

Louviere, J., Hensher, D., and Swait, J. (2000). *Stated Choice Methods: Analysis and Applications*. Cambridge University Press, New York.

Lynch, S. and Smith, K.R. (1994). Lean, mean, and green... designing farm support programs in a new era. Department of Agricultural Economics.

MacLaren, N. M. (1992). A Limit on the Usable Length of a Pseudorandom Sequence. *Journal of Statistical Computing and Simulation*, 42:47–54.

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge.

Manski, C.F. (2004). Measuring expectations. *Econometrica*, 72:1329–1376.

Marsaglia, G. (1965). Ratios of normals variables and ratios of sums of uniform variables. *Journal of the American Statistical Association*, 60:193–204.

Marsaglia, G. (1996). DIEHARD: A Battery of Tests of Randomness. Available at *http://stat.fsu.edu/pub/diehard/*.

Marsaglia, George (2000). The monster, A random number generator with period over 10^{2857} times as long as the previously touted longest-period one. Technical report, Florida State University, Tallahassee, FL, USA.

Martin, M. (1985). Design of a food intake study in two Bambara villages in the Segou Region of Mali with preliminary findings. In Hill, A., editor, *Population, Health and Nutrition in the Sahel*. KPI Limited, London, U.K.

Matsumoto, M. and Kurita, Y. (1992). Twisted GFSR Generators. *acmt*, 2:179–194.

Matsumoto, M. and Kurita, Y. (1994). Twisted GFSR Generators II. *acmt*, 4:254–266.

Matsumoto, M. and Nishimura, T. (1998). A 623-Dimensionally Equidistributed Uniform Pseudorandom Number Generator. *acmt*, 8(1):3–30.

McCulloch, R. and Rossi, P. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, 64:207–240.

McCullough, B. D. (1999a). Econometric Software Reliability: EViews, LIMDEP, SHAZAM and TSP. *Journal of Applied Econometrics*, 14(2):191–202.

McCullough, B.D. (1999b). Assessing the Reliability of Statistical Software: Part II. *The American Statistician*, 53(1):149–159.

McCullough, B.D. (2000). Is It Safe to Assume That Software is Accurate? *International Journal of Forecasting*, 16(3):349–357 (doi:10.1016/S0169–2070(00)00032–7).

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–42. Academic Press, New York.

McFadden, D. (1978). Modeling the choice of residential location. In Karlqvist, A., Lundqvist, L., Snickars, F., and Weibull, J., editors, *Spatial Interaction Theory and Planning Models*, pages 75–96. North-Holland, Amsterdam.

McFadden, D. (1981). Econometric models of probabilistic choice. In Manski, C. and McFadden, D., editors, *Structural Analysis of Discrete Data with Econometric Applications*, chapter 5, pages 198–272. MIT Press, Cambridge, MA, USA.

McFadden, D. (1989). A method of simulated moments for estimation of discrete choice models without numerical integration. *Econometrica*, 57(5):995–1026.

McFadden, D. (1994). Contingent Valuation and Social Choice. *American Journal of Agricultural Economics*, 76:689–708.

McFadden, D. and Ruud, P. (1994). Estimation by simulation. *Review of Economics and Statistics*, 76(4):591–608.

McFadden, D. and Train, K. (1996). Mixed MNL models for discrete response. Technical report, Department of Economics, Berkeley.

McFadden, D. and Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15:447–470.

McLeod, D. M. and Bergland, O. (1999). Willingness to pay estimates using the double bounded dichotomous choice contingent valuation format: a test for validity and precision in a bayesian framework. *Land Economics*, 75:115–25.

McLeod, K., Boxall, P., Adamowicz, W., Williams, M., and Louviere, J. (1993). The incorporation of non-timber goods and services in integrated resource management. Project Report 93-12, Department of Rural Economy, University of Alberta.

McPeak, J. (2003). Analyzing and addressing localized degradation in the commons. *Land Economics*, 79(4):515–536.

Meilby, H., Strange, N., and Thorsen, B. J. (1999). A comparison of simulated annealing, tabu search and selforganising algorithms for land use optimization, in P. Lohmander (ed.): Proceedings of the Biennial Meeting of Scandinavian Society of Forest Economics. Umeå, Sweden, 2 6-28 May, 1998. *Scandinavian Forest Economics*, 37:173–185.

Merrill, A. (1928). Frequency distribution of an index when both the components follow the normal law. *Biometrika*, 20A:53–63.

Metropolis, N. and Ulam, S. M. (1949). The Monte Carlo method. *Journal of the American Statistical Association*, 44(247):335–341.

Metropolis, N. C., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092.

Moeltner, K. and Layton, D. (2002). A censored random coefficients model for pooled survey data with application to the estimation of power outrage costs. *Review of Economics and Statistics*, 84:552–561.

Montgomery, C. A., Brown, G. M. Jr., and Adams, D. M. (1994). The marginal cost of species preservation: The northern spotted owl. *Journal of Environmental Economics and Management*, 26:111–128.

Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). *Introduction to the Theory of Statistics*. McGraw-Hill, New York.

Morrison, M., Bennett, J., Blamey, R., and Louviere, J. (2002). Choice modeling and tests of benefit transfer. *American Journal of Agricultural Economics*, 84:161–170.

Mrkaic, M. (2002). Policy Iteration Accelerated with Krylov Methods. *Journal of Economic Dynamics and Control*, 26:517–45.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46:69–85.

Munizaga, M. A. and Alvarez, R. (2001). Mixed logit vs. nested logit and probit models. 5th tri-annual Invitational Choice Symposium Workshop: Hybrid Choice Models, Formulation and Practical Issues.

Munizaga, M. A., Heydecker, B. G., and de Dios Ortúzar, J. (2000). Representation of heteroskedasticity in discrete choice models. *Transportation Research B*, 34(3):219–240.

Nakamura, M., Takahashi, T., and Vertinsky, I. (2001). Why Japanese firms choose to certify: A study of managerial responses to environmental issues. *Journal of Environmental Economics and Management*, 43:23–52.

Nelsen, R. B. (1999). *An Introduction to Copulas. Lecture Notes in Statistics.* Springer-Verlag, New York, U.S.A.

Newey, W. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In Griliches, Z. and Intriligator, M., editors, *The Handbook of Econometrics, Volume 4*, pages 2111–2245. Elsevier Science Piblishers, Amsterdam.

Newey, W. K. and West, K. D. (1987). A simple, semi-positive definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708.

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56:3–48.

Nyquist, H. (1990). Optimal Designs of Discrete Response Experiments in Contingent Valuation Studies. *Review of Economics and Statistics*, 74:559–563.

OECD (1997). Environmental indicators for agriculture. Technical report, OECD, Paris.

OECD, Organisation for Economic Co-operation and Development (1997). OECD proposed guidelines for collecting and interpreting technological innovation data. Technical report, OECD/Eurostat, Paris, France.

Ortúzar, J. de D. and Willumsen, L.G. (2001). *Modelling Transport*. John Wiley and Sons, Chichester, U.K., third edition.

Papola, A. (2004). Some developments on the cross-nested logit model. *Transportation Research*, 38B(9):833–854.

Park, Stephen K. and Miller, Keith W. (1988). Random number generators: Good ones are hard to find. *Communications of ACM*, 31(10):1192–1201.

Park, T., Loomis, J. B., and Creel, M. (1991). Confidence intervals for evaluating benefits estimates from dichotomous choice contingent valuation studies. *Land Economics*, 67(1):65–73.

Parsons, G., Jakus, P., and Tomasi, T. (1999). A comparison of welfare estimates from four models for linking seasonal recreation trips to multinomial logit models of choice. *Journal of Environmental Economics and Management*, 38:143–157.

Pascoe, S., Andersen, J. L., and de Wilde, J-W. (2001). The impact of management regulation on the technical efficiency of vessels in the dutch beam trawl industry. *European Review of Agricultural Economics*, 28:187–206.

Pascoe, S. and Coglan, L. (2002). The contribution of unmeasurable inputs to fisheries production: an analysis of technical efficiency of fishing vessels in the english channel. *American Journal of Agricultural Economics*, 84:585–597.

Pascoe, S., Hassaszahed, P., Anderson, J., and Korsbrekkes, K. (2003). Economic versus physical input measures in the analysis of technical efficiency in fisheries. *Applied Economics*, 35:1699–1710.

Pavitt, K. (1984). Sectoral patterns of technical change: Towards a taxonomy and a theory. *Research Policy*, 13:343–373.

Phaneuf, D. J. (1999). A dual approach to modeling corner solutions in recreation demand. *Journal of Environmental Economics and Management*, 37:85–105.

Phaneuf, D. J. and Herriges, J. A. (1999). Choice set definition in a Kuhn-Tucker model of recreation demand. *Marine Resource Economics*, 14:343–355.

Phaneuf, D. J., Kling, C. L., and Herriges, J. A. (2000). Estimation and welfare calculations in a generalized corner solution model with an application to recreation demand. *Review of Economics and Statistics*, 82:83–92.

Phaneuf, D. J. and Smith, V. K. (2005). Recreation demand models. forthcoming, *Handbook of Environmental Economics*, eds. K. Mäler and J. Vincent.

Poirier, D. (1994). Jeffrey's prior for logit models. *Journal of Econometrics*, 63:327–329.

Poirier, D. (1996). A bayesian analysis of nested logit models. *Journal of Econometrics*, 75:163–181.

Poirier, D. J. (1988). Frequentist and subjectivist perspectives on the problems of model building in economics (with discussion and reply). *Journal of Economic Perspectives*, 2:121–170.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1995). *Numerical Recipes in C*. Cambridge University Press, Cambridge, second edition.

Provencher, B. (1995a). An Investigation of the Harvest Decisions of Timber Firms in the Southeast United States. *Journal of Applied Econometric*, 10:S57–74.

Provencher, B. (1995b). Structural Estimation of the Stochastic Dynamic Decision Problems of Resource Users: An Application to the Timber Harvest Decision. *Journal of Environmental Economics and Management*, 29:321–338.

Provencher, B. and Bishop, R. C. (1997). An Estimable Dynamic Model of Recreation Behavior with an Application to Great Lakes Angling. *Journal of Environmental Economics and Management*, 33:107–127.

Pudney, S. (1989). *Modelling Individual Choice: The econometrics of corners, kinks, and holes*. Basil-Blackewell, New York.

Raftery, A. E. and Lewis, S. M. (1992). How many iterations in the Gibbs sampler? In et al., J.M. Bernardo, editor, *Bayesian Statistics*, pages 763–773. Oxford University Press, Oxford, U.K.

Randall, A. (1998). Beyond the crucial experiment: mapping the performance characteristics of contingent valuation. *Resource and Energy Economics*, 20:197–206.

Rennings, K. (2000). Redefining innovation - Eco-innovation research and the contribution from ecological economics. *Ecological Economics*, 32:319–332.

Rennings, K., Ziegler, A., Ankele, K., Hoffmann, E., and Nill, J. (2003). The influence of the EU environmental management and auditing scheme on environmental innovations and competitiveness in Germany. An analysis on the basis of case studies and a large-scale survey. Discussion Paper 02-14, ZEW, Mannheim, Germany.

Rennings, K., Ziegler, A., and Zwick, T. (2004). The effect of environmental innovations on employment changes: An econometric analysis. *Business Strategy and the Environment*, 13(6):374–387.

Rennings, K. and Zwick, T. (2002). The employment impact of cleaner production on the firm level - Empirical evidence from a survey in five European countries. *International Journal of Innovation Management*, 6(3):319–342.

Revelt, D. (1999). *Three Discrete Choice Random Coefficients Papers and One Police Crime Study*. PhD thesis, University of California, Berkeley.

Revelt, D. and Train, K. (1998). Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of Economics and Statistics*, 80(4):647–657.

Ripley, B. D. (1987). *Stochastic Simulation*. John Wiley, New York.

Robert, C. P. (2001). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer, New York, NY, U.S.A., 2nd edition.

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer Verlag, New York, NY, U.S.A.

Roe, B., Boyle, K. J., and Mario, T. F. (1996). Using conjoint analysis to derive estimates of compensating variation. *Journal of Environmental Economics and Management*, 31:145–159.

Rust, J. (1987). Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurchner. *Econometrica*, 55:993–1033.

Rust, J. (1994a). Do People Behave According to Bellman's Principle of Optimality? Manuscript, University of Wisconsin-Madison.

Rust, J. (1994b). Structural estimation of markov decision processes. In Engle, R. and McFadden, D., editors, *Handbook of Econometrics*, pages 3081–143. North-Holland, Amsterdam.

Rust, J. (1997a). Dealing with the Complexity of Economic Calculations. Economics Working Paper Archive at WUSTL.

Rust, J. (1997b). Using randomization to break the curse of dimensionality. *Econometrica*, 65:487–516.

Ruud, P. A. (1983). Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of distribution in multinomial discrete choice models. *Econometrica*, 51:225–228.

Ruud, P. A. (1996). Simulation of the multinomial probit model: An analysis of covariance matrix estimation. Working Paper,Department of Economics, University of California.

Samuelson, W. and Zeckhauser, R. (1988). Status-quo bias in decision-making. *Journal of Risk and Uncertainty*, 24(1):7–59.

Sandor, Z. and Train, K. (2004). Quasi-random simulation of discrete choice models. *Transportation Research Part B*, 38:313–327.

Sawitzki, G. (1985). Another random number generator which should be avoided. *Statistical Software Newsletter*, 11:81–82.

Sawtooth Software (1999). The CBC/HB module for hierarchical Bayes. Technical Paper available on the web site of Sawtooth Software.

Scarpa, R. and Bateman, I. (2000). Efficiency gains afforded by improved bid design versus follow-up valuation questions in discrete choice CV studies. *Land Economics*, 76:299–311.

Scarpa, R., Spalatro, F., and Canavari, M. (2002). Investigating preferences for environment friendly production practices. AgEcon Search Working Paper 03-3, http://agecon.lib.umn.edu/.

Scarpa, R., Willis, K. G., and Acutt, M. (2005). Individual-specific welfare measures for public goods: a latent class approach for residential customers of Yorkshire Water. In Koundouri, P., editor, *Econometrics Informing Natural Resource Management*, chapter 14. Edward Elgar Publisher, London, U.K.

Schaller, R.R. (1997). Moore's Law: Past, Present, and Future. *IEEE Spectrum*, 34:53–59.

Schmidt, P. and Sickles, R. C. (1984). Production frontiers and panel data. *Journal of Business and Economic Statistics*, 2:367–374.

Schwab, Matthias, Karrenbach, Martin, and Claerbout, Jon (1995). Making scientific computations reproducible. *IEEE MultiMedia*, 2(6):61–67.

Schwartz, H., Shaabani, S., and Walther, D. (1991). Range Management Handbook of Kenya Volume II-1, Marsabit District. Technical report, Republic of Kenya, Ministry of Livestock Development, Nairobi.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components*. Wiley, New York.

Sillano, M. and Ortúzar, J. de D. (2004). Willingness-to-pay estimation with mixed logit models: some new evidence. *Environment and Planning A*, in press.

Sklar, A. (1959). Fonctions de répartition á n dimensions et leurs marges. *Publ. Inst. Statis. Univ. Paris*, 8:229–231.

Sklar, A. (1996). Random variables, distribution functions, and copulas - a personal look backward and forward. In Rüschendorf, L., Schweizer, B., and Taylor, M.D., editors, *Distributions with Fixed Marginals and Related Topics*. Institute of Mathematical Statistics, Hayward, CA, U.S.A.

Small, K. (1987). A discrete choice model for ordered alternatives. *Econometrica*, 55:409–424.

Smith, V.K. (1992). Environmental Costing for Agriculture: Will it be Standard Fare in the Farm Bill of 2000? *American Journal of Agricultural Economics*, 74:1076–1088.

Sonnier, G., Ainslie, A., and Otter, T. (2003). The influence of brand image and product style on consumer brand valuations. Working Paper, Anderson Graduate School of Management, University of California, Los Angeles.

Squires, D., Grafton, R. Q., Alam, M. F., and Omar, I. H. (2003). Technical efficiency in the Malaysian gill net artisanal fishery. *Environment and Development Economics*, 8:481–504.

Squires, D. and Kirkley, J. (1999). Skipper skill and panel data in fishing industries. *Canadian Journal of Fisheries and Aquatic Sciences*, 56:2011–2018.

Stern, S. (1992). A method for smoothing simulated moments of discrete probabilities in multinomial probit models. *Econometrica*, 60:943–952.

Stern, S. (1997). Simulation-based estimation. *Journal of Economic Literature*, 35(December):2006–2039.

Stern, S. (2000). Simulation-based inference in econometrics: Motivation and methods. In Mariano, R., Schuermann, T., and Weeks, M. J., editors, *Simulation-Based Inference in Econometrics: Methods and Applications*, pages 9–37. Cambridge University Press, Cambridge, U.K.

Strazzera, E., Scarpa, R., Calia, P., Garrod, G. D., and Willis, K. G. (2003). Modelling zero values and protest responses in contingent valuation surveys. *Applied Economics*, 35:133–138.

Student (1908). The probable error of a mean. *Biometrika*, 6:1–25.

Sullivan, M. J., Suddeth, B. N., Vardell, T., and Vojdani, A. (1996). Interruption costs, customer satisfaction and expectations for service reliability. *IEEE Transactions on Power Systems*, 11:989–995.

Swait, J. D. and Adamowicz, W. (2001). Choice environment, market complexity and consumer behaviour: a theoretical and empirical approach for incorporating decision complexity into models of consumer choice. *Organizational Behaviour and Human Decision Processes*, 86:141–167.

Tanner, M.A. and Wong, W.H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82:528–550.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics*, 22:1701–62.

Train, K. (1998). Recreation demand models with taste variation. *Land Economics*, 74:230–239.

Train, K. (1999). *Halton sequences for Mixed Logit*. Working Paper. Department of Economics, University of California, Berkeley, California.

Train, K. (2000). Halton sequences for mixed logit. Working Paper No. E00-278, Department of Economics, University of California, Berkeley.

Train, K. (2001). A comparison of hierarchical bayes and maximum simulated likelihood for mixed logit. Working Paper, Department of Economics, University of California, Berkeley.

Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, MA, U.S.A.

Train, K. and Hudson, K. (2000). The impact of information in vehicle choice and the demand for electric vehicles in california. Project report, National Economic Research Associates.

Train, K. and Sonnier, G. (2005). Mixed logit with bounded distributions of correlated partworths. In Scarpa, R. and Alberini, A., editors, *Applications of simulation methods in Environmental and Resource Economics*. Chapter 7 of this volume. Springer, Boston, MA, U.S.A.

Train, K. and Winston, C. (2004). Vehicle choice behavior and the declining market of us automakers. Working Paper, Department of Economics, University of California, Berkeley.

Tsionas, E.G. (2002). Stochastic frontier models with random coefficients. *Journal of Applied Econometrics*, 17:127–147.

Ulam, S. M. (1977). *Adventures of a Mathematician*. Scribners, New York.

U.S. Department of Commerce (2002). Fisheries of the United States; Current Fisheries Statistics No. 2000. Technical report, National Oceanic and Atmospheric Administration, National Marine Fisheries Service.

van Kooten, G.C., Young, D.L., and Krautkraemer, J.A. (1997). A Safety-First Approach to Dynamic Cropping Decisions. *European Review of Agricultural Economic*, 24:47–63.

Vijverberg, W.P.M. (1997). Monte Carlo evaluation of multivariate normal probabilities. *Journal of Econometrics*, 76:281–307.

Vinod, H. D. (2000). Review of GAUSS for Windows, including its numerical accuracy. *Journal of Applied Econometrics*, 14(2):211–220.

Vinod, H. D. (2001). Care and feeding of reproducible econometrics. *Journal of Econometrics*, 100(1):87–88 (doi:10.1016/S0304–4076(00)00064–6).

Viswanathan, K. and Omar, K. (2001). Fishing skill in developing country-fisheries: the Kedah, Malaysia trawl fishery. *Marine Resource Economics*, 16:293–314.

von Haefen, R. H. (2003). Incorporating observed choice into the construction of welfare measures from random utility models. *Journal of Environmental Economics and Management*, 45:145–165.

von Haefen, R. H. (2004a). Empirical strategies for incorporating weak complementarity into continuous demand system models. Working Paper, Department of Agricultural and Resource Economics, University of Arizona.

von Haefen, R. H. (2004b). Latent consideration sets and continuous demand system models. Working Paper, Department of Agricultural and Resource Economics, University of Arizona.

von Haefen, R. H. and Phaneuf, D. J. (2003). Estimating preferences for outdoor recreation: A comparison of continuous and count data demand systems. *Journal of Environmental Economics and Management*, 45:612–630.

von Haefen, R. H., Phaneuf, D. J., and Parsons, G.R. (2004). Estimation and welfare calculation with large demand systems. *Journal of Business and Economic Statistics*, 22:194–205.

Vovsha, P. (1997). Application of a cross-nested logit model to mode choice in Tel Aviv, Israel, metropolitan area. *Transportation Research Record*, 1607:6–15.

Vovsha, P. and Bekhor, S. (1998). The link-nested logit model of route choice: overcoming the route overlapping problem. *Transportation Research Record*, 1645:133–142.

Wacker, G., Subramaniam, R. K., and Billington, R. (1985). Using cost of electric service interruption surveys in the determination of a composite customer damage function. *International Journal of Energy Systems*, 5:100–104.

Walden, J. B., Kirkley, J. E., and Kitts, A. W. (2003). A limited economic assessment of the northeast groundfish fishery buyout program. *Land Economics*, 79:426–439.

Waldman, D. (1985). Computation in duration models with heterogeneity. *Journal of Econometrics*, 28:127–134.

Wales, T. and Woodland, A. (1983). Estimation of consumer demand systems with binding non-negativity constraints. *Journal of Econometrics*, 21:63–285.

Walker, J. L. (2001). *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables*. PhD thesis, MIT, Cambridge, MA.

Wang, Y. (2003). *Estimation of Recreation Demand Models Using Dual Approaches*. Ph.d. dissertation, North Carolina State University.

Warwick, W. M. (2000). A primer on the deregulation and restructuring of the U.S. Electric utility market. Technical report, Pacific Northwest National Laboratory, Richland, WA.

Watkins, K. B., Lu, Y., and Huang, W. (1998). Economic and Environmental Feasibility of Variable Rate Nitrogen Fertilizer Application with Carry-Over Effects. *Journal of Agricultural and Resource Economics*, 23:401–26.

Wen, C.-H. and Koppelman, F. (2001). The generalized nested logit model. *Transportation Research B*, 35:627–641.

Werner, M. (1999). Allowing for zeros in dichotomous-choice contingent-valuation models. *Journal of Business and Economic Statistics*, 17:479–486.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25.

Whitehead, J.C. (2002). Incentive incompatibility and starting point bias in iterative valuation questions. *Land Economics*, 78:285–97.

Wilen, J. (1979). Fisherman behaviour and the design of efficient fisheries regulation programmes. *Journal of Fisheries Research Board of Canada*, 36:855–858.

Williams, H.C.W.L. (1977). On the formulation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning A*, 9(3):285–344.

Willis, K.G. and Scarpa, R. (2002). Yorkshire Water customers' utility and willingness-to-pay for changes in water service factor levels: A stated choice analysis. Report to Yorkshire Waters.

Willis, K.G., Scarpa, R., and Acutt, M. (2004). Assessing water company customer preferences and willingness-to-pay for service improvements: a stated choice analysis. Forthcoming, *Water Resources Research*.

Woo, C.-K. and Pupp, R. L. (1992). Costs of service disruptions to electricity consumers. *Energy*, 17:109–126.

Woo, C.-K., Pupp, R. L., Flaim, T., and Mango, R. (1991). How much do electric customers want to pay for reliability? New evidence on an old controversy. *Energy Systems and Policy*, 15:145–159.

Woodward, R.T., Wui, Y.S., and Griffin, W.L. (2005). Living with the curse of dimensionality: Closed-loop optimization in a large-scale fisheries simulation model. *American Journal of Agricultural Economics*, 87:48–60.

Yu, P. L. H. and Chan, L.K.Y. (2001). Bayesian analysis of wandering vector models for displaying ranking data. *Statistics Sinica*, 11:445–461.

Zellner, A. (1996). *An Introduction to Bayesian Inference in Econometrics*. John Wiley and Sons, Wiley Classics Library, New York, NY, U.S.A.

Ziegler, A. (2001). Simulated z-tests in multinomial probit models. Discussion Paper 01-53, ZEW, Mannheim, Germany.

Ziegler, A. and Eymann, A. (2001). Zur Simulated Maximum Likelihood. Schätzung von Mehrperioden-Mehralternativen-Probitmodellen. *Allgemeines Statistisches Archiv*, 85(3):319–342.

Ziegler, A. and Rennings, K. (2004). Determinants of environmental innovations in Germany: Do organizational measures matter? A discrete choice analysis at the firm level. Discussion Paper 04-30, ZEW, Mannheim, Germany.

Index