

Causal Analysis in Population Studies

THE SPRINGER SERIES ON DEMOGRAPHIC METHODS AND POPULATION ANALYSIS

Series Editor

KENNETH C. LAND

Duke University

In recent decades, there has been a rapid development of demographic models and methods and an explosive growth in the range of applications of population analysis. This series seeks to provide a publication outlet both for high-quality textual and expository books on modern techniques of demographic analysis and for works that present exemplary applications of such techniques to various aspects of population analysis.

Topics appropriate for the series include:

- General demographic methods
- Techniques of standardization
- Life table models and methods
- Multistate and multiregional life tables, analyses and projections
- Demographic aspects of biostatistics and epidemiology
- Stable population theory and its extensions
- Methods of indirect estimation
- Stochastic population models
- Event history analysis, duration analysis, and hazard regression models
- Demographic projection methods and population forecasts
- Techniques of applied demographic analysis, regional and local population estimates and projections
- Methods of estimation and projection for business and health care applications
- Methods and estimates for unique populations such as schools and students

Volumes in the series are of interest to researchers, professionals, and students in demography, sociology, economics, statistics, geography and regional science, public health and health care management, epidemiology, biostatistics, actuarial science, business, and related fields.

For other titles published in this series, go to
www.springer.com/series/6449

Causal Analysis in Population Studies

Concepts, Methods, Applications

Edited by

Henriette Engelhardt

University of Bamberg, Germany

Hans-Peter Kohler

University of Pennsylvania, Philadelphia, PA, USA

and

Alexia Prskawetz

Vienna Institute of Demography, Austrian Academy of Sciences
and Vienna University of Technology, Austria

 Springer

Editors

Prof. Henriette Engelhardt
University of Bamberg
Faculty of Social and Economic Sciences
Lichtenhaidestr. 11
96045 Bamberg
Germany
henriette.engelhardt-woelfler@uni-bamberg.de

Prof. Hans-Peter Kohler
University of Pennsylvania
Dept. Sociology
Population Studies Center
3718 Locust Walk
Philadelphia PA 19104
USA
hpkohler@pop.upenn.edu

Prof. Alexia Prskawetz
Vienna University of Technology
Institute for Mathematical Methods in Economics
Argentinierstr. 8/4/105-3
A-1040 Vienna, Austria
afp@econ.tuwien.ac.at

“Supported by the City of Vienna, Cultural Department, Science and Research Promotion”.

ISBN 978-1-4020-9966-3

e-ISBN 978-1-4020-9967-0

DOI 10.1007/978-1-4020-9967-0

Library of Congress Control Number: 2009920980

© Springer Science+Business Media B.V. 2009

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Contents

1 Causal Analysis in Population Studies	1
Henriette Engelhardt, Hans-Peter Kohler and Alexia Prskawetz	
2 Issues in the Estimation of Causal Effects in Population Research, with an Application to the Effects of Teenage Childbearing	9
Robert A. Moffitt	
3 Sequential Potential Outcome Models to Analyze the Effects of Fertility on Labor Market Outcomes	31
Michael Lechner	
4 Structural Modelling, Exogeneity, and Causality	59
Michel Mouchart, Federica Russo and Guillaume Wunsch	
5 Causation as a Generative Process. The Elaboration of an Idea for the Social Sciences and an Application to an Analysis of an Interdependent Dynamic Social System	83
Hans-Peter Blossfeld	
6 Instrumental Variable Estimation for Duration Data	111
Govert E. Bijwaard	
7 Female Labour Participation with Concurrent Demographic Processes: An Estimation for Italy	149
Gustavo De Santis and Antonino Di Pino	
8 New Estimates on the Effect of Parental Separation on Child Health	167
Shirley H. Liu and Frank Heiland	

9 Assessing the Causal Effect of Childbearing on Household Income in Albania 201
Francesca Francavilla and Alessandra Mattei

10 Causation and Its Discontents 233
Herbert L. Smith

Index 243

Contributors

Govert E. Bijwaard Netherlands Interdisciplinary Demographic Institute (NIDI), P.O.Box 11650, The Hague, The Netherlands, bijwaard@nidi.nl

Hans-Peter Blossfeld Faculty of Social and Economic Sciences, University of Bamberg, Lichtenhaiderstr. 11, Bamberg, Germany, hpb3007@t-online.de

Henriette Engelhardt Department for Population Studies, University of Bamberg, Lichtenhaiderstr. 11, Bamberg, Germany, henriette.engelhardt-woelfler@uni-bamberg.de

Francesca Francavilla University of Westminster, 50 Hanson Street, London W1W6UP, UK, f.francavilla@psi.org.uk

Frank Heiland Department of Economics and Center for Demography and Population Health, Florida State University, 32306-2180 Tallahassee, FL, USA, fheiland@fsu.edu

Hans-Peter Kohler Department of Sociology, University of Pennsylvania, Philadelphia, PA, USA, hpkohler@pop.upenn.edu

Michael Lechner Swiss Institute of Empirical Economics Research, University of St. Gallen, Varnbühlstr. 14, St. Gallen, Switzerland, Michael.Lechner@unisg.ch

Shirley H. Liu Department of Economics, University of Miami, P.O.Box 248126, Coral Gables, FL, USA, s.liu2@miami.edu

Alessandra Mattei Department of Statistics, University of Florence, Viale Morgagni, 59, Firenze, Italy, mattei@ds.unifi.it

Robert A. Moffitt Department of Economics, John Hopkins University, 3400 N. Charles St., Baltimore, MD, USA, moffitt@jhu.edu

Michel Mouchart Catholic University of Leuven, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium, michel.mouchart@uclouvain.be

Antonino Di Pino University of Messina, Via T. Cannizzaro, 278, 98100 Messina, Italy, dipino@unime.it

Alexia Prskawetz Vienna University of Technology, Institute for Mathematical Methods in Economics, Argentinierstr. 8/4//105-3, 1040 Vienna, Austria and Austrian Academy of Sciences, Vienna Institute of Demography, Wohllebengasse 12–14, 6th floor, Austria, afp@econ.tuwien.ac.at

Federica Russo Catholic University of Leuven, Place Cardinal Mercier 14, 1348 Louvain-la-Neuve, Belgium, federica.russo@uclouvain.be

Gustavo De Santis Department of Statistics, University of Florence, Viale Morgagni, 59, Firenze, Italy, desantis@ds.unifi.it

Herbert L. Smith Population Studies Center, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA, USA, hsmith@pop.upenn.edu

Guillaume Wunsch Institute of Demography, University of Louvain, Place Montesquieu 1/17, Louvain-la-Neuve, Belgium, Wunsch@demo.ucl.ac.be

Chapter 1

Causal Analysis in Population Studies

Henriette Engelhardt, Hans-Peter Kohler and Alexia Prskawetz

1.1 Introduction

An important hallmark of empirical research in population studies and demography has traditionally been a focus on careful description of population trends and changes using representative micro- or large-scale macro-data. For example, much effort has been devoted to describing the trends and variations in the core demographic processes – fertility, mortality and migration – and how the size and structure of a population are affected by these underlying processes. A core aspect of demographic methods therefore has been on the construction of vital rates, life-course measures of the tempo and quantum of demographic events, life table analysis and its extension to multi-state processes, and the decomposition of population differences in terms of rates and proportions (Vaupel 2001). Building on the methods and insights of these descriptive analyses, demographers have also developed sophisticated means for population projections (e.g. Lutz et al. 1999) and for investigating the relationships between mortality, fertility and migration in stable populations (e.g. Preston et al. 2001). In recent years, however, the field of population studies has grown increasingly diverse. While maintaining its traditional focus on formal demography (e.g. Feichtinger 1979), the discipline has strengthened its connections to other fields of science. As a result, demographers are increasingly adopting theories, concepts, and methods from sociology, economics, biology, medicine, anthropology, ecology, agriculture, geography, as well as mathematics, statistics, and econometrics, and demographic research increasingly addresses topics or questions that used to be within the domain of other disciplines.

With this broadening of view, a central aim of many research papers in population studies and demography is now to explain cause-effect relationships among variables or events. That is, demographic research is increasingly trying to address the causal mechanisms generating trends and variation in the core demographic processes of fertility, mortality and migration, and demographers are increasingly

H. Engelhardt (✉)
Department for Population Studies, University of Bamberg, Lichtenhaidenstr. 11,
Bamberg, Germany
e-mail: henriette.engelhardt-woelfler@uni-bamberg.de

interested in understanding how these processes, and their changes over time, affect other social and economic aspects of populations. The identification of genuine causes is thus increasingly viewed as essential for understanding the basic driving forces underlying demographic phenomena such as marriage, fertility, divorce, longevity, mortality and migration. Within this perspective, causal judgements are made to explain the occurrence of events and help to understand why particular events take place. However, it is important to realize that the causal understanding of demographic processes is not only important in its own right. In contrast, it is the understanding of the causal mechanisms that often facilitates the prediction of events or new observations in the future and allows for a certain amount of control over events. Moreover, causal knowledge is essential for demographers to provide effective and accurate policy recommendations.

Despite the recognised importance of identifying causes, relatively little attention has been paid by population researchers to considering what causality actually means and how knowledge of causes is acquired (in the sociological literature see, e.g. Abbott 1998; Goldthorpe 2001; Marini and Singer 1988). Reasoning is usually guided by an intuitive idea of causality, i.e., vague ideas of the possible determinant(s) of the event of interest, leading to attempts to consider directional relationships that are not merely spurious. However, the “causal” effects estimated in population studies (and in other social sciences) often do not provide much causal understanding, not only because of a faulty methodology but also because the attribution of hypothesised effects is merely based on an intuitive causality. Of late, there is an increased awareness in population studies and demography. For example, at the 2003 Annual Meeting of the Population Association of America (PAA) a symposium was held on causal analysis in the population sciences (see Bachrach and McNicoll 2003; Fricke 2003; Moffitt 2003, 2005; Smith 2003).

In the last two decades causality and causal inference has undergone a major transformation: from a philosophically overloaded concept to a mathematical object with well-defined semantics and a well-founded logic. For a long period, practical problems regarding causal inference with traditional regression based approaches were regarded as either metaphysical or unmanageable. These problems can now be reduced by using new statistical techniques (see e.g. Holland 1986; Sobel 1995, 1996; Winship and Morgan 1999). The contributions of statistics to causal inference are unquestionably most substantial when it comes to counterfactual analysis. The counterfactual account of causation pertains to establish the effects of causes (e.g. “What is the effect of womens’ labour force participation on fertility?”) instead of the causes of effects (e.g. “What accounts for the decline of fertility rates?”).

For decades, population scientists have concentrated their efforts on estimating the causes of effects by applying standard cross-sectional and dynamic regression techniques, with regression coefficients routinely being understood as estimates of causal effects. The regression approach to causality has loomed large in population research as well as in other social sciences because it seems to fit well with the way in which empirical social research proceeded. Much of population sciences proceeds by a researcher positing a causal theory of how and why a phenomenon occurs, im-

plying that the presence of attribute X causes an outcome Y. Data on observed values of X and Y are then collected. If a correlation between X and Y is observed, it is seen as supporting the causal theory. However, it does not demonstrate causality because there are other possible explanations of the correlation, notably Y might be causing X or a third (set of) attributes Z might be causing X and Y. Therefore, elaborating the causes of an observed effect is a theoretical or a philosophical problem. There is no epistemological basis for a statistical solution.

The standard approach to infer the effects of causes in natural sciences and in psychology is to conduct controlled randomised experiments. In population studies and most other social sciences, experimental designs are frequently infeasible and most research continues to be based on non-experimental designs (also called observational designs or survey designs). Social experiments are often too expensive or too challenging to implement and some experimental designs may be difficult to reconcile with the guidelines for ethical research. The subjects in randomized experiments may also be unwilling to follow the experimental protocol, and the treatment of interest may not be directly manipulable. While important exceptions exist, such as for example the Progressa Project in Mexico, experimental designs therefore remain the exception among demographic studies. In lieu of suitable randomized experiments, however, demographers, along with economists and other social scientists, have started to utilize quasi experiments or natural experiments to estimate treatment effects (e.g. the effects of social programs or public policy).

Unlike randomised experiments, quasi-experimental and non-experimental designs suffer from the problem of non-compliance. Inferring the effects of causes – or treatment effects – from other than experimental data is tricky (e.g. Manski 1995). All efforts to infer treatment effects from quasi-experimental data and observational data must confront the fact that the data are inherently censored. One wants to compare outcomes across different treatments or causes, but on the one hand each unit of analysis, whether a survey respondent or a quasi-experimental subject, experiences only one of the treatments. On the other hand individuals with different treatments or causes differ in many observable and unobservable respects which select them into treatment.

However, treatment effects can be inferred from non-experimental data with a counterfactual approach. Contemporary counterfactual analyses of causal effects are in strong analogy to experimental methods but are adaptable to observational studies as well. They start from the idea that the randomised experiment with perfect compliance and no missing data is the golden standard for estimating the causal effects of treatment interventions (or, generally, any other kind of causes). In this counterfactual perspective, causal effects are defined as the difference between the potential outcome disregarding of whether an individual had received a certain treatment (or experienced a certain cause) or not. The counterfactual approach to estimate effects of causes from quasi-experimental data or from observational studies was first proposed by Rubin (1974) in the context of what later became known as the Rubin causal model. Other important contributions to this approach include the work of James Heckman and his collaborators (e.g. Heckman et al. 1999) and that of Charles Manski and his collaborators (e.g. Manski and Nagin 1998).

These rapid strides of recent decades in econometrics and statistics give population scientists much to think about. They have surely moved population researchers closer to the elusive goal of discovering valid causal explanations of key demographic processes. At the same time, the advances have also stimulated greater attention to the limitations of high-powered statistical analysis – and, more fundamentally, current models of causality – in and for the social sciences. The importance of these contributions is also to caution against too great a faith in the promise of causal modelling of singular cause-and-effect-relationships in complex population processes. The ambition to understand social processes – implying a causal explanation – requires an indispensable eclectic mode, integrating findings of formal counterfactual analysis with those of more qualitative, interpretative research, however with both drawing on theory.

In December 2006 a conference with the theme “Causal Analysis in Population Studies: Concepts, Methods, Application” was organized by the Vienna Institute of Demography at the Austrian Academy of Sciences in collaboration with the University of Bamberg, Department of Population Studies and the University of Pennsylvania, Department of Sociology and Population Studies. The aim of the conference was to bring together leading researchers working on causal analysis in population sciences and to discuss the new developments in theoretical and statistical causality as well as empirical applications. A selection of the conference contributions, covering a wide range of research issues, is arranged in the present proceedings.

1.2 Structure of the Volume

In Chapter 2 Robert Moffitt presents his paper on “Issues in the estimation of causal effects in population research, with an application to the effects of teenage child-bearing”. This essay surveys the recent literature in economics on the estimation of causal effects, with a focus on the method of instrumental variables. Conditions for the validity of instrumental variable methods are discussed along with the proper interpretation of the resulting estimates. Several difficult issues with the method are outlined, including the problem of external validity, reconciling the differences in estimates when different instruments are used, and detecting instrument validity.

Chapter 3 by Michael Lechner is on “Sequential potential outcome models to analyse the effect of fertility on subsequent labour-market outcomes”. This paper proposes to use dynamic treatment models to analyze the effects of fertility on labour market interactions when large data sets are available. The main advantages are (i) its flexibility due to its nonparametric nature, (ii) its potential of allowing careful consideration of the selection issues coming from the dynamic interaction between fertility and labour market outcomes; and (iii) the possibility of defining relevant parameters of interest in a precise and detailed way. Based on artificial data that mimic important features of real data sets, the approach is implemented and issues that come up in any practical application of this approach are discussed.

Michel Mouchart, Federica Russo and Guillaume Wunsch write in Chapter 4 on “Structural modelling, exogeneity and causality” in the social sciences. They first present a conceptual framework where causal analysis is based on a rationale of

variation rather than on Humean regularities. They then develop a formal framework for causal analysis by means of structural modelling. Within this framework they approach causality in terms of exogeneity in a structural conditional model based on (i) good model fit, (ii) invariance under a large variety of environmental changes, and (iii) congruence with background knowledge and theory. They also tackle the issue of confounding and show how latent confounders can play havoc with exogeneity. Staying at the level of knowledge, this framework avoids making untestable metaphysical claims about causal relations and yet remains useful for cognitive and action-oriented goals.

In Chapter 5 Hans-Peter Blossfeld gives an overview of “Causation as a Generative Process. The Elaboration of an Idea for the Social Sciences and an Application to an Analysis of an Interdependent Dynamic Social System”. In the social sciences, two understandings of causation have guided the empirical analysis of causal relationships: (1) “Causation as robust dependence” and (2) “causation as consequential manipulation.” Both approaches clearly have strengths and weaknesses for the social sciences as described in detail in this chapter. Based on this discussion, a third understanding of “causation as generative process,” proposed by David Cox, is then further developed. This idea seems to be particularly valuable for modern social sciences because it can easily be combined with a narrative in terms of actor’s objectives, knowledge, reasoning, and decisions (methodological individualism). Using event history models, this approach is then applied to the causal analysis of an interdependent dynamic social system. Based on separate applications in West and East Germany, Canada, Latvia, and the Netherlands, the usefulness of the approach of “causation as generative process” is demonstrated by analyzing two highly interdependent family processes: entry into marriage (for individuals in a consensual union) as the dependent process and first pregnancy/childbirth as the explaining one. The authors then move to more substantive explanations, including the importance of actors, probabilistic causal relations, preferences and negotiation, observed and unobserved decisions and the problem of conditioning on future events.

Govert Bijwaard in Chapter 6 presents a paper on “Instrumental variable estimation for duration data”. In this article he focuses on duration data with an endogenous variable for which an instrument is available. In duration analysis the covariates and/or the effect of the covariates may vary over time. Another complication of duration data is that they are usually heavy censored. The hazard rate is invariant to censoring. Therefore, a natural choice is to model the hazard rate instead of the mean. Govert Bijwaard develops an Instrumental Variable estimation procedure for the Generalized Accelerated Failure Time (GAFT) model. The GAFT model is a duration data model that encompasses two competing approaches to such data; the (Mixed) Proportional Hazard (MPH) model and the Accelerated Failure Time (AFT) model. He discusses the large sample properties of this Instrumental Variable Linear Rank (IVLR) estimation based on counting process theory. He shows that choosing the right weight function in the IVLR can improve its efficiency and discusses the implementation of the estimator and applies it to the Illinois re-employment bonus experiment.

In Chapter 7 Gustavo De Santis and Antonino Di Pino analyse “Female labour participation with concurrent demographic processes: An estimation for Italy”. Fe-

male labour participation influences, but is also influenced by, several demographic processes (like fertility or couple formation and dissolution, for instance), and this endogeneity, if uncontrolled for, deprives regression analyses of most of their interest. Refinements that lead close to a causal analysis are possible using instrumental variables, and accounting for selectivity and treatment effects. In this paper, they show an application on a data source that is cross sectional, but with retrospective questions: the 2002 Bank of Italy Survey on Household Income and Wealth. The main results are: endogeneity exists and affects residuals, but its effects can be at least partly eliminated; well educated and less family-oriented women work more than others, as expected, but marked differences exist in this respect between the north and the south of Italy.

In Chapter 8 Shirley H. Liu and Frank Heiland present “New Estimates on the Effect of Parental Separation on Child Health”. This study examines the causal link between parental separation and the health status of young children. Using a representative sample of children all born to unwed parents drawn from the Fragile Families and Child Wellbeing Study (FFCWS), the authors investigate whether separation between unmarried biological parents has a causal effect on a child’s likelihood of developing asthma by age three. Comparing children with similar observable characteristics who differ only in terms of whether their parents separate, they find that parental separation increases the probability that a child develops asthma by age three by seven percentage points, relative to children whose parents remained romantically involved.

Chapter 9 by Francesca Francavilla and Alessandra Mattei is on “Assessing the causal effect of childbearing on household income in Albania”. This paper analyzes to what extent births may lead to changes in economic wellbeing. In contrast to most previous studies on this issue the authors apply appropriate econometric techniques based on longitudinal micro data in order to identify the causal effects of child bearing events on income. The analyses are performed on longitudinal data from the Albanian Living Standard Measurement Survey. They take a quasi experimental approach, that is, they consider the experience of a childbearing event as the treatment variable, and their measure of wellbeing as the outcome variable. In order to deal with the confounding due to the presence of systematic differences in background characteristics between the treatment groups, the authors first fit a multiple linear regression model that includes relevant background characteristics as well as an indicator variable for the treatment (i.e., childbearing). This estimation is then compared and contrasted with a matching approach, based on the bias-corrected matching estimator introduced by Abadie and Imbens (2002). The analysis suggests that there is some evidence that childbearing events can in fact increase household wellbeing in Albania. In addition, the treatment effect is highly heterogeneous with respect to observable characteristics such as the woman’s working status and the woman’s parity. All the results appear to be robust with respect to the estimated equivalence scale: changing the equivalence scale leaves the childbearing effect on income positive and non-significant.

The volume concludes with a critical discussion of causality in population studies by Herb Smith: “Causation and Its Discontents”. Here, the author puts forward his

thoughts against the enthusiastic view that causation is the most precious thing that population scientists possess or could acquire.

References

- Abadie, A. and G.W. Imbens (2002). Simple and Bias-Corrected Matching Estimators. *Mimeo*. Department of Economics, UC Berkeley.
- Abbott, A. (1998). The Causal Devolution. *Sociological Methods & Research* 27: 148–181.
- Bachrach, C. and G. McNicoll (2003). Causal Analysis in the Population Sciences: A Symposium. *Population and Development Review* 29: 443–447.
- Feichtinger, G. (1979). *Demographische Analyse und populationsdynamische Modelle: Grundzüge der Bevölkerungsmathematik*. Wien: Springer.
- Fricke, T. (2003). Culture and Causality: An Anthropological Comment. *Population and Development Review* 29: 470–479.
- Goldthorpe, J.H. (2001). Causation, Statistics, and Sociology. *European Sociological Review* 17: 1–20.
- Heckman, J.J., R.J. LaLonde and J.A. Smith (1999). The Economics and Econometrics of Active Labor Market Programs. In: *Handbook of Labor Economics*, Vol. 3A, eds. O. Ashenfelter and D. Card. Amsterdam: Elsevier, 1865–2097.
- Holland, P. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association* 81: 945–960.
- Lutz, W., J.W. Vaupel and D.A. Ahlburg (eds.) (1999). *Frontiers of Population Forecasting. Population and Development Review*, Suppl. 24.
- Manski, C.F. (1995). *Identification Problems in the Social Sciences*. Cambridge, MA: Harvard University Press.
- Manski, C.F. and D.S. Nagin (1998). Bounding Disagreements about Treatment Effects: A Case Study of Sentencing and Recidivism. *Sociological Methodology* 28: 99–137.
- Marini, M.M. and B. Singer (1988). Causality in the Social Sciences. *Sociological Methodology* 18: 347–409.
- Moffitt, R. (2003). Causal Analysis in Population Research: An Economist's Perspective. *Population and Development Review* 29: 448–458.
- Moffitt, R. (2005). Remarks on the Analysis of Causal Relationships in Population Research. *Demography* 42: 91–108.
- Preston, S.H., P. Heuveline and M. Guillot (2001). *Demography: Measuring and Modeling Population Processes*. Oxford: Blackwell.
- Rubin, D.B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology* 66: 688–701.
- Smith, H.L. (2003). Some Thoughts on Causation as it Relates to Demography and Population Studies. *Population and Development Review* 29: 459–469.
- Sobel, M.E. (1995). Causal Inference in the Social and Behavioral Sciences. In: *Handbook of Statistical Modelling for the Social and Behavioral Sciences*, eds. G.A. Arminger, C.C. Clogg and M.E. Sobel. New York: Plenum Press.
- Sobel, M.E. (1996). An Introduction to Causal Inference. *Sociological Methods & Research* 24: 353–379.
- Vaupel, J.W. (2001). *Analysis of Population Changes and Differences: Methods for Demographers, Statisticians, Biologists, Epidemiologists, and Reliability Engineers*. Unpublished Manuscript, Rostock: Max Planck Institute for Demographic Research.
- Winship, C. and S.L. Morgan (1999). The Estimation of Causal Effects from Observational Data. *Annual Review of Sociology* 25: 659–706.

Chapter 2

Issues in the Estimation of Causal Effects in Population Research, with an Application to the Effects of Teenage Childbearing

Robert A. Moffitt

2.1 Introduction

Population research has a distinguished history of empirical work on a wide variety of important topics related to population growth, the components of demographic trends, estimation of vital rates, life table construction, investigation into causes of historical population developments, and many others. However, one branch of population research that has seen increasing interest has been in the area of social demography, where the determinants of individual behavior regarding fertility, marriage, and related areas has been studied. It is in that branch that issues of causal inference have arisen, and with which this essay is concerned.

This development in population research coincides with a more general interest in causal inference in statistics and in other social sciences such as economics. In statistics, the development of the Rubin Causal Model (Rubin 1974) as a framework for studying causal questions has become a dominant paradigm even though, at the same time, there is considerable work by other statisticians using somewhat different frames. The development in statistics, while having many historical antecedents in the field, by and large occurred only in the 1970s and 1980s. Prior to that time, randomized experiments were regarded as the only method for true causal inference. However, randomized experiments are generally not possible in many fields, including population research, and hence methods for the analysis of causation using observational data are needed.

In economics, while causality has a much longer history dating to the development of the simultaneous equations model, which saw its fullest development in the Cowles Commission work in the 1950s, renewed interest in the issue has arisen since the 1980s and 1990s as more subtle issues have been addressed. Other social science disciplines such as sociology and political science are following the developments in statistics and in economics, with new developments adapted to their unique sets of questions and issues.

R.A. Moffitt (✉)

Department of Economics, John Hopkins University, 3400 N. Charles St., Baltimore, MD, USA
e-mail: moffitt@jhu.edu

This essay will review the issues in causal modeling using primarily the framework adapted in economics, and will apply those modeling issues to the study of population questions. The economics framework is, when boiled down to essentials, observationally equivalent to the Rubin Causal model in statistics, although the interpretation and language used to describe the two are quite different. In addition, their practical empirical implementation is often quite different, with economists leaning toward modeling by the use of regression equations with explicitly represented error terms, an approach different from that in statistics.

While the causal modeling developments in economics have taken many directions, the vast majority of applications in the field use the method of instrumental variables (IV) to estimate causal effects. Therefore, this essay will also concentrate on that method, outlining both its rationale and advantages and the pitfalls and weaknesses associated with its use. Brief mention will be made of other methods such as panel data fixed effects methods and matching.

The running example in the essay is the question of whether teenage childbearing has a deleterious effect on female economic outcomes such as income and earnings. The increase in rates of teen childbearing in the U.S. has been a source of public concern not only because much of that childbearing is nonmarital but also because of the widespread perception that women who begin their childbearing at a very young age run the risk of harming their educational progress and their later economic and social success. There has been a great deal of research on this issue with, surprisingly, much less support for this conventional view as might be expected. But the literature has also generated much discussion of the method of causal effects and of the effects of using different instruments for estimation. Thus, this particular literature can be used to illustrate a number of the issues in causal modeling in population research in general.

The first section below lays out the general causal model in economics and discusses a number of the main issues. The method of instrumental variables is then outlined, followed by a categorization of the types of instruments most often used. Additional issues in the use of instrumental variables are then reviewed, followed by a set of conclusions. Some of the points made in the essay are (i) a tradeoff between internal validity and external validity is often faced by analysts using the method; (ii) multiple instruments or instruments with multiple values can be used to learn more about effects in heterogeneous populations than binary instruments; and (iii) use of theory is important to determine mechanisms by which treatments affect outcomes and how differing instruments interact with those mechanisms.¹

2.2 The Basic Causal Model

The basic causal model in economics dates to the Cowles Commission work on simultaneous equations in economics and, later, its adaptation to individual actions represented in the switching regression model (Heckman 1978; Lee 1979).

¹ See Moffitt (2003, 2005) for earlier reviews.

Heckman and Robb (1985) and Björklund and Moffitt (1987) made the connection between that model and newer thinking in causal modeling as well as introducing the notion of heterogeneity to be discussed momentarily. Heckman et al. (2006) provide a recent overview of the model.

The prototype linear regression model used in this literature is

$$y_i = \alpha_i T_i + X_i \beta + \varepsilon_i \quad (2.1)$$

$$T_i^* = X_i \gamma + Z_i \delta + v_i \quad (2.2)$$

$$T_i = \begin{cases} 1 & \text{if } T_i^* \geq 0 \\ 0 & \text{if } T_i^* < 0 \end{cases} \quad (2.3)$$

where y_i is the value of the outcome for individual i , T_i is a dummy variable for whether an individual has received the “treatment,” α_i is the effect of the treatment on y for individual i , X_i is a vector of exogenous covariates and β is its associated coefficient vector, Z_i is a vector of exogenous variables affecting the probability of receiving treatment but which do not affect y directly, and ε_i and v_i are mean-zero error terms. This two-equation model, consisting of an outcome equation as a function of treatment and a second, selection equation (Equations (2.2) and (2.3) together) representing the determinants of treatment, is a special case of the general switching regression model. The selection equation is often not written down explicitly in the studies in the literature, and does not have to have the latent index structure shown in (2.2) and (2.3), but this is the most common interpretation. In addition, when estimating some of the objects of interest such as the marginal treatment effect or the local average treatment effect discussed below, an explicit representation of the selection equation is particularly helpful in interpretation.

The most important difference between this model and the classic linear simultaneous equations model in economics is that the effect of receiving treatment on the outcome (α_i) varies across individuals and hence treatment effects are “heterogeneous” in the population. In the older literature, this effect was assumed to be fixed. Allowing the effect to be heterogeneous has critical implications for interpretation and estimation. Most obviously, it requires a reconsideration of the object of estimation. The parameter α_i has a distribution in the population and one could imagine attempting to estimate different features of that distribution. One could attempt to estimate the mean of α_i , $E(\alpha_i)$, commonly called the average treatment effect, which is the average effect on y if all women in the population had a teen birth. Or one could attempt to estimate the average α_i for a subset of women observed, in a particular sample, to have had a teen birth. This object is $E(\alpha_i | T_i = 1)$ and is called the effect of the treatment on the treated. Two other possible objects of interest, the marginal treatment effect and the local average treatment effect, will be discussed below.

The assumptions on the covariance matrix of the error terms are that $E(\varepsilon_i v_i) \neq 0$ and $E(\alpha_i v_i) \neq 0$.² If y is individual earnings at some age like 25 and T is a dummy

² Conditioning on X and Z is left implicit.

variable for whether a woman had a teenage birth, $E(\varepsilon_i v_i) \neq 0$ if, for a group of women who have the same X , those women who had a teen birth would have had different future earnings than those women who did not, even if they had not had a teen birth. For example, those who had a teen birth may have come from disadvantaged backgrounds in unobserved ways. If, on the other hand, those who had a teen birth have a lower (i.e., less negative) impact of such a birth on future earnings than that of women who did not have a teen birth, then $E(\alpha_i v_i) \neq 0$. For example, women from disadvantaged backgrounds, who are more likely to have a teen birth, may be less affected by having a teen birth than women from less disadvantaged backgrounds because they have a lower payoff to human capital investments in the first place.

The implications for OLS of the two covariance assumptions are different. If $E(\varepsilon_i v_i) \neq 0$, OLS is biased for any object of estimation one might be interested in (i.e., any feature of the distribution of α_i).³ OLS compares the y of women who had a teen birth to the y of women who did not, and this comparison is faulty because the y of women who did not have a teen birth does not equal the value of y for the women who did have a teen birth, if they had not had one. If $E(\alpha_i v_i) \neq 0$ but $E(\varepsilon_i v_i) = 0$, however, OLS returns a coefficient on T_i of $E(\alpha_i | T_i = 1)$, the effect of the treatment on the treated; but this is biased for the average treatment effect, if that is the object of interest. If both $E(\varepsilon_i v_i) \neq 0$ and $E(\alpha_i v_i) \neq 0$, again OLS produces nothing of interest.

Equation (2.1) is formulated as a linear regression function with additively separable X and with the vector of variables within X assumed to have effects through a linear index, $X\beta$. Thus, no nonlinearities in X or interactions between X and T are allowed. These restrictions can be relaxed by introducing such interactions into the model. Alternatively, a fuller nonparametric method could be used which allows T , and X to enter the model in arbitrary ways, although those methods typically lack power unless sample sizes are large. The method of matching is in this family, for matching is a method which nonparametrically estimates the effect of T on y , allowing X to affect y in an arbitrary fashion and allowing arbitrary interaction of T with X . However, the matching method is, explicitly or implicitly, an OLS method with an extended set of nonlinearities and interactions added, and hence requires that the error term in (2.1) be unrelated to v_i (see Imbens (2004) for a review).⁴ More formally, the assumption necessary for the method to produce unbiased estimates is the assumption of conditional independence, $E(\varepsilon_i | X_i) = 0$. Thus the method of matching is designed to address a different question than the standard selection model; the latter is designed to address the problem of unbiased estimation when selection is on unobservables, while the former is designed to address the problem

³ OLS produces inconsistent and biased estimates because T is an “endogenous” independent variable, defined as one which is correlated with the error term in the equation even after controlling for X .

⁴ That is, a fully nonparametric regression of y on X and T is equivalent to matching except that matching also imposes a “balancing” requirement to make the distributions of X in the $T = 0$ and $T = 1$ samples the same, whereas the nonparametric regression would not impose this.

of nonlinearities and interactions in the functional form of Equation (2.1) when selection is only determined by observables in the data (X).⁵

2.3 Instrumental Variables

The method of instrumental variables (IV) is designed to address the problem of selection on unobservables. It relies on the existence of the vector Z and, in its traditional two-stage least-squares form, simply involves first regressing T on X and Z and then regressing y on X and the predicted T from that first-stage equation. However, it can be equivalently represented in other forms which provide more intuition for what is being done and what is being estimated. To illustrate this point, let us simplify the model in (2.1)–(2.3) by omitting the control variables X and assume that Z is represented by a single binary (dummy) variable:

$$y_i = \beta_0 + \alpha_i T_i + \varepsilon_i \quad (2.4)$$

$$T_i^* = \gamma_0 + \delta Z_i + \nu_i \quad (2.5)$$

$$T_i = \begin{cases} 1 & \text{if } T_i^* \geq 0 \\ 0 & \text{if } T_i^* < 0 \end{cases} \quad (2.6)$$

For consistent estimation, the instrumental variable Z_i is assumed to have three properties: $E(\varepsilon_i | Z_i) = 0$, $E(\alpha_i | Z_i) = \bar{\alpha}$, and $E(T_i Z_i) \neq 0$.⁶ The first two are “validity” restrictions that require that the instrument be mean-independent of the two unobservables in the y equation. The third is a “relevance” restriction which requires that the instrument be correlated with the probability of receiving treatment; the instrument must be “relevant.”

When studying the effect of teenage fertility on later earnings, for example, the availability of contraceptives in a woman’s geographic area might satisfy these conditions. For example, contraceptives might be more available in one area than another because of different governmental decisions to provide them. Those governmental decisions may be unrelated to the unobserved earnings levels of the women in each area (ε_i) or to the effects of teen childbearing on earnings of those women (α_i). The assumption would fail, however, if governments provided more contraceptives when the women in the area are more economically disadvantaged. Availability of contraceptives is likely to be relevant since it presumably affects teen childbearing.

A true experiment would guarantee that the conditions were met. Suppose that Z is a dummy for whether a women is randomly assigned to an experimental group or

⁵ In principle, there is no reason that attention to nonlinearities and interactions in Equation (2.1) cannot be addressed at the same time as addressing selection on unobservables.

⁶ These assumptions are stronger than what is needed and can be relaxed, but make the exposition particularly simple.

a control group. The experimental group is offered additional contraceptives while the experimental group is not (this is called an “offer” experiment; not all women in the experimental group use the contraceptives). This experiment should affect the treatment variable T , i.e., whether a birth occurs, and hence Z should be relevant. Because of the randomization, Z should be unrelated to the earnings functions of the women in the experiment in the absence of the treatment, i.e., the women in the experimental and control groups should be identical in all observed ways (ε_i and α_i). Thus Z should also be valid. The search for a valid and relevant Z in observational data is essentially a search for a “natural experiment” where Z is effectively randomly assigned to different groups within the population.

An important question is whether there is any way to test whether a particular variable is a suitable instrument and meets the criteria of validity and relevance. Relevance can be ascertained to some degree by determining how significant and strong a determinant of T a particular variable is, either by examining the t-statistic or F-statistic for the coefficient on Z when estimating T as a function of Z . Validity, however, cannot be tested if there is only one Z being examined (the case of a “just-identified” model). To do so requires that a sample estimate of the covariance between Z and ε (in a simple linear model) be obtained, and it is not possible to obtain a consistent estimate of ε without the assumption that Z is valid in the first place. Determining whether a variable is a valid instrument therefore requires appeals to theory and priori arguments for why Z is likely to be randomly assigned, usually based on arguments for why the determinants of Z are likely to be uncorrelated with the unobservable determinants of y , as well as empirical investigations into the observable determinants of Z which, while not constituting proof of its lack of correlation with unobservables, nevertheless can give clues to the process of its determination.

When Z is a dummy variable, the two-stage least-squares estimation of the model yields a coefficient on T in the y equation equal to:

$$\hat{\alpha}_{IV} = \frac{\bar{y}_1 - \bar{y}_0}{\bar{T}_1 - \bar{T}_0} \quad (2.7)$$

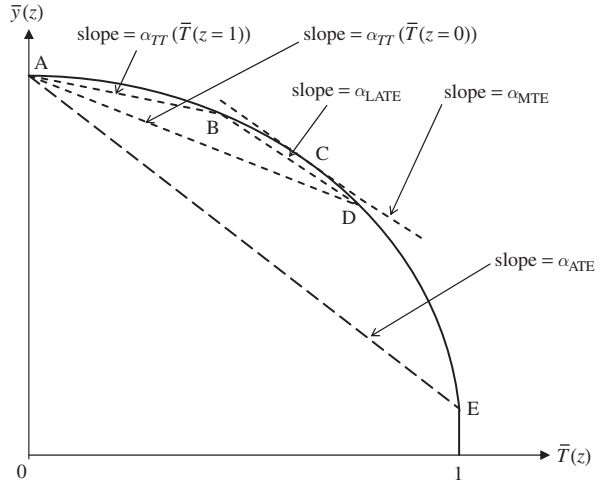
where \bar{y}_1 is the mean of y_i over all the observations for which $Z_i = 1$, \bar{y}_0 is the mean of y_i over all the observations for which $Z_i = 0$, \bar{T}_1 is the mean of T_i over all the $Z_i = 1$ observations, and \bar{T}_0 is the mean of T_i over all the $Z_i = 0$ observations. This formulation of the IV coefficient, emphasized by Imbens and Angrist (1994) and termed the local average treatment effect (LATE) for reasons to be explained below, is instructive in understanding how IV effects are estimated. In the case of a binary Z , the numerator of (2.7) represents the difference in the value of y for those in the “experimental group” ($Z = 1$) and those in the “control group” ($Z = 0$). In each of these groups, the mean of y is a weighted average of those with $T = 1$ and $T = 0$ in the group. The denominator represents the difference in the fraction of each group which “participates” by “taking up” the offer of treatment.

While Equation (2.7) involves only overall means in the two groups, it can be interpreted as representing the change in y for the subset of observations who changed

their value of T as a result of the differing Z . For example, if contraceptives are more available in one area ($Z = 1$) and less available in another ($Z = 0$), the denominator represents the difference in the fraction of women who have a teen birth because of greater contraceptive availability. If the instrument is relevant, this change will be nonzero, that is, contraceptive availability affects teen birth probabilities. The numerator represents the difference in mean earnings in the two areas and is assumed to be solely a result of the change in teen births; this follows from the assumption of validity, which implies that earnings in the two areas are identical aside from Z and hence any difference in earnings can be ascribed to the effects of differing Z . The change in earnings has to be “inflated” by the change in the teen birth fractions between the areas. So, for example, if the difference in average earnings between the areas is \$1,000 per year and the change in the teen birth fraction as a result of increased contraceptive availability was -0.10 , then the change in y for the 10 percent of women who changed their birth behavior must have been \$10,000. That is because 90 percent of women did not change their value of T at all, and hence must have had no change in their y .

This formulation of the IV method when the instrument is binary underscores the limited nature of what has been learned. The effect being estimated is the average effect of treatment, such as having a teen birth, for the 10 percent of women who switched from not having a teen birth to having a teen birth (called “switchers” or, in the language of Angrist et al. 1996, “compliers”). The effect of having a teen birth on outcomes cannot be assumed to be the same for any other group. For example, if the difference in contraceptive availability in the two areas in the data changes the teen birth rate from 0.30 to 0.20, its effect on earnings cannot be assumed to equal what would happen if the birth changed from 0.20 to 0.10, or for any other rate. If α_i is heterogeneous, as is assumed here, then the effects of teen birth on earnings will differ for different groups. If, for example, those women who have teen births when there are few contraceptives available include many for whom the negative consequences on earnings are particularly large, then expanding contraceptive availability and lowering the birth rate may have a gradually smaller and smaller effect on earnings, as those who continue to have birth rates even when contraceptive availability is high are those who have the smallest negative consequences for earnings.

This case is illustrated in Fig. 2.1. The mean of earnings in an area is plotted against the teen birth rate in the area as line ABCDE. As the birth rate rises, mean earnings fall as a larger fraction of the population has a reduction in earnings as a result of a teen birth. However, the figure is drawn such that the slope of the curve increases (becomes more negative) as the fraction with a teen birth rises, as would be the case if those who have the largest negative earnings reductions are the “last” ones to have a teen birth. The LATE estimate of the effect of teen births on earnings when the binary instrument lowers the teen birth rate from $\bar{T}(Z = 0)$ to $\bar{T}(Z = 1)$ is the slope of the dotted line connecting points B and D. It is termed the local average treatment effect because it applies only to a “local” area of the curve (that between B and D) and is only an average of those who change teen birth behavior in that region. Obviously, a LATE estimate will differ depending on where the two points

Fig. 2.1 Treatment effects

induced by the instrumental variable are on the curve and, in this sense, there is no longer a single effect of treatment on outcomes; one cannot speak of “the” effect of teen births on earnings, for “the” effect depends on the population affected. The LATE estimate will not differ across ranges of fractions treated only if the curve in Fig. 2.1 is a straight line, with constant slope.

The other two possible objects of interest mentioned above, the average treatment effect (ATE) and the effect of the treatment on the treated (TT), are also shown in the figure. The ATE is the slope of the line connecting points A and E, representing the change in earnings that would occur if the entire population went from no teen births to all teen births. Two values of the TT are shown by the slopes of the dotted lines AB and AD, which show the difference in earnings that would arise if the fraction of women having a teen birth in the population were $\bar{T}(Z = 1)$ or $\bar{T}(Z = 0)$, respectively, rather than no teen births. The IV estimate obtained from a binary instrument Z will not equal a TT unless one is lucky enough to have an instrument for which $\bar{T} = 0$ (e.g., an area where contraceptives are so available that the teen birth rate is almost zero) and will not estimate the ATE unless one is even luckier to have an instrument for which the teen birth rate is zero for some values of the instrument and is equal to one (everyone has a teen birth) for other values of the instrument. In most applications, this is extremely rare.

Learning only the effect of the treatment in a local area may not be limiting if the other areas of the curve are not particularly relevant. No populations or subpopulations defined by demographic groups have teen birth rates close to 100 percent, so the lack of knowledge in that region of the population may not be discouraging. If the instrument in question moves the teen birth rate over a region which is very similar to that of most other populations and periods one is interested in, it may not be very disadvantageous.

In some cases, the researcher may be willing to extrapolate beyond the data available and draw conclusions about the effect of treatment on other populations.

Such extrapolation is a special case of the problem of “external validity” originally discussed in the context of classical experiments, where the question is whether the results of an experiment can be generalized beyond the specific type of program examined and beyond the specific type of population enrolled in the experiment. Extrapolation is always possible, but it is necessary to be clear that to do so requires additional assumptions (e.g., on the shape of the curve outside the range of the data) than were necessary to obtain the initial estimates, and a clear separation between the two needs to be made.

Learning the effect of teen births over more points on the curve requires more instruments or more variation in a single instrument. Sometimes this can arise across studies using different instruments, in which case one could imagine piecing together the curve from different investigations. Alternatively, if multiple instruments or instruments with more than a two values are available, the effects can be estimated over greater portions of the curve. Carneiro et al. (2006) and Moffitt (forthcoming) have considered estimation in these circumstances, both noting that Heckman and Vytlacil (2005) showed that the curve in Fig. 2.1 can be represented by the regression equation

$$y_i = X_i\beta + g [T (Z_i)] + \varepsilon_i \tag{2.8}$$

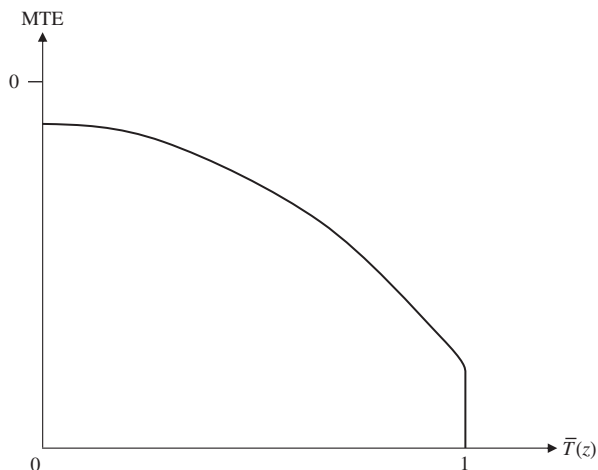
where a vector of other variables (X_i) has been reintroduced. The function g is intended to represent the curve ABCDE in Fig. 2.1, and can be specified as a polynomial, piecewise-linear function, or estimated completely nonparametrically.⁷ To estimate it requires a first-stage estimate of $T(Z_i)$, which is the same as the first-stage estimation in traditional two-stage least-squares estimation of these models. The predicted probability of $T = 1$, as a function of Z , is then inserted into (2.8) and estimation can proceed.

Heckman et al. and Moffitt, following on the terminology of Björklund and Moffitt (1987) and Heckman and Vytlacil (2005) term the slope of the “ g ” function the “marginal treatment effect” (MTE). It is simply the slope of the curve in Fig. 2.1, and will vary over the range of \bar{T} . The slope of the dotted line at point C in Fig. 2.1 shows the MTE at that point, and Fig. 2.2 shows the MTE over the entire range. The MTE represents the effect of the treatment on the outcome for the marginal person “brought into” treatment by a small increase in the fraction treated. Estimates of Equation (2.8) may reveal, on the contrary, that y is linearly related to $T(Z)$, in which case the curve in Fig. 2.1 is a straight line and it can be concluded that there is no heterogeneity of response in the population.

As before, how much can be learned from such an exercise depends on the range of \bar{T} induced by the range of instrumental variables in the data. If the entire range of \bar{T} from 0 to 1 is not induced, only a portion of the curve can be estimated. Nevertheless, the general lesson is clearly that instruments which induce a wider

⁷ Carneiro et al. proposed applying the partial-linear regression model to obtain nonparametric estimates, while Moffitt proposed applying series methods.

Fig. 2.2 Marginal treatment effect



range of fractions treated are more desirable than those which induce a smaller range.

If multiple instruments or instruments with more than two values are available, application of IV (e.g., in its traditional two-stage least squares form) to Equation (2.1) without allowing for nonlinearity of $T(Z)$ will produce a single coefficient which is a weighted average of the different MTE's over the range of the data (Angrist and Imbens 1995; Angrist and Krueger 1999; Heckman and Vytlačil 1999). This weighted average can be more difficult to interpret than either the LATE for a binary instrument or the varying MTE estimated in the Heckman-Moffitt approach because it is unclear what range of population it applies to. Nevertheless, it roughly characterizes the average effect in a loose sense.

2.4 Types of Instrumental Variables

The range of types of instrumental variables used in the literature in economics, demography, and other fields is very wide, and hence any attempt to group them into different types must necessarily be only approximate. But with this caveat in mind, the large majority of instruments can be classified into one of four types: cross-sectional ecological variables including area fixed effects instruments, population-segment fixed effects instruments, sibling and related instruments, and a residual category of “natural experiment” variables.

Cross-sectional ecological variables are variables which affect the environment in which individuals make choices and are most often measured at the aggregate level, most commonly for a geographic area. The variable for availability of contraceptives discussed in the teen birth case is an example of this type of instrument (Klepinger et al. 1999). Similar instruments used in this literature are state restrictive abortion laws and state family planning services (Klepinger et al. 1999) and the availability

of gynecologists in the individuals' area (Ribar 1994). More generally, variables measuring differences in laws, labor markets, social structure, prices and availability of services (e.g., child care) in an area are used. The argument for the validity of such instruments is that they are "external" to the individual's own behavior, and therefore can be argued to be unrelated to the individual's individual determination of his or her outcome, y . Unlike individual-level characteristics such as family background or related measures, for example, which are quite likely to be direct determinants of y and hence not excludable, the higher-level instruments could be argued to not directly appear in the y equation.⁸

The most common objection to ecological variables is the well-known problem of unobserved ecological correlations. One common type of correlation arises when individuals who live in different areas are different in unobserved ways, as might arise through residential sorting. Another is that, even if residential sorting does not take place, unobserved area-level factors may be present which cause differences in individual outcomes across areas, even holding individual characteristics fixed. However, both of these problems cause difficulties only if the unobserved differences in question are correlated with the area-level instrument being employed. To consider this requires investigating the determinants of the instrument and why its value varies across areas. For example, if the availability of contraceptives is partly affected by the level of teen births in an area, implying that any unobservable affecting teenage fertility is correlated with contraceptive availability, the instrument will be invalid. In many cases, the political process (e.g., in the case of laws) has to be considered and, in some studies, a rather detailed study of the reasons for passage of a particular piece of legislation is provided in order to prove that the reasons for passage were purely "political" and not related to the value of y in the area.

In many cases, objections to the validity of ecological instruments on the basis that those instruments are correlated with area-level unobservables can be addressed if the instruments change in value over time differently in different areas, and if data are available on samples of individuals over time as well. In this case, an area fixed-effects model can be estimated, most simply by estimating the model in first-difference form.⁹ In this case, the main equation is formulated as one in which Δy is assumed to be a function of ΔT , and the instrumental variable used is ΔZ .¹⁰ Estimation in this form will eliminate any area-level unobservables that are fixed over time, which will be differenced out. For example, examining how the change in earnings over time across areas is related to the change in contraceptive availability,

⁸ The multi-level model in social statistics bears a relationship to ecological instruments but is aimed at a different problem, for the multi-level model is aimed at obtaining correct and efficient standard errors rather than addressing the problem of endogeneity of a treatment variable.

⁹ As in the standard fixed effects model, however, the "within" estimator is more efficient than the first-difference estimator.

¹⁰ If panel data are available, this model can be estimated directly. If only repeated cross-sections are available, the model has to be formulated slightly differently, with the waves of the data pooled and an interaction term between time dummies and T and Z entered.

working through changes in teen birth rates, may be a more reliable method. Once again, however, one must carefully consider why contraceptive availability changed in different ways across areas, to insure that it did not change in response to trends in the teen birth rate in the area in question.

It is worth noting that estimation of the panel data fixed effects model without adjustment for endogeneity with instrumental variables is no longer favored within economics, whereas it was initially thought by some researchers to be an acceptable solution to the problem. Simply examining whether a change in an independent variable (T) for an individual is correlated with a change in outcome (y), which will eliminate individual-level fixed effects, is now thought to be problematic because individuals change their actions (T) over time for reasons that are usually related to changes in their situation that are also affecting y . It is still necessary to find some determinant of the change in T that can be more plausibly argued to be unrelated to the forces at the individual level that are driving changes in y .

A second type of instrument can be termed, for lack of a better term, the population-segment fixed effects variable. In this case, changes in outcomes y for two groups which experience different changes in T and Z are compared, and the difference in those changes are assumed to be a result of the changes in T and Z which also occur over time. The difference with the area fixed effects model is that the groups are not defined by geographic location but rather by demographic or economic group. For example, suppose that contraceptives are more or less freely available to all higher-income and more-educated groups at all times, but become more available to lower-income families over time (in the nation as a whole; no geographic variation is assumed to occur). In that case, the variation in the change in Z over time arises from differences by income or education group. The validity of the instrument depends on the accuracy of the assumption that both groups would have had the same change in their earnings (controlling for changes in observables, X) in the absence of a change in contraceptive availability for the lower-income, less-educated group. Put differently, the assumption for validity is that the earnings of both groups are trending at the same rate in the absence of the change in Z . This may not be the case if there are other time-trending unobservable factors affecting the two groups differently.

There have been no applications of this method in the teenage birth case, but it has been used frequently in studies examining nationwide changes in governmental policy such as changes in the welfare system (Moffitt and Ver Ploeg 2001, Chapter 4). There, changes in earnings, fertility, or other outcomes over time for groups primarily affected by welfare reform (e.g., less-educated single mothers) are compared to those changes for groups presumably unaffected or less affected (more educated single mothers, single childless women, married mothers, men). The assumption that the different demographic groups, or population segments, trend at the same rate in terms of earnings, fertility, and other outcomes would appear to be a very strong one both because so many other social, economic, and political forces are typically changing over time that may affect the groups differently but also because, in some cases, the policy in question may affect the characteristics that define the groups (e.g., marriage or fertility).

A third type of instrumental variable used frequently in recent years is based on sibling or twin differences. Assuming data are available on a sample of individuals, some of whom are twins or siblings, Equation (2.1) can be estimated on the pooled sample of all individuals. The instrument in this case is $(T_{if} - \bar{T}_f)$, where T_{if} is the treatment value for individual i in family f and \bar{T}_f is the average of T_{if} over all individuals in family f . Thus the instrument is the deviation of each individual's T from the family-specific mean. The assumption needed for validity in this case is that that deviation is independent of the deviation of each individual's ε from its family-specific mean. In the teen childbearing case, for example, the necessary assumption is that the fact that one sibling has a teen birth and another does not is unrelated to their future earnings or any determinant of future earnings such as ability, motivation, or other factors related to economic success. What the mean eliminates as a source of problem are the differences that arise from different levels of disadvantage across families, which is almost surely related both to teen childbearing and to later earnings. Nevertheless, the assumption remains a very strong one, for there are well-known differences in sibling development and in parental treatment of siblings that could cause the necessary assumption to fail (for a discussion of these issues by economists, see Bound and Solon (1999)).

Even though the model can be estimated by IV, most often it is estimated instead in reduced form. Assuming that Equations (2.2) and (2.3) are “substituted” in for T_i in Equation (2.1) (ignoring the nonlinearity involved in the substitution), a reduced form equation is obtained specifying y_i as a function of X_i and Z_i . Using the within-family deviation on T for Z , OLS estimation of such an equation is equivalent to estimating the model in within-family differences (that is, regressing the within-family deviation of y on the similarly-defined X and Z). Geronimus and Korenman (1992) were the first to apply this method in the teen birth literature, and found it to have a large effects on the results relative to OLS. Hoffman et al. (1993) and Ribar (1999) have provided further discussion of the method and the results.

A fourth category of instrument, really a residual category comprising several different types of approaches, is the use of “natural experiments” as instruments. The term refers to occurrences of “random” events that arise in “nature” (that is, not in a controlled laboratory setting) which can be arguably unrelated to unobservables in many individual outcomes (Angrist and Krueger 2001). In fact, defining the term this broadly would include all the other instruments already discussed, so the term in this case is more narrowly defined. One type included here is really a subset of the area fixed effects model and is applied whenever there is a law or policy change that applies to a narrow group of the population in very similar circumstances. For example, a law which affects only children in a particular age range and in a particular income in one state but not another states could be arguably unrelated to various later outcomes for the children (see Currie and Gruber (1996) for a related example). This differs from the general area fixed-effects model only by virtue of using a much narrower demographic and geographic segment of the population. Another category is what Rosenzweig and Wolpin (2000) have called “natural natural experiments” which arise when a possibly random demographic event occurs such as the birth of

twins, the month of the year in which a child is born, or whether a miscarriage occurs for a woman who is pregnant. The birth of twins has been argued to cause a random increase in the number of children and hence may be arguably used to estimate the effects of childbearing on a wide range of outcomes (e.g. Angrist and Evans 1998). The month in which a child is born has been argued to affect when a child can enter school and when a child is legally eligible to drop out of school, and hence is arguably a random determinant of years of education (Angrist and Krueger 1991). Miscarriage can be argued to also randomly affect fertility, or at least its timing, and has been used as an instrument for the likelihood of a teen birth (Hotz et al. 1997a, b). The age at menarche is another possibly random variable which should affect the age at which a first birth can occur and hence the probability of a teen birth (Ribar 1994; Chevalier and Viitanen 2003). Yet another type of instrument which may be termed a natural experiment are instruments based on so-called regression discontinuity designs (Cameron and Trivedi 2005; Imbens and Lemieux, forthcoming). This approach makes use of cases where there is an important variable which affects outcomes, and therefore is not excludable, but there also exists a policy or other event which generates a discontinuous change in that variable at a discrete point in the range. For example, if a family planning organization provided free contraceptives and other services strictly only to those with incomes below the poverty line, then a comparison of the outcomes of families just above and just below the poverty line should come close to measuring the effects of free contraceptives, other things being equal, because the two groups of families are “almost” identical in terms of income.

The validity of natural experiment instruments must be considered on a case-by-case basis, for there are very few generalizations possible. In the case of differing laws across states, the same issues have to be considered as in the general area fixed-effects model. For the natural natural experiments, threats to validity are often based on the suspicion that the demographic event in question has direct effects on the outcomes of interest. For example, a miscarriage may affect the mother’s attitudes toward future fertility or her educational and economic outcomes directly, and not simply through the consequent postponement of fertility. There is also the possibility that miscarriage is related to underlying physiological or health factors that might be related to later economic success. Likewise, age at menarche might be related to unobserved health factors. Some demographers have argued that month of birth is correlated with other variables affecting the timing of fertility and the probability of conception, and therefore may have some indirect correlation with later educational outcomes, for example.

A rather different concern with natural experiments is that they are rather limited in their external validity. The narrow-population law differences necessarily apply only to a small segment of the population, for example, whose effects may not generalize to other parts of the population or to other laws. The regression discontinuity design necessarily can estimate impacts only for those around the point of discontinuity, e.g., only those with incomes just around the poverty line. Again, those effects may not generalize to other groups in the population. These issues of external validity are in addition to those discussed in the prior section related to varying MTE over the range of fractions population treated; the issue here is simpler and

more traditional because the groups whose effects are estimated are characterized by standard socioeconomic observables (age of the child, family income, etc).

The issue of external validity raised in the natural experiment literature also serves to illustrate a seeming tradeoff between internal and external validity. Internal validity, defined in this case to be the validity of an instrument in an IV context, is attempted in the natural experiment literature by focusing on narrowly defined groups for whom it is plausible not only that observables are equivalent but also unobservables. But maximization of internal validity comes at the cost of sacrificing external validity, i.e., generalizability. It is possible that a “partially valid” instrument applied to a larger population could generate estimates which are biased but have acceptable mean-squared error and yet are more generalizable. This issue is difficult, if not impossible, to address because it is not clear how this tradeoff can be formalized. As a result, much of the literature, particularly that on natural experiments, has moved toward maximizing internal validity to at least learn something definite even if on a narrow population.

2.5 Additional Issues

The discussion of types of instruments provided in the last section is sufficient in and of itself to illustrate many of the issues with instrumental variable methods for estimating causal effects. Some of these issues bear further discussion and emphasis, and some additional issues should be introduced.

2.5.1 Heterogeneity

The issues discussed earlier in the paper surrounding the existence of heterogeneous effects and varying MTE, and the importance of identifying the specific population and point on the response curve from which estimates are being generated, has not penetrated most of the applied IV literature at this writing. In a large majority of the cases, binary instruments are used, which only permit a single LATE estimate to be obtained. Only rarely are comparisons made across studies in an attempt to piece together a fuller picture of the entire response curve (Card (2001) is an example of such an attempt). In addition, even in cases where multiple-valued instruments or multiple instruments are used, typically a single IV coefficient is estimated when more could be learned by estimating some portion of the MTE response curve. The literature is still young in this regard.

2.5.2 Differences in Estimates Across Instruments

One of the more difficult issues in the literature is that different instruments appear to generate differences in estimates for reasons not apparent. One possible reason is that just noted, that different instruments are moving the fraction of the population that is treated across different ranges of the population response curve. However, this often does not appear to explain the main differences in effects across instruments.

For the teen childbearing case, for example, Reinhold (2007) conducted an investigation into the reasons for differences in the effects of teen childbearing on high school completion using miscarriage as an instrument, which implies that childbearing has no effect if not a positive one, versus age at menarche, which implies a negative effect (albeit one smaller than OLS). Reinhold found that the differences could not be explained by the heterogeneous response curve, for the instruments yielded different estimates even when estimated at the same point on the curve (although the estimates also had large standard errors). Reinhold's investigation revealed that miscarriage occurs disproportionately among very young women, possibly because of immature physical development at that age, and that a comparison of women who miscarry with those who do not involves a comparison of women with different ages at the time of pregnancy. If those who become pregnant at young ages have more negative effects of teen childbearing on completed education than those who become pregnant at older ages, this lack of proper control for age could lead to small or zero effects of teen childbearing on education when using miscarriage as an instrument.

More generally, a possible cause of differences in IV estimates for different instruments is that they affect different types of individuals. Miscarriage affects women who have expressed a desire to have a child when young (assuming the child is desired) whereas increased contraceptive availability is likely to affect women who desire to have a child later and want to avoid having one early. These two types of women may not be affected in the same way by a postponement of childbearing, although it is not clear on a priori grounds which group would have the greater impact on later earnings. Similar reasoning could apply to other instruments.

A somewhat related reason for different estimates across instruments is that different instruments may work through different mechanisms, or channels. Indeed, the effect of teen childbearing on later earnings can have many different pathways. It could affect educational outcomes, which in turn affect earnings; the presence of young children could affect the ability to work after leaving school, even if education is unaffected; it could affect the types of jobs that a woman can and is willing to take, for similar reasons; or it could affect the probability of marrying, and marital status is known to have a significant effect on labor force participation and earnings. The increased control of fertility made possible by more availability of contraceptives, for example, may allow a woman to proceed with marriage to a suitable partner, knowing that fertility can still be controlled. Miscarriage may not have the same effects on marriage probabilities if fertility is relatively uncontrolled, and marriage may be delayed. To some extent, these differences in instrument effects can be examined by studying different mechanisms, of course, but this is only occasionally done.¹¹

¹¹ In all of these cases, the “compliers” – that is, the women whose decisions are changed by the variation in the instrument – are different. See Angrist and Imbens (1995) for a discussion of IV with multiple instruments.

That the particular mechanism through which a treatment has an effect may matter for IV estimation is an illustration of the more general point of Rosenzweig and Wolpin (2000) that a careful specification of theory is necessary prior to applying instrumental variables. They provide several illustrations where the interpretation of IV estimates differ dramatically depending on what other variables are controlled for in the equation and by what channels treatment has an effect.

2.5.3 Relevance of Instruments to Policies of Interest

Yet another issue of some importance in the IV literature is whether the instruments used are relevant to the policies or programs that might be the ultimate goal of the analysis. This problem appears most often with experiments of the “natural natural” type, where twins, month of birth, miscarriage, age at menarche, and related variables are used. These variables do not directly relate to any public policy and hence may be difficult to use to learn what the effects of policies might be. This is particularly true when there are multiple mechanisms by which the treatment can affect outcomes, for particular government policies may work in different ways. If teen childbearing were addressed by providing extra subsidies to stay in school, for example, that could have a different effect on later earnings than the effect of a miscarriage.

Studies which use policy-based instruments are preferable from this point of view because they provide direct information on the effects of at least one concrete policy. Even here, however, it is unclear how to use the results of a policy instrument study to forecast the effects of some other government policy which works in different ways. Both of these examples suggest that some attention should be paid to the specification of the selection equation and to using theory to account for the different mechanisms by which variables affect the treatment. For example, the monetary cost of staying in school is, according to economic theory, one possible determinant of voluntary teen childbearing. A study which represented that cost explicitly in the teen birth probability equation and which estimated its effects would allow a “mapping” of the effect of the instrument into the effect of schooling costs, and thereby permit an estimate of the effect of such alternative policies.

2.5.4 Reduced Form Versus Structural Form

Some of the studies in the IV literature estimate reduced forms rather than structural forms; that is, they estimate models for y as a function of X and Z directly. In the case of a binary instrument, this effect is simply the numerator of Equation 2.7 and therefore simply equals the IV coefficient multiplied by the change in the fraction of the population treated, so there is no real difference between them. However, estimation of the reduced form alone is generally conducted only when the instrument in question is of direct policy interest, for the results of reduced form estimation will

yield an estimate of the effect of the policy even if it does not work through the particular T specified in the model. On the contrary, simply estimating the effects of, say, having twins on later female earnings is not in and of itself very interesting unless it is interpreted as working through effects on family size. Nevertheless, even in the former case, most analysts believe that estimation of the structural form is of the greater theoretical interest because learning the mechanism by which policies have their effects, and that knowledge of this mechanism is necessary to design new policies which work through the same mechanism.¹²

2.5.5 Weak Instruments

The criterion of relevance for an instrument is an important one, for in many cases an instrument, while having a large asymptotic t-statistic in the estimation of the effect of the instrument on T , may nevertheless have small explanatory power for T . In that case, the instrument is said to be “weak” and the IV estimate of the effect of T on y can be shown to be biased toward OLS, and to have much larger confidence intervals than produced by the usual formulas (Cameron and Trivedi 2005). Rules of thumb have been developed for detecting weak instruments based on the F-statistic for a single instrument in the T equation, e.g., that it should be at least 10 (Stock et al. 2002; Stock and Yogo 2005) as well as more formal methods, and there are also formal methods for calculating more accurate confidence intervals for effect estimates when instruments are weak. The implication for practice is that a somewhat higher standard for instruments must be applied, for they not only must be valid in the standard asymptotic sense, but they must be sufficiently “strong.” In many applications, instruments which are arguably exogenous on theoretical grounds or which appear to have a significant coefficient in the T equation nevertheless are only weakly related to the fraction of the population treated, in which case usually a search for a stronger instrument is required.

2.5.6 No Instruments Available

The thinking about instrumental variables described in this essay has led to a higher standard for the choice of instruments than existed in earlier years of research, when the attitude toward instruments was more casual. This has made the search for a suitable instrument more difficult and in some cases no credible instrument exists either conceptually or in the available data sources. Particularly if attention is restricted to pure natural experiments of the type described above, the relative infrequency of

¹² Heckman and Vytalacil (2001) have argued that the reduced form estimates, which they term “policy relevant treatment effects,” are also useful because they do not require that there be no “defiers” in the language of Angrist et al. (1996). Defiers are individuals who change T in the opposite direction to that intended by the policy, e.g., who have more teen births after the increased availability of contraceptives.

such events may greatly restrict the set of research questions that can be studied. It would not be useful for scientific advance if questions where no instrument is available were simply left unstudied.

A variety of approaches are possible in this case. One is simply to apply OLS to the (y, T, X) relationship and to make a priori arguments on the degree of bias expected. These arguments will necessarily turn on how well the vector X is capable of capturing the main determinants of y and whether there are likely to be unobservables left out which are correlated with T . The direction and magnitude of bias from any remaining unobservables is often something that can be partially assessed on the basis of intuition and outside evidence. The method of matching, described above, can also be applied to determine whether the functional form of the estimated equation is affecting the conclusions drawn about the causal effect of T on y .

Another approach is to apply more formal sensitivity tests to the model to assess how much the estimated effect of T on y would be affected by different degrees of bias. In the case where all error terms are assumed to be multivariate normal, for example, the bias is captured by a control variable termed the Heckman lambda term (Barnow et al. 1980), and a single parameter – the correlation between the errors in the y equation and the T equation – determines the degree of bias in the coefficient on T . Fixing the correlation coefficient at different values and estimating the model with this restriction can be used to assess how the estimate of the effect of T is affected by the magnitude of the correlation (Robins et al. 2000). At another extreme, one can apply an analysis which determines the maximum degree of bias that might arise, a method of “bounds” analysis most formally developed within economics by Manski (1995). This “worst case” analysis can sometimes show that even in the maximal bias case, the estimated effect of T on y is still of reasonable magnitude on a scientific or policy basis. If the maximal bias results in a reversal of results, however, more restrictions on those bounds are needed to obtain more useful results.

2.6 Summary and Conclusions

Much progress has been made in understanding the estimation and interpretation of causal effects with observational data and how exclusion restrictions, which are an implicit assumption that an experiment exists in nature, can be used to identify those effects. Nevertheless, while a deeper understanding has been achieved, the difficulty of the problem has also become better understood. Most importantly, the criteria for valid and relevant instruments have been shown to be particularly stringent, and the scope of what is learned from instruments which are based on narrow populations is now seen to be possibly quite limited. Assessing the validity of instruments is also particularly problematic, as there are no formal tests for validity in the just-identified case, and resolving why different instruments yield different effect estimates can also be quite challenging. Studying the mechanism by which instruments affect

the fraction of the population treated and how that interacts with the mechanism by which treatment affects outcomes is now also recognized as important. Much therefore remains to be done.

References

- Angrist, J. and W. Evans (1998). Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *American Economic Review* 88 (June): 450–477.
- Angrist, J. and G. Imbens (1995). Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity. *Journal of the American Statistical Society* 90 (June): 431–442.
- Angrist, J., G. Imbens and D. Rubin (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association* 91 (June): 444–472.
- Angrist, J. and A. Krueger (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *Quarterly Journal of Economics* 106 (November): 979–1014.
- Angrist, J. and A. Krueger (1999). Empirical Strategies in Labor Economics. In: *Handbook of Labor Economics*, Vol. 3A, eds. O. Ashenfelter and D. Card. Amsterdam: North-Holland.
- Angrist, J. and A. Krueger (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives* 15 (Fall): 69–85.
- Barnow, B., G. Cain and A. Goldberger (1980). Issues in the Analysis of Selectivity Bias. In: *Evaluation Review Studies Annual*, eds. E. Stromsdorfer and G. Farkas. Beverly Hills: Sage.
- Björklund, A. and R. Moffitt (1987). The Estimation of Wage and Welfare Gains in Self-Selection Models. *Review of Economics and Statistics* 69: 42–49.
- Bound, J. and G. Solon (1999). Double Trouble: On the Value of Twins-Based Estimation of the Return to Schooling. *Economics of Education Review* 18 (April): 169–182.
- Cameron, A.C. and P. Trivedi (2005). *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.
- Card, D. (2001). Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica* 69 (September): 1127–1160.
- Carneiro, P.J., J. Heckman and E. Vytlačil (2006). *Estimating Marginal and Average Returns to Education*. New York: Mimeo.
- Chevalier, A. and T. Viitanen (2003). The Long-Run Labour Market Consequences of Teenage Motherhood in Britain. *Journal of Population Economics* 16: 323–343.
- Currie, J. and J. Gruber (1996). Health Insurance Eligibility, Utilization of Medical Care, and Child Health. *Quarterly Journal of Economics* 111 (May): 431–466.
- Geronimus, A. and S. Korenman (1992). The Socioeconomic Consequences of Teen Childbearing Reconsidered. *Quarterly Journal of Economics* 107 (November): 1187–1214.
- Heckman, J. (1978). Dummy Endogenous Variables in a Simultaneous Equation System. *Econometrica* 46: 931–960.
- Heckman, J. and R. Robb (1985). Alternative Methods for Evaluating the Impact of Interventions. In: *Longitudinal Analysis of Labor Market Data*, eds. J. Heckman and B. Singer. Cambridge: Cambridge University Press.
- Heckman, J., S. Urzua and E. Vytlačil (2006). Understanding Instrumental Variables in Models with Essential Heterogeneity. *Review of Economics and Statistics* 88 (August): 389–432.
- Heckman, J. and E. Vytlačil (1999). Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects. *Proceedings of the National Academy of Sciences* 96 (April): 4730–4734.
- Heckman, J. and E. Vytlačil (2001). Policy-Relevant Treatment Effects. *American Economic Review* 91 (May): 107–111.

- Heckman, J. and E. Vytlacil (2005). Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica* 73 (May): 669–738.
- Hoffman, S., E.M. Foster and F. Furstenberg (1993). Re-evaluating the Costs of Teenage Childbearing. *Demography* 30 (February): 1–13.
- Hotz, V.J., S. McElroy and S. Sanders (1997a). The Impacts of Teenage Childbearing on the Mothers and the Consequences of those Impacts for the Government. In: *Kids Having Kids: Economic Costs and Social Consequences of Teen Pregnancy*, ed. R. Maynard. Washington: Urban Institute Press.
- Hotz, V.J., C. Mullin and S. Sanders (1997b). Bounding Causal Effects Using Data from a Contaminated Natural Experiment: Analysing the Effect of Teenage Childbearing. *Review of Economic Studies* 64 (October): 575–603.
- Imbens, G. (2004). Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review. *Review of Economics and Statistics* 86 (February): 4–29.
- Imbens, G. and J. Angrist (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica* 62: 467–76.
- Imbens, G. and T. Lemieux (2008). Regression Continuity Designs: A Guide for Practice. *Journal of Econometrics* 142 (February): 615–35.
- Klepinger, D., S. Lundberg and R. Plotnick (1999). How Does Adolescent Fertility Affect the Human Capital and Wages of Young Women? *Journal of Human Resources* 34 (Summer): 421–448.
- Lee, L.F. (1979). Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables. *Econometrica* 47: 977–996.
- Manski, C. (1995). *Identification Problems in the Social Sciences*. Cambridge: Harvard University Press.
- Moffitt, R. (2003). Causal Analysis in Population Research: An Economist’s Perspective. *Population and Development Review* 29 (September): 448–458.
- Moffitt, R. (2005). Remarks on the Analysis of Causal Relationships in Population Research. *Demography* 42 (February): 91–108.
- Moffitt, R. (Forthcoming). Estimating Marginal Treatment Effects in Heterogeneous Populations. *Annales d’Economie et de Statistique*.
- Moffitt, R. and M. Ver Ploeg (eds.) (2001). *Evaluating Welfare Reform in an Era of Transition*. Washington: National Research Council.
- Reinhold, S. (2007). *Essays in Demographic Economics*. Unpublished Ph.D. dissertation, Johns Hopkins University.
- Ribar, D. (1994). Teenage Fertility and High School Completion. *Review of Economics and Statistics* 76 (August): 413–424.
- Ribar, D. (1999). The Socioeconomic Consequences of Young Women’s Childbearing: Reconciling Disparate Evidence. *Journal of Population Economics* 12 (November): 547–565.
- Robins, J., A. Rotnitzky and D. Scharfstein (2000). Sensitivity Analysis for Selection Bias and Unmeasured Confounding in Missing Data and Causal Inference Models. In: *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, eds. E.M. Halloran and D. Berry. New York: Springer-Verlag.
- Rosenzweig, M. and K. Wolpin (2000). Natural ‘Natural Experiments’ in Economics. *Journal of Economic Literature* 38 (December): 827–874.
- Rubin, D. (1974). Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies. *Journal of Educational Psychology* 66: 688–701.
- Stock, J., J. Wright and M. Yogo (2002). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business and Economic Statistics* 20 (October): 518–529.
- Stock, J. and M. Yogo (2005). Testing for Weak Instruments in Linear IV Regression. In: *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, eds. D. Andrews and J. Stock. New York: Cambridge University Press.

Chapter 3

Sequential Potential Outcome Models to Analyze the Effects of Fertility on Labor Market Outcomes

Michael Lechner

3.1 Introduction

This paper proposes to use dynamic treatment models to analyze the effects of fertility on labor market interactions. It argues that when large data sets are available the dynamic potential outcome model is an interesting modeling framework because it allows the careful consideration of the selection issues coming from the interaction of fertility and labor market decisions at different ages. It allows explicitly considering their dependence on the labor market and fertility history realized up to that period. There is no need to collapse the ‘endogeneity’ problem into a static setting since the dynamic nature and timing of the interaction can be explicitly addressed. Furthermore, the paper argues that this approach allows defining relevant parameters of interest in a more precise way. Based on artificial data, the approach is implemented and issues that may come up in practical applications of this approach are discussed.

The literature on the effect of fertility on labor market outcomes can be organized along the dimensions of model structure and time. The first strand, the so-called structural approach, uses fully structural behavioral models typically based on some sort of utility maximization subject to time and budget constraints. Usually, these models are fully parametrically specified. An early example of this approach is Moffitt (1984). There is in fact a considerable literature which is based on rather sophisticated structural modeling of the individual decision problems. These parametric models are combined with econometric modeling of the uncertainties in the model or the data. The resulting moment conditions or likelihood functions are then the basis for a parametric estimation of the parameters that are of particular interest in the specific application. This literature is surveyed in Arroyo and Zhang (1997) and Hotz et al. (1997), for example. Recent papers based on this approach are Del Bocca and Sauer (2005), Francesconi (2002) and Klepinger et al. (1999), among others.

M. Lechner (✉)

Swiss Institute of Empirical Economics Research, University of St. Gallen, Varnbühlstr. 14,
St. Gallen, Switzerland

e-mail: Michael.Lechner@unisg.ch

An alternative approach to structural modeling is reduced form modeling, i.e. deriving the equation to estimate not directly from some mathematically formulated theoretical model, but instead specifying the empirical model to be estimated by appealing to some general properties of various theoretical models of interest in a more or less ad-hoc way. A prominent paper based on this approach analyzing the timing and spacing of births in Sweden is Heckman and Walker (1990) who estimate duration models. Recent papers by Adsera (2005a, b) are based on a fairly similar approach.

A related approach that is more explicit on the various endogeneity problems that almost naturally occur in the previously mentioned group of papers models labor market outcomes and fertility decisions jointly based on selection type models. Examples for this approach are Hotz and Miller (1988), Di Tommaso (1999), and Troske and Voicu (2004).

However, even when the different decisions are analyzed within a joint modeling framework, it remains questionable whether the different effects can really be identified independently of each other. Therefore, Angrist and Evans (1998) propose an instrumental variable approach that allows analyzing the labor supply reaction when family size is varied exogenously. The latter variation comes from the observation that parents are more likely to get a third child, if the first two children have the same sex. Using an explicit non- or semiparametric causal framework, like in the analyses of post-unification fertility in East and West Germany by Lechner (1998, 2001a), is a rather unusual approach in that literature. Reduced form models appear to be typical as well for the demographic literature. For example, Rosenfeld (1996) and Connelly (1996) describe the work-fertility interaction without strict behavioral models, be it mathematical or not, and without a comprehensive empirical analysis. Note that none of the mentioned econometric or demographic reduced form analyses uses a dynamic framework.

The virtue of the structural approach is that the behavioral assumptions of the model are very clear. The drawback is however that the tight parametric functions are restrictive and usually rejected by the data, at least when the sample is large enough. In contrast, the virtue of the reduced form approach, particularly in its non-parametric causal version, is that it usually does not impose more than the required just identifying restrictions on the data. The drawback may be that those just identifying conditions are not explicitly derived from a formal mathematical model of utility maximization. Therefore, the conditions for their validity sometimes may be less clear than in the case of structural models.

The second dimension 'time' is straightforward. Some paper collapse individual fertility and labour market histories into one observation at a specific point in time, while other empirical approaches follow the realization of fertility and labor market outcomes over time. The virtue of the former approach is clearly its simplicity, while the latter approach, which requires more and better data, allows taking into account the important time dimension of fertility and labor market events.

Next, consider recent advances in econometric methodology. One of the recent important developments in econometrics is the increased emphasis on discovering

causal as opposed to associational relations from the data and clarifying the conditions required for the causal interpretation of the estimators used. Only causal relations are useful for policy advice, because they contain the reaction of the economic variables of interest.¹ Econometrics developed two different ways to define what a causal effect is. One concept originated in time series econometrics. The other concept comes from the sphere of microeconometrics and statistics.

The concept used in time series econometrics is due to Wiener (1956), Granger (1969), and Sims (1972) (see the review article by Geweke (1984) for an overview). Their basic idea is that (non-) causality is very similar, if not the same, than (non-) predictability. Therefore, they consider one variable not to cause another variable, if the current value of the causing variable does not help to predict future values of the variables that might capture the effects of this cause. This statement is conditional on the information set available at each point in time. This concept is in principle (technically) applicable if one cross-sectional unit (e.g. a country) is observed for a sufficiently long period.

The alternative concept popular in microeconometrics, particularly and most explicitly in the program evaluation literature (e.g. Heckman et al. 1999), is based on the idea that the relevant comparison is between different states of the world, each of which relates to a value of the causing variable. If causation is absent, then the outcomes that would have been realized if those potential states had actually been occurred would be the same (in some probabilistic sense). To relate this concept of different states of the world to data, it is necessary to observe different sample units in the different states. Then, so-called identifying assumptions are employed to relate the observed data to the distribution of the potential outcome variables, so that causal effects can be inferred from the ‘real world’ that is reflected in the data. The statistical formulation of the resulting inference problem is probably due to Neyman (1923) and was extended and popularized by Rubin (1974). Recently, dynamic versions of the potential outcome approach were suggested by Robins (1986) and extended by Lechner and Miquel (2001) that will be explained below.² Lechner (2009) proposes matching type estimators for this model.

The paper proceeds as follows: Section 3.2 outlines the dynamic causal framework. The notation is introduced and the basic identification conditions are restated. The estimation problem is explained in Section 3.3 and sequential matching estimation is reviewed. Section 3.4 relates the model explicitly to the substantive questions at stake in the pseudo-empirical example. Section 3.5 details the artificial data. Section 3.6 presents the ‘empirical’ examples and Section 3.7 concludes.

¹ See the excellent account of the historical developments in econometrics by Heckman (2000).

² Pötter and Blossfeld (2001) discuss concepts of causality and the use of the time dimension in sociology. However, they do not consider the research potential of dynamic potential outcome models.

3.2 The Dynamic Causal Model – Notation, Effects, and Identification

3.2.1 Introduction

Robins (1986) first suggested an explicitly dynamic causal framework based on potential outcomes that allows the definitions of causal effects of dynamic interventions and systematically addresses this type of selection problem. His approach was subsequently applied in epidemiology and biostatistics (e.g. Robins (1989, 1997, 1999) and Robins et al. (1999) for discrete treatments; Gill and Robins (2001) for continuous treatments) to define the effect of treatments in discrete time. Identification is achieved by sequential randomization assumptions (see the very comprehensible summary by Abbring (2003)). The effects are frequently estimated using parametric models.

Recently, Lechner and Miquel (2001, LM01 further on) extend Robins' (1986) framework to comparisons of more general sequences, different parameters, and different selection processes. Focusing on the case when all elements that influence selection and outcomes at each stage of the sequence are observable, LM01 discuss different identification conditions required for particular dynamic causal effects. Since the assumptions used in LM01 bear enough similarity to the selection on observables or conditional independence assumption (CIA) that is prominent in the static evaluation literature, Lechner (2009, L04 further on) proposed matching and inverse probability weighting estimators that are dynamic extensions of similar estimators used in the static model. These estimators retain most of the flexible and convenient properties of the static methods that made them the workhorse in empirical econometric evaluation studies (see the excellent survey by Imbens (2004)). Lechner (2008b) discusses some operational characteristics of this approach in the context of the evaluation of labor market programs. Although this approach has been applied by Lechner and Wiehler (2007) to analyze the effects of the timing and order of Austrian active labor market programs, generally applications of this explicit dynamic causal framework based on potential outcomes are however rare in econometrics so far.³

The following sections briefly repeat the definitions of the dynamic causal model as well as the identification results derived by Lechner and Miquel (2001) for the case of sequential selection on observables. To ease the notational burden, I use a three-period-two-treatments model to discuss the most relevant issues that distinguish the dynamic from the static model, although in the application more periods and more treatments are considered. As usual in the econometric evaluation literature, I use the standard statistics terminology based on treatments (fertility) and potential outcomes (labor market indicators) to define causal effects.

³ A further exception is Ding and Lehrer (2003) who use this framework and related work by Miquel (2002, 2003) to evaluate a sequentially randomized class size study using difference-in-difference-type estimation methods.

3.2.2 Basic Structure of the Model

For the fertility example, it is appropriate to treat age (in years) as the relevant time dimension, and consider calendar time as an individual exogenous attribute. Suppose that there is an initial period in which everybody is in the same treatment, plus two subsequent periods in which different treatment states are realized. The initial period is tied to the subpopulation of interest. Suppose further that interest is in the effects of births on the women aged 25 who did not give birth so far. In this case, the initial period would relate to age 25 and the population of interest (which then leads to the sample used in the estimation) would be the women of age 25 who did not give birth so far. However, no restrictions will be imposed on the population that relate to the periods after age 25. Periods are indexed by t or τ ($\tau \in \{0, 1, 2\}$). The treatment is defined over all periods. It is described by a vector of Bernoulli random variables (RV), $S = (S_1, S_2)$. S_t measures the occurrence of a birth in period t . For notational convenience, the treatment of the initial period ($S_0 = 0$) is sometimes not mentioned explicitly. A particular realization of S_t is denoted by $s_t \in \{0, 1\}$. Denote the history of variables up to period t by a bar below that variable, i.e. $\underline{s}_2 = (s_1, s_2)$.⁴ Since effect heterogeneity is not restricted over time, it makes sense to define potential outcomes in terms of sequences of potential states of the world. Thus, in period one, a woman is observed in exactly one of two treatments. In period two, the treatment will be described by two potential outcomes depending on what happened in period 1. Therefore, she is part of one of four treatments defined by the sequences (0,0), (1,0), (0,1), and (1,1). Thus, every individual is observed in exactly one sequence defined by s_1 and another sequence defined by the same value s_1 and a value s_2 . To sum up, in the two (plus one)-period-two-treatments example there are six different overlapping potential outcomes corresponding to two mutually exclusive states defined by treatment status in period 1 only, plus four mutually exclusive states defined by treatment status in periods 1 and 2 together. Such states could be characterized for example by one birth in period 1, zero births in period 2, or by no births at all, or any other sequence of births.

Variables used to measure the effects of the treatment in period t , i.e. the potential outcomes, are indexed by treatments and denoted by $Y_t^{s_1}$ ($t \geq 1$) or $Y_t^{s_2}$ ($t \geq 2$). They are measured at the end of each period, whereas treatment status is measured in the beginning of each period. For each sequence length (1 or 2 periods), one of the potential outcomes is observable and denoted by Y_t . Here, the potential outcomes measure individual labor market status, like labor market participation or earnings. To link the potential outcomes of the causal model to the data, the following observation rules are defined in Equation (3.1):

⁴ To differentiate between different sequences, sometimes a letter (e.g. j) is used to index a sequence, as in s_t^j . As a further convention, capital letters usually denote random variables, whereas small letters denote specific values of the random variables. When I deviate from this convention, the intended meaning should be obvious.

$$\begin{aligned}
Y_1 &= S_1 Y_1^1 + (1 - S_1) Y_1^0; \\
Y_2 &= S_1 Y_2^1 + (1 - S_1) Y_2^0 \\
&= S_1 S_2 Y_2^{11} + (1 - S_1) S_2 Y_2^{01} + S_1 (1 - S_2) Y_2^{10} + (1 - S_1)(1 - S_2) Y_2^{00}. \quad (3.1)
\end{aligned}$$

Finally, variables that may influence fertility behavior and (or) potential labor market outcomes are denoted by X . The K -dimensional vector X_t may contain functions of Y_t and is observable at the same time as Y_t .

3.2.3 Defining the Estimand: Average Causal Effects

Although it has already been stated that interest is in finding the effects of fertility on post-fertility labor market outcomes, it remains to formulate this estimand in terms of the dynamic causal model. First, assume for the sake of the simplified example that interest is in the effects of births at age 26 and 27 on the labor market outcome at age 28 (or any later age). It is important to note at this stage that this analytical framework allows for complete effect heterogeneity, i.e. different individuals may react differently to the same treatment sequence. Therefore, I will define different average treatment effects for different subpopulations based on the comparison of the same treatment sequences. Those subpopulations may be characterized by exogenous characteristics or more importantly by the treatment sequences themselves, i.e. the fertility sequences the women actually experienced in those two years.

As in the static model, the potential outcomes are used to define several average causal effects. Equation (3.2) defines the causal effect (for period t) of a sequence of treatments up to period 1 or 2 (τ , τ') compared to an alternative sequence of the same or a different length for a population defined by one of those sequences or a third sequence:

$$\begin{aligned}
\theta_t^{s_\tau^k, s_{\tau'}^l} (s_{\tilde{\tau}}^j) &= E \left(Y_t^{s_\tau^k} | \underline{s}_{\tilde{\tau}} = s_{\tilde{\tau}}^j \right) - E \left(Y_t^{s_{\tau'}^l} | \underline{s}_{\tilde{\tau}} = s_{\tilde{\tau}}^j \right), \quad 0 \leq \tilde{\tau}; \\
1 &\leq \tau, \tau' \leq 2, \quad \tilde{\tau} \leq \tau', \tau; \\
k &\neq 1, k \in (1, \dots, 2^\tau), l \in (1, \dots, 2^{\tau'}), j \in (1, \dots, 2^{\tilde{\tau}}). \quad (3.2)
\end{aligned}$$

The treatment sequences indexed by k , l , and j may correspond to (0) or (1) if τ (or τ') denotes period 1, or to the longer sequences (0,0), (0,1), (1,0), or (1,1) if τ (or τ') equals two. LM01 call $\theta_t^{s_\tau^k, s_{\tau'}^l}$ the dynamic average treatment effect (DATE). Accordingly, $\theta_t^{s_\tau^k, s_{\tau'}^l} (s_{\tilde{\tau}}^k)$, as well as $\theta_t^{s_\tau^k, s_{\tau'}^l} (s_{\tilde{\tau}}^l)$ are termed DATE on the treated (DATET) and DATE on the nontreated. There are cases in-between, like $\theta_t^{s_\tau^k, s_{\tau'}^l} (s_{\tilde{\tau}}^l)$, for which the conditioning set is defined by a sequence shorter than the one defining the causal contrast. Note that the effects are by definition symmetric for the same population ($\theta_t^{s_\tau^k, s_{\tau'}^l} (s_{\tilde{\tau}}^k) = -\theta_t^{s_{\tau'}^l, s_\tau^k} (s_{\tilde{\tau}}^k)$). This feature, however, does not restrict effect heterogeneity across individuals ($\theta_t^{s_\tau^k, s_{\tau'}^l} (s_{\tilde{\tau}}^k) \neq \theta_t^{s_\tau^k, s_{\tau'}^l} (s_{\tilde{\tau}}^l)$).

3.2.4 Identification

Having analytically defined the causal framework and the objects of interest, assumptions that are plausible in our application are required to find consistent estimators for those estimands, i.e. the estimands have to be linked to the data.

Assume that a large sample $\{s_{1i}, s_{2i}, x_{0i}, x_{1i}, y_{1i}, y_{2i}\}_{i=1:N}$ of size N is available, randomly drawn from a large population defined by $S_0 = 0$. The latter is characterized by the corresponding random variables $(S_1, S_2, X_0, X_1, Y_1, Y_2)$.⁵ Furthermore, assume that all conditional expectations that are of interest in the remainder of this paper exist. To ease notation further, assume that interest is in the effects of sequences of length two only. If the variables that jointly influence selection at each stage as well as the outcomes are observable, some average treatment effects are identified (weak conditional independence assumptions):

*Weak dynamic conditional independence assumption (W-DCIA)*⁶

- a) $Y_2^{00}, Y_2^{10}, Y_2^{01}, Y_2^{11} \perp\!\!\!\perp S_1 \mid X_0 = x_0$;
- b) $Y_2^{00}, Y_2^{10}, Y_2^{01}, Y_2^{11} \perp\!\!\!\perp S_2 \mid \underline{X}_1 = \underline{x}_1, S_1 = s_1$;
- c) $1 > P(S_1 = 1 \mid X_0 = x_0) > 0, 1 > P(S_2 = 1 \mid \underline{X}_1 = \underline{x}_1, S_1 = s_1) > 0$;
 $\forall \underline{x}_1 \in \underline{X}_1, \forall s_1 : s_1 \in \{0, 1\}$.

$\underline{X}_1 = (\chi_0, \chi_1)$ denotes the support of X_0 and X_1 . Part a) of W-DCIA states that the potential outcomes are independent of treatment choice in period 1 (S_1) conditional on X_0 . This is the standard version of the static CIA (e.g. Rubin 1974). Part b) states that conditional on the treatment in period 1, on observable outcomes of period 1 (which may be part of X_1), and on the confounding variables from periods 0 and 1 (\underline{X}_1), potential outcomes are independent of participation in period 2 (S_2).

To see whether such an assumption is plausible in this application, the question is which variables influence births as well as subsequent labor market outcomes and whether such variables are observable. If the answer to the latter question is yes, and if there is common support (defined in part c) of W-DCIA), i.e. there are individuals with the same observable characteristics that are observed in both treatment sequences of interest, then there is identification, even if some or all conditioning variables in period 2 are influenced by the labor market and fertility outcomes of period 1. LM01 show that, for example, quantities like $E(Y_2^{11})$, $E(Y_2^{11} \mid S_1 = 0)$, $E(Y_2^{11} \mid S_1 = 1)$, or $E[Y_2^{11} \mid \underline{S}_2 = (1, 0)]$ are identified, but that $E[Y_2^{11} \mid \underline{S}_2 = (0, 0)]$

⁵ To simplify the notation further, I consider period 2 as the only period relevant for the outcome of interest. However, for all what follows Y_2 should be considered as measured at some point in time after treatment 2 occurred. The exact timing is determined by the substantive interest of the researcher conducting the empirical study. Here, it could be any labor market outcome after the last day of age 27.

⁶ $A \perp\!\!\!\perp B \mid C = c$ means that *each element* of the vector of random variables B is independent of the random variable A conditional on the random variable C taking a value of c in the sense of Dawid (1979).

or $E[Y_2^{11} | S_2 = (0, 1)]$ are not identified. Thus, $\theta_2^{s_2^k, s_2^l}$ and $\theta_2^{s_2^k, s_2^l}(s_1^j)$ are identified $\forall s_1^k, s_2^k, s_1^l, s_2^l, s_1^j, s_2^j \in \{0, 1\}$, but $\theta_2^{s_2^k, s_2^l}(s_2^j)$ is not identified if $s_1^l \neq s_1^k$, or $s_1^l \neq s_1^j$, or $s_1^k \neq s_1^j$. This result states that pair-wise comparisons of all birth sequences are identified, but only for groups of women defined by their birth status in periods 0 or periods 0 and 1 together. The relevant distinction between the populations defined by fertility states in period 1 and subsequent periods is that in period 1, treatment choice is random conditional on exogenous variables, which is the result of the initial condition stating that $S_0 = 0$ holds for everybody. However, in the second period, randomization into these treatments is conditional on variables already influenced by the first part of the treatment. W-DCIA has an appeal for applied work as a natural extension of the static framework. However, W-DCIA does not identify the classical treatment effects on the treated which would define the population of interest using one of the complete sequences (for all three periods), if the sequences of interest differ in period 1.

LM01 show that to identify all treatment parameters, W-DCIA must be strengthened by essentially imposing that the confounding variables used to control selection into the treatment of the second period are not influenced by the selection into the first-period treatment. This can be summarized by an independence condition like $Y_2^{s_2} \perp\!\!\!\perp S_2 | X_1$ (LM01 call this the strong conditional dynamic independence assumption, *S-DCIA*). Note that the conditioning set includes the outcome variables from the first period. This is the usual conditional independence assumption used in the multiple treatment framework (with four treatments; see Imbens 2000, and Lechner 2001). In other words, when the control variables are not influenced by the previous treatments, the dynamic problem collapses to a static problem of four treatments with selection on observables. This result shows that by treating the dynamic selection process explicitly, the identifying assumption can be relaxed and still interesting effects are identified. In the example, the strong conditional independence assumption amounts to assuming that (intermediate) labor market outcomes at age 25 and 26 are not influenced by the contemporaneous birth events, which clearly is unrealistic.⁷

To sum up, note that the dynamic concept based on sequential conditional independence assumptions allows for weaker conditions on the selection process than using a static model: whereas the static model requires a selection on observables assumption to hold for all elements of the sequence at once, the dynamic model still works fine in most dimensions if the selection on observables assumption is only valid in each period for the next selection step conditional on the past selections steps, past outcomes, and other past confounders. However, both strategies will break down if there are unobservable characteristics that influence the selection steps as well as the potential outcomes. In this case, IV methods, like the one suggested by Miquel (2003) become relevant.

⁷ Note that if this assumption does not hold, then the conditioning variable X_1 would become endogenous and thus change the meaning of the estimand in some way that is hard to interpret (see Lechner 2008a).

Any attempts of nonparametrically estimating these effects face the problem that distributional adjustments based on a potentially high-dimensional vector of characteristics and intermediate outcomes (X) are required (details below). Therefore, in the applied static matching literature balancing scores are a popular device to reduce the dimension of the estimation problem (see Rosenbaum and Rubin 1983). Similar properties hold for the dynamic model as well:

Balancing score property for W-DCIA

If the conditions of W-DCIA hold, then:

- a) $Y_2^{00}, Y_2^{10}, Y_2^{01}, Y_2^{11} \perp\!\!\!\perp S_1 \mid b_1(X_0) = b_1(x_0)$ holds for all $b_1(x_0)$ such that $E[p^{s_1}(x_0) \mid b_1(X_0) = b_1(x_0)] = p^{s_1}(x_0)$; $p^{s_1}(x_0) := P(S_1 = s_1 \mid X_0 = x_0)$.
- b) $Y_2^{00}, Y_2^{10}, Y_2^{01}, Y_2^{11} \perp\!\!\!\perp S_2 \mid b_2(\underline{X}_1, S_1) = b_2(\underline{x}_1, s_1)$ holds for all $b_2(\underline{x}_1, s_1)$ such that $E[p^{s_2|s_1}(\underline{x}_1) \mid b_2(\underline{X}_1, S_1) = b_2(\underline{x}_1, s_1)] = p^{s_2|s_1}(\underline{x}_1)$; $p^{s_2|s_1}(\underline{x}_1) := P(S_2 = s_2 \mid \underline{X}_1 = \underline{x}_1, S_1 = s_1)$.

A low-dimensional choice for balancing scores consists of conditional transition probabilities in combination with the variable indicating the treatment in the previous period (which of course can be ignored in the first period): $b_1(x_0) = p^{s_1}(x_0)$, $b_2(\underline{x}_1, s_1) = [p^{s_2|s_1}(\underline{x}_1), s_1]$.

3.3 Estimation

3.3.1 Structure of Sequential Estimators

Lechner (2009) shows that these scores are convenient for constructing sequential propensity score matching estimators to correct for selection bias under W-DCIA. I focus on this particular estimator because of its simplicity and because it is the workhorse of empirical evaluation studies. Other static matching-type estimators can be adapted to the dynamic context in a similar way (see Imbens (2004) for an overview of available estimators). I refrain from discussing estimation based on the S-DCIA explicitly, because that assumption is not relevant here.

Using the balancing scores suggested above, the following estimand results for quantities identified under W-DCIA:

$$E\left(Y_2^{s_2^k} \mid S_1 = s_1^j\right) = E_{p_1^k(X_0)} \left\{ E_{p_2^k|s_1^k(\underline{X}_1)} \left[E\left(Y_2 \mid \underline{S}_2 = s_2^k, \underline{p}^{s_2^k|s_1^k, s_1^k}(\underline{X}_1)\right) \mid S_1 = s_1^k, p^{s_1^k}(X_0) \right] \mid S_1 = s_1^j \right\},$$

$$\underline{p}^{s_2^k|s_1^k, s_1^k}(\underline{X}_1) := \left[p^{s_2^k|s_1^k}(\underline{X}_1), p^{s_1^k}(X_0) \right], \quad s_1^k, s_2^k, s_1^j, s_1 \in \{0, 1\}. \quad (3.3)$$

To learn the counterfactual outcome for the population participating in s_1^j (the target population) had they participated in sequence \underline{s}_2^k , women with \underline{s}_2^k must be reweighted to make them comparable to the women in the target population (s_1^j). The dynamic, sequential structure of the causal model restricts the possible ways to do so. Intuitively, for the members of the target population, women in the first element of the sequence of interest (s_1^k) should be reweighted such that they have the same distribution of $p^{s_1^k}(X_0)$ as the target population. Call this artificially created group comparison group 1. Yet, to estimate the effect of the full sequence, the outcomes of women in \underline{s}_2^k instead of s_1^k are required. Thus, an artificial subpopulation of women in \underline{s}_2^k that has the same distribution of characteristics of $p^{s_1^k}(X_0)$ and $p^{s_2^k|s_1^k}(\underline{X}_1)$ as the artificially created comparison group 1 is required. The same principle applies for dynamic average treatment effects in the population (DATE).

All proposed estimators in L04 have the same structure: They are computed as weighted means of the outcome variables observed in subsample $\underline{S}_2 = \underline{s}_2^k$. The weights, $w(\cdot)$ depend on the specific effects of interest and are functions of the balancing scores.

$$E\left(\widehat{Y_2^{s_2^k} \mid S_1 = s_1^j}\right) = \sum_{i \in \underline{s}_2^k} w_i^{s_2^k, s_1^j} \left(\underline{p}^{s_2^k|s_1^k, s_1^k}(\underline{x}_{1,i}), s_1^k \right) y_i;$$

$$w_i^{s_2^k, s_1^j} \geq 0; \quad \sum_{i \in \underline{s}_2^k} w_i^{s_2^k, s_1^j} = 1; \quad (3.4)$$

$$E\left(\widehat{Y_2^{s_2^k}}\right) = \sum_{i \in \underline{s}_2^k} w_i^{s_2^k} \left(\underline{p}^{s_2^k|s_1^k, s_1^k}(\underline{x}_{1,i}), s_1^k \right) y_i; \quad w_i^{s_2^k} \geq 0; \quad \sum_{i \in \underline{s}_2^k} w_i^{s_2^k} = 1. \quad (3.5)$$

Note that in the case of more than two treatments, the balancing scores for (3.4) and (3.5) will differ with respect to the participation probability for the first period. For Equation (3.4), the required quantity is $P(S_1 = s_1^j | X_0 = x_0, S_1 \in \{s_1^k, s_1^j\})$, whereas in Equation (3.5), in which all of the population is the target, $P(S_1 = s_1^k | X_0 = x_0)$ is appropriate.

3.3.2 Sequential Matching Estimators (SM)

Lechner (2009) propose to extend the simple pair-matching estimators that are highly popular in applied studies to the dynamic context. The idea is to perform the required adjustments by sequentially choosing close pairs of observations in the various steps, so as to mimic the sequential conditional expectations appearing in expressions (3.4) and (3.5). The first step is the same for both effects and consists in finding for every women in $\underline{S}_1 = \underline{s}_1^k$ a women in $\underline{S}_2 = \underline{s}_2^k$ with very similar (the same) values of $p^{s_2^k|s_1^k}(\underline{x}_{1,i})$ and $p^{s_1^k}(x_{0,i})$. Note that matching must be with replacement, because the target population may be larger than the treatment

population. In the second step, every women in $S_1 = s_1^j$ (Equation 3.4) or $S_0 = 0$ (Equation 3.5) is to be paired with a women observed with $S_1 = s_1^k$ with very similar (same) values of $p^{s_1^k}(x_{0,i})$. The positive weights that are attached to some or all women in $\underline{S}_2 = \underline{s}_2^k$ coming from step 1 are then updated depending on how often a women in $\underline{S}_2 = \underline{s}_2^k$ is matched to a women of the target population via the intermediate matching step. This procedure leads to the following weights:

$$w_i^{s_2^k, s_1^j} = \frac{1}{N^{s_1^j}} \sum_{n \in s_1^j} \sum_{m \in s_1^k} v_1 \left[p^{s_1^k}(x_{0,n}), p^{s_1^k}(x_{0,m}); \cdot \right] \\ v_2 \left[\underline{p}^{s_2^k | s_1^k, s_1^k}(\underline{x}_{1,m}), \underline{p}^{s_2^k | s_1^k, s_1^k}(\underline{x}_{1,i}), s_1^k; \cdot \right]; \forall i \in \underline{S}_2 = \underline{s}_2^k, \quad (3.6)$$

$$w_i^{s_2^k} = \frac{1}{N} \sum_{n=1}^N \sum_{m \in s_1^k} v_1 \left[p^{s_1^k}(x_{0,n}), p^{s_1^k}(x_{0,m}); \cdot \right] \\ v_2 \left[\underline{p}^{s_2^k | s_1^k, s_1^k}(\underline{x}_{1,m}), \underline{p}^{s_2^k | s_1^k, s_1^k}(\underline{x}_{1,i}), s_1^k; \cdot \right]; \forall i \in \underline{S}_2 = \underline{s}_2^k. \quad (3.7)$$

$N^{s_1^j}$ denotes the number of observations for which $S_1 = s_1^j$. The function $v_1[p^{s_1^k}(x_{0,n}), p^{s_1^k}(x_{0,m}); \cdot]$ is defined to be one if $p^{s_1^k}(x_{0,m})$ is closest to $p^{s_1^k}(x_{0,n})$ among all observations belonging to the subsample defined by $S_1 = s_1^k$, and zero otherwise. Similarly, $v_2[\underline{p}^{s_2^k | s_1^k, s_1^k}(\underline{x}_{1,m}), \underline{p}^{s_2^k | s_1^k, s_1^k}(\underline{x}_{1,i}), s_1^k; \cdot]$ is one if observation i is closest to observation m (with $s_{1,m} = s_1^k$) in terms of $\underline{p}^{s_2^k | s_1^k}(\underline{x}_{1,i})$ and $p^{s_1^k}(x_{0,i})$, and zero otherwise. The Mahalanobis metric (a quadratic form of the variables defining the distance weighted by the inverse of their sample covariance matrix) is a frequently used measure for similarity. Note that the weight of observation i is 0, if it is not matched to any member of the target population. On the other extreme, if observation i is matched to every member of the target population its weight would be 1. A specific variant of this estimator is shown in Table 3.1 for the example of estimating $\theta_i^{s_2^1, s_2^0}(s_1^1)$.

Some remarks about this protocol that are already contained in L04 are worth repeating: First, matching is with replacement. Every step of the matching sequence is essentially the same as for matching in a static framework. However, sequential propensity score matching involves several probabilities in the second period matching step. Second, some issues arise from the sequential nature of matching. By choosing observations as matches with *similar* values of the probabilities instead of the *same* values (because such observations may not be available), it may happen that the probabilities attached to observations in earlier matching steps (relating to transitions in early periods) change over different sequential matching steps due to imprecise matching. To prevent this from happening, every matched comparison observation in period 2 is recorded with the values $\hat{p}_i^{s_1^1}$ of the observation it was matched to in period 1, instead of its own (\hat{p} denotes a consistent estimate of p).

Table 3.1 A sequential matching estimator for $\theta_{t^2, s_2^0}^{s_1^0} (s_1^1)$ based on propensity scores

Step 0: Sample reduction	Delete all observations not belonging to s_1^1, s_2^0 , or s_2^1 .
Step A: Match $s_2^0 = (s_1^0, s_2^0)$ to s_1^1 $(E(Y_t^{s_2^0} S_1 = s_1^1))$	
A.1.0	Define a weight $w_t^{s_2^0} = 0$ for every observation in s_2^0 .
A.1.P	Estimate a probit for $P(S_1 = s_1^0 X_0 = \underline{x}_0) \rightarrow p_t^{s_1^0}(x_{0,t}) = p_t^{s_1^0}$.
A.1.CS	Delete all obs. of s_1^1 with lower or higher values of $p_t^{s_1^0}$ than obs. in s_2^0 .
A.1.M	For every obs. in s_1^1 not deleted in A.1.CS find the obs. in s_1^0 that is closest in terms of $p_t^{s_1^0}$ (a match).
A.1.C	For the matched obs. keep the value of $p_t^{s_1^0}$ of the obs. in s_1^1 they have been matched to.
A.2.R	Some obs. in s_1^0 may appear many times in this matched sample.
A.2.P	Define a sample of obs. In s_1^0 .
A.2.CS	Estimate a probit for $P(S_2 = s_2^0 S_1 = s_1^0, X_1 = \underline{x}_1) \rightarrow p_t^{s_2^0 s_1^0}(x_{1,t}) =: p_t^{s_2^0 s_1^0}$.
A.2.M	Delete all obs. of the matched comparison sample of s_1^1 (defined in A.1.C) (as well as the corresponding elements of the target population s_1^1) with lower or higher values of $p_t^{s_2^0 s_1^0}$ and $p_t^{s_1^0 s_1^0}$ than obs. in s_2^0 .
	For every obs. in the matched comparison sample of s_1^1 not deleted in A.2.CS find an obs. in s_2^0 that is closest in terms of $p_t^{s_2^0 s_1^0}$ and $p_t^{s_1^0 s_1^0}$ using the Mahalanobis metric (covariance computed in s_1^1). Every time an obs. in s_2^0 is matched, its weight $w_t^{s_2^0}$ is increased by 1.
B.1.0	Define a weight $w_t^{s_2^1} = 0$ for every obs. in s_2^1 .
B.2.R	Reduce sample to participants in s_1^1 still in the common support.

Table 3.1 *Contd.*

Step 0: Sample reduction	Delete all observations not belonging to s_1^1, s_2^1 , or s_2^0 .
B.2.P	Estimate a probit for $P(S_2 = s_2^1 S_1 = s_1^1, X_1 = x_1) \rightarrow p^{s_2^1 s_1^1}(x_1, i) =: p_i^{s_2^1 s_1^1}$.
B.2.CS	Delete all obs. of s_1^1 with lower or higher values of $p_i^{s_2^1 s_1^1}$ than obs. in s_2^1 .
B.2.M	For every obs. in s_1^1 not deleted in B.2.CS find the member of s_2^1 that is closest in terms of $p_i^{s_2^1 s_1^1}$. Every time an obs. in s_2^1 is matched, its weight $w_i^{s_2^1}$ is increased by 1.
C.1	Reduce $w_i^{s_1^1}$ by 1 for every obs. i matched to an obs. in s_1^1 deleted in A.1.CS or A.2.CS.
C.2	Reduce $w_i^{s_2^0}$ by 1 for every obs. i matched to an obs. in s_1^1 deleted in B.2.CS.
D.1	$\hat{\theta}_t^{s_2^1, s_2^0}(s_1^1) = \frac{1}{\sum_{i \in s_2^1} w_i^{s_2^1}} \sum_{i \in s_2^1} w_i^{s_2^1} y_i - \frac{1}{\sum_{i \in s_2^0} w_i^{s_2^0}} \sum_{i \in s_2^0} w_i^{s_2^0} y_i$
D.2	$\widehat{Var} \left[\hat{\theta}_t^{s_2^1, s_2^0}(s_1^1) \right] = \frac{\sum_{i \in s_2^1} \left(\frac{s_2^1}{w_i^{s_2^1}} \widehat{Var}(Y_i S = s_2^1, w = w_i^{s_2^1}) \right) + \sum_{i \in s_2^0} \left(\frac{s_2^0}{w_i^{s_2^0}} \widehat{Var}(Y_i S = s_2^0, w = w_i^{s_2^0}) \right)}{\sum_{i \in s_2^1} (w_i^{s_2^1})^2} + \frac{\sum_{i \in s_2^0} (w_i^{s_2^0})^2}{\sum_{i \in s_2^0} (w_i^{s_2^0})^2}$
	$\widehat{Var} \left(Y_i S = s_2^j, w = w_i^{s_2^j} \right) = \frac{1}{N} \frac{I \left(\frac{s_2^j}{s_2^j, w_i^{s_2^j}} \right)}{I \left(\frac{s_2^j}{s_2^j, w_i^{s_2^j}} \right)} \sum_{i \in I \left(\frac{s_2^j}{s_2^j, w_i^{s_2^j}} \right)} \left[y_i - \bar{y} \right]^2,$
	$I \in I \left(\frac{s_2^j}{s_2^j, w_i^{s_2^j}} \right) = \sum_{i \in I \left(\frac{s_2^j}{s_2^j, w_i^{s_2^j}} \right)} 1, \bar{y} = \frac{I \left(\frac{s_2^j}{s_2^j, w_i^{s_2^j}} \right)}{N} = \frac{1}{N} \frac{I \left(\frac{s_2^j}{s_2^j, w_i^{s_2^j}} \right)}{I \left(\frac{s_2^j}{s_2^j, w_i^{s_2^j}} \right)} \sum_{i \in I \left(\frac{s_2^j}{s_2^j, w_i^{s_2^j}} \right)} y_i.$

The set $I \left(\frac{s_2^j}{s_2^j, w_i^{s_2^j}} \right)$ is determined as the $2\sqrt{N}$ closest neighbors w.r.t. the value of $w_i^{s_2^j}$ of observations with same observed treatment as observation i .

Note: $t > 1$. Changes required for the case $t=1$ are obvious. The number of neighbors in the k-NN estimation of the conditional variance is the one used in L04. In the empirical application below, the results appear not to be sensitive to small deviations from this value.

Hence, the ‘history’ of the match, or, in other words, the characteristics of the reference distribution, does not change when the next match occurs in the subsequent period.

Third, to compute $E(Y^{s_2^k} | S_1 = s_1^l)$ the only information that is needed for the $N^{s_1^l}$ participants in s_1^l is $\hat{p}_i^{s_1^k}$. Similarly, for participants in s_2^k , all probabilities of the type $\hat{p}_i^{s_2^k | s_1^k, s_1^k}$ are required. For participants in s_1^k , but not in s_2^k , only $\hat{p}_i^{s_1^k}$ is needed, and so on. To estimate $E(Y^{s_2^l} | S_1 = s_1^l)$ instead of $E(Y^{s_2^k} | S_1 = s_1^l)$, the only change in the matching protocol is that the initial matching step on $\hat{p}_i^{s_1^l}$ is redundant. When interest is in the average effect in the population ($E(Y^{s_2^k})$), then the whole population plays the role of the first reference group (instead of s_1^l). In this case, in the matching step based on $\hat{p}_i^{s_1^k}$, all participants in s_1^k are matched to themselves. In addition selected participants in s_1^k are matched to participants in the remaining treatments in the first period.

When matching is on the propensity score instead of directly on the confounding variables, there is the issue of selecting a probability model. It seems that, so far, even in the static model the literature has not addressed this thoroughly. So far the consensus seems to be that a flexibly specified (and extensively tested) parametric model is sufficiently rich and that the choice of the model does not really matter (e.g., see the Monte Carlo results by Zhao (2004)). Similarly, the suggestions in the literature to guide the specification choice by the ability to achieve balancing of the respective covariates (e.g. Rosenbaum and Rubin 1984; Rubin 2004) can be applied here as well (in each step).

Next, there is the issue of consistent estimation of the standard errors that is not yet resolved for the static matching literature. Based on the simulation results presented in L04, the standard errors are computed conditional on the weights. In other words, the fact that the weights are estimated quantities is ignored. Furthermore, the outcomes may show heteroscedasticity. However, heteroscedasticity is only relevant in this context if related to the weights. Therefore, a simple k-nearest neighbor estimator is used as in L04 to adjust for any such heteroscedasticity. Although such an estimator performed well in L04, there is potential for improvement.

The final remark about the matching protocol concerns the common support. The region of common support – defined on the reference distribution for which the effect is desired – has to be adjusted period by period with respect to the conditioning variables of that period. The matching estimator makes it easy to trace back the impact of this procedure on the reference distribution.

3.3.3 Multiple Treatments and Many Periods

The main issue concerns the specification of the propensity scores: For example, when specifying the probability of participating in s_2^k conditional on participating in s_1^k , is it necessary to take account of the fact that not participating in s_2^k implies a range of possible other states in period 2? The answer is no, because in each step the independence assumption relates only to a binary comparison, e.g. $Y_2^{s_2^k} \perp\!\!\!\perp \mathbb{1}(S_2 =$

s_2^k) $| S_1 = s_1^k, \underline{X}_1 = \underline{x}_1$, and $Y_2^{21} \prod \mathbb{1}(S_1 = s_1^k) | S_1 \in \{s_1^j, s_1^k\}, X_0 = x_0$ (s_1^j being the target population as before). Therefore, the conditional probabilities of not participating in the event of interest conditional on the history are sufficient.⁸ Hence, as already noted $P(S_2 = s_2^k | S_1 = s_1^k, \underline{X}_1 = \underline{x}_1)$ and $P[S_1 = s_1^k | X_0 = x_0, S_1 \in \{s_1^l, s_1^k\}]$ may be used in the matching step in period 1. The multiple treatment feature of the problem does not add to the dimension of the propensity scores.

3.4 Specifying Causal Parameters of Interest

3.4.1 General Issues

Since the causal model is formulated in discrete time, the first issue that arises concerns the concept of time to be used and the related question about the length of a period.

An important distinction is between process time, i.e. the clock starts running at some specific event that is related to the object of interest, or calendar time. The choice between the two concepts depends on the application, i.e. whether interesting causal effects are naturally defined in process time (like age specific birth patterns) or more naturally defined in calendar time. If one of those concepts is chosen, usually the other dimension of the problem will be controlled for in the estimation. For example, if process time is given by age, than any analysis based on calendar time will control for age effects and vice versa. In our current example concerning the labor market impacts of differential fertility behavior over the life-cycle, it appears natural to specify the model in terms of process time, here age. Doing so, allows specifying meaningful sequences as well as maximum flexibility in modeling relevant selection processes.

In an ideal world with very large sample sizes and high frequency data, the type of selection problem expected is one of the key determinants of the desirable length of a period. If it is assumed that short term events have important influences on the fertility decisions of the respective women, then shorter time periods allow more flexibility than longer periods. The price to pay for such flexibility is a loss in the precision of the estimates, because if the sequences cover only very specific events, then not many observations will be observed in any such sequence. In addition, it will be hard to interpret the (noisy) effects of sequences that have a very similar economic interpretation (like a birth in the first or second month after the 20th birthday). Therefore, more parsimonious models leading to more precise estimates can be obtained by using longer time periods. Since the systematic factors of fertility decisions are based on longer term considerations, the application considers two years as one period and counts the number of births within such a period. With smaller samples, further aggregation of periods may be advantageous.

⁸ Imbens (2000) and Lechner (2001) develop the same argument to show that in static multiple treatment models conditioning on appropriate one-dimensional scores is sufficient.

Another issue concerns the length of the sequence specified. The longer the sequences, i.e. the treatments are fixed over a longer time horizon, the more precise is the specification of the contrasts of interest, which could be important in some situation. However, as before, the price to pay will be a larger number of parameters to be estimated with a smaller number of observations.

3.4.2 Number of Children

First, consider the issue of estimating the effects of the number of children on labor market outcomes. For example, if interest is in the effect of one child compared to no child at all, one may want to consider sequences that cover (almost) all of the fertile ages of women. The contrasts involve one sequence with zeros everywhere compared to an alternative sequence with one child in some period. Since the timing of birth is not relevant in the above formulated causal question, one may want to aggregate the effects for all possible sequences describing one birth based on some weighting scheme. A plausible weighting scheme could for example be based on the number of observations in each such one-birth-sequence.

If this question is only interesting for example for births up to the age of 30, then the specified one-birth and zero-birth sequences would only cover the ages up to age 30. The same principles apply when comparing two to zero or two to one birth, and so on. Section 3.6 presents concrete examples.

3.4.3 The Effects of Timing and Spacing

A frequent question that arises is whether early births lead to different labor market outcomes than late births. This can be easily analyzed comparing sequences with the same total number of total births occurring at different ages. Varying the ages would also allow getting an estimate of the effect of an incremental postponement. Similarly, the spacing of the births, i.e. the time between the first and the second birth may be analyzed by comparing sequences with first births occurring at the same age in both sequences, but with the second birth occurring at a different age. Whereas such sequences could cover (most) of the fertile ages, one may be interested in the effects of 'starting' later or earlier without necessarily keeping the overall number of births fixed, but rather considering it as being determined by the early or late start. In this case, the sequences to be specified are much shorter. They would cover only the ages until the first birth of interest. This age can be varied to understand the full pattern of this effect.

Again, if interest is in a combination of different sequences, all effects for birth sequences with the same distance between births may be aggregated.

3.4.4 The Role of Confounding Variables

Compared to the static potential outcome model, the dynamic potential outcome model is a powerful tool if the selection process into (and the dropout from) the

sequences is determined by time varying variables that are related to the outcome of interest. Those variables, termed intermediate outcomes in the previous sections, play an important role in addition to any selection variables that are constant over the time window considered. Therefore, the data available must be informative about such variables. In the second part of Section 3.5, the selection processes into and out of some selected sequences are documented.

3.5 Data

To illustrate the methods, I generated a rather large sample of 100000 ‘women’ with ‘yearly’ information on labor market status, fertility, and some background characteristics that show similarity to typical variables measuring education, vocational degree and labor demand. These women are observed from the age of 16 to the age 45. For simplicity the data do not contain calendar time effects.⁹ Before discussing the ‘data’ in more detail, note that they are simulated for the purpose of illustrating some of the potentials of the suggested dynamic causal methods. They are neither meant to reflect any real-life data nor are they supposed to reflect some specific insights about the connection between fertility and socio-economic variables.

Quite to the contrary, when investigating some of the descriptive statistics for specific subsamples to be presented below, it is obvious that some of the statistics show rather extreme properties that are however not detrimental to the purpose intended in this paper. It should also be pointed out that although the actual sample size is rather large compared to real life data, the suggested approach can be used with smaller sample sizes as well.

In the simulated data, the number of births per year is modeled by a latent index model of the probit type. Choices depend on observables that also appear in the outcome processes as well as on normally distributed unobservables that are independent of observables and unobservables appearing in the outcome equations. Note obviously that normality or an index model assumption is not required for W-DCIA to hold, but they are convenient choices for a simulation exercise. Such observables are schooling, vocational degree, regional indicators, as well as lagged labor market indicators and the birth history. The labor market indicators, such as earnings and employment, are modeled as dynamic processes influenced by the exogenous variables as well as by their own past and current and past fertility behavior. All selection processes fulfill W-DCIA, but not D-CIA.

Table 3.2 contains descriptive statistics as well as a characterization of the type of variable for the most important time-dependent and time-independent variables. They are the usual types of variables with typical codings, means, and standard errors. A full set of statistics for all variables are available on request from the author. All descriptive statistics are shown for particular subsamples defined by selected treatment sequences. Since these are simulated data, I refrain from any interpretation

⁹ Since space constraints do not allow reproducing the data generating process explicitly, a Gauss 8.0 program is available from the author on request. Furthermore, a Gauss 8.0 program performing the estimation is available as well.

Table 3.2 Descriptive statistics for selected variables and selected subsamples

Variable	Type						Subsamples defined by fertility status between age 21 up to age 42						
	all		1		0		101		0000000110		1100000000		
	mean	std.	mean	std.	mean	std.	mean	std.	mean	std.	mean	std.	
Monthly earnings in EUR (0 for non-workers)													
Age 25	C	1133	1078	190	596	1393	1034	35	242	1846	450	241	699
35	C	1108	1172	362	839	1314	1166	248	674	1040	905	962	1124
45	C	1239	1445	409	970	1465	1475	259	708	1511	1068	949	1257
Out of labor force in %													
Age 25	I	38	—	83	—	25	—	93	—	0.1	—	75	—
35	I	42	—	77	—	32	—	82	—	13	—	48	—
45	I	40	—	75	—	31	—	80	—	14	—	48	—
Employment in %													
Age 25	I	57	—	10	—	71	—	2	—	99	—	12	—
35	I	53	—	17	—	63	—	13	—	61	—	46	—
45	I	54	—	18	—	64	—	13	—	75	—	43	—
Total number of children per women													
Age 25	D	.7	.9	1.9	.5	.3	.6	1.7	.4	0	—	2	—
35	D	1.5	1.2	2.6	.7	1.1	1.2	2.6	.6	.5	.5	2	—
45	D	1.7	1.3	2.9	.8	1.5	1.2	2.8	.6	2.1	.3	2.1	.3
Other variables													
Schooling (8–12)	D	10.1	1.7	10.5	1.2	9.9	1.3	9.9	1.1	10.2	1.2	10	1.2
Vocational degree (0, 1, 2)	D	.9	.4	.9	.6	.9	.6	.7	.5	.8	.6	.7	.6
Regional share of service sector	C	56	15	45	12	59	14	43	10	65	9.2	50	13
Regional share of production sector	C	30	11	33	12	29	10	33	12	80	8.9	32	11
Sectoral UE rate	C	12	4	12	4	12	4	12	4	12	4	12	4
Occupational UE rate	C	12	5	12	5	12	5	12	5	12	5	12	5
Observations		85843		16856		67327		2908		1409		2526	

Note: *I*: Binary indicator variable (0, 1); *D*: Discrete variable; *C*: Continuous variable. For indicator variables the share of ones is given in %. The number of observations in the full sample is 100,000. The table is based on the subsample of women who had no birth before the age of 21. Descriptive statistics for the variables and subsamples not included are available on request from the author.

of the descriptive statistics. However, it is important that the variation of the values of the covariates and intermediate outcomes across sequences reveals considerable selection effects as well as considerable differences in the outcome variables.

The selection process along a sequence is modeled by a binary probit model. The actual probit models underlying the results presented in the next sections are slightly misspecified, but in ways that remain largely undetected by conventional specification tests. The misspecification relates to the functional form as well as to the omission of some covariates that are highly correlated with covariates included in the sample. In this respect as well, the artificial data seem to exhibit similar problems and questions as real data sets usually do. The results for selected sequences are given in Table 3.3.

Following the specification over time, it becomes obvious that all probits include the same set of time constant variables, but differ with respect to time varying

Table 3.3 Estimated coefficients of sequential probit models (first part of a selected sequence)

Variable	0 vs. 1	00 if 0	001 if 00	0011 if 001	00110 if 0011	001100 if 00110	0011000 if 00110
Schooling (8–12)	<i>-.33</i>	<i>-.31</i>	<i>.24</i>	<i>.26</i>	<i>-.44</i>	<i>-.43</i>	<i>-.44</i>
Vocational degree (0,1, 2)	<i>-.31</i>	<i>-.22</i>	<i>.12</i>	<i>.14</i>	<i>-.45</i>	<i>-.31</i>	<i>-.34</i>
Regional share of service sector	<i>.005</i>	<i>.009</i>	<i>-.01</i>	<i>-.01</i>	<i>.003</i>	<i>.01</i>	<i>.007</i>
Regional share of production sector	<i>-.001</i>	<i>-.001</i>	<i>-.000</i>	<i>0.002</i>	<i>-.003</i>	<i>-.004</i>	<i>-.004</i>
Sectoral UE rate	<i>-.002</i>	<i>-.000</i>	<i>-.001</i>	<i>.002</i>	<i>.005</i>	<i>.006</i>	<i>.01</i>
Regional UE rate	<i>.001</i>	<i>-.000</i>	<i>.002</i>	<i>.002</i>	<i>.005</i>	<i>.008</i>	<i>.01</i>
Employed at age 20	<i>.42</i>	0	0	0	0	0	0
Out-of labor force at age 20	<i>-.24</i>	0	0	0	0	0	0
Earnings at age 20	<i>.0007</i>	0	0	0	0	0	0
Employed at age 22	X	<i>.49</i>	0	0	0	0	0
Earnings at age 22	X	<i>.0008</i>	0	0	<i>-.000</i>	<i>-.0002</i>	<i>.000</i>
Employed at age 24	X	X	<i>-.50</i>	0	0	0	0
Earnings at age 24	X	X	<i>-.0007</i>	0	0	0	0
Out of labor force at age 26	X	X	X	<i>.62</i>	0	0	0
Earnings at age 26	X	X	X	<i>-.0007</i>	0	0	0
Out of labor force at age 28	X	X	X	X	<i>-.73</i>	0	0
Out of labor force at age 30	X	X	X	X	X	<i>-.73</i>	0
Out of labor force at age 32	X	X	X	X	X	X	<i>-1.12</i>
Subsample	1 and 0	0	00	001	0011	00110	001100
Number of observations in subsample	84183	67327	55131	6504	4787	4260	3933
Dependent variable	0	00	001	0011	00110	001100	0011000
Mean of dep. variable in subsample	0.8	.8	.14	.74	.89	.93	.86

Note: Binary probit model estimated on the respective subsample. All specifications include an intercept. If not stated otherwise, all information in the variables relates to age 20. *Exclusion restrictions*: 0: Variables omitted from specification. X: Variable not temporarily prior to the dependent variable. **Bold** letters in *italics* denote significance at the 1% level. **Bold** letters denote significance at the 5% level. *Italics* denote significance at the 10% level.

variables. According to the theory outlined so far, the number of time varying variables to be included in the probit should increase along the sequences, since each estimation step should include all variables previously included as well as those variables directly prior to the current treatment. Doing so would however lead to multicollinearity problems due to overparametrization. This problem becomes more binding over time, as the number of women still in the sequence decreases. Since it is not the purpose of this estimation step to obtain consistent estimates of probit coefficients, but of the respective probabilities instead, deleting variables because of their almost perfect correlation with other variables already in the model does not harm consistent estimation of the treatment effects.

3.6 Estimation Results

3.6.1 *Quantity*

Based on the considerations of Section 3.4, Table 3.4 presents some example on how to estimate the labor market effects of additional children. In the left part of the table two different ways of modeling the effect of one additional child are shown: The first comparison is for a birth at the age 21/22 and no birth for the next four years (sequence 100) compared to no births over six years from age 21–26 (000). Since it may be true that the first child has a different effect than the second one, a second specification is added which compares the one birth at 21/22 case to another sequence with an additional birth at 23/24 (110). Note that the way the effect is specified it is left open whether there will be or will not be more births after the age of 26. In this sense it is a minimal specification with respect to time periods that has been chosen because of sample size considerations (sequences with one child only are very rare in the simulated data). If there are enough observations in the individual sequences one may want to compare three children to two children as well as consider different timings of births. Furthermore, one may consider longer time periods without birth as is done in the final specifications contained in that table. The latter compare two or three early children with no children. The coverage of these sequences is now from age 21 to 40 and thus covers almost all of women's fertile period.

The particular target population and its size after imposing the common support criterion is shown in the line just below the sequences.¹⁰ The effects are estimated for different subpopulations that are defined by the first diverging element in each of the sequences that are compared. These effects are identified by the W-DCIA. A more stringent specification of those subpopulations (i.e. longer sequences for the target population) would require the S-DCIA to hold instead of the weaker W-DCIA,

¹⁰ Members of the target population, for whom no comparable observations exist for any step in any of the two sequences, are removed. They violate W-DCIA. This procedure changes the estimand, but is common practice in station.

Table 3.4 Labor market effects of birth sequences at age 45 (quantity)

Outcome	1 additional birth			2 births			3 births		
	Employment	Earnings	Employment	Earnings	Employment	Earnings	Employment	Earnings	
Sequences: $s^1 - s^0$	100-000		110-100		1100000000 - 0000000000		1110000000 - 0000000000		
Target population s (obs. after common support, N_s)	1 (15563)		11 (9634)		1 (14847)		1 (14484)		
Estimated outcome:									
$E(Y_T^1 S = s)$	28	377	16	348	25	547	13	308	
	(2)	(39)	(-4)	(10)	(2)	(51)	(2)	(51)	
Estimated outcome:									
$E(Y_T^0 S = s)$	23	520	15	312	90	1490	68	1196	
	(3)	(89)	(2)	(43)	(15)	(537)	(22)	(837)	
Effect: $\theta_T^{0,s^1}(s)$	-5	-142	1	35	-64	-943	-56	-889	
	(-3)	(98)	(2)	(45)	(15)	(540)	(23)	(838)	
$s(N_s)$	0 (64696)		10 (5068)		0 (60467)		0 (46092)		
$E(Y_T^1 S = s)$	57	1207	24	489	57	1144	24	508	
	(2)	(54)	(1)	(22)	(4)	(96)	(11)	(298)	
$E(Y_T^0 S = s)$	64	1354	27	565	90	1740	74	1436	
	(-4)	(9)	(1)	(18)	(4)	(106)	(8)	(173)	
$\theta_T^{0,s^1}(s)$	-6	-147	-3	-75	-33	-596	-50	-928	
	(2)	(55)	(1)	(29)	(5)	(143)	(13)	(345)	

Note: Each element of a sequence covers two years. The first element of each sequence refers to the age group 21-22. All sequences are defined such that there are no births before age 21. Outcome variables are measured at age 45. Standard errors are in parentheses. **Bold and italics**: Effect is significant at 1% level. **Bold**: Effect is significant at 5% level. *Italics*: Effect significant at 10% level. N_s is the sample size of the target population after imposing common support.

which is however not true with this simulated data, and usually not plausible in such applications as well.

The results of the estimations are given below the line for the target population. They show the estimated values for the (counterfactual) potential outcomes in the target population as well as their difference, which is the desired causal effect of the comparison of the two sequences. Standard errors are in brackets.

With respect to the results, all effects have the expected sign. For the larger target population of those women without a birth at age 21/22 all effects appear to be significant at conventional levels, whereas for the smaller subpopulation precision becomes a problem for the one-child comparisons, but not so for the comparisons involving more than one child. Comparing the effects across populations reveals some indication for possible effect heterogeneity.

3.6.2 Timing

Table 3.5 shows 8 different comparisons that all relate to timing issues, like when to have the birth, the time between first and second birth, and when to start fertility. The first 6 comparisons are based on rather long sequences (20 years if sample sizes permit) to investigate the various dimensions of early versus late kids. Therefore, the overall number of kids is kept constant over those 20 years. The remaining two comparisons investigate the effects of timing the first birth only ('starting early' vs. 'starting late'). Thus, the remaining birth history is not specified and considered as being part of the effect.

According to the data generating process, later births should have positive effects on labor market outcomes compared to earlier kids. Those effects should be somewhat smaller than the effects for different quantities of births. This result is generally confirmed by the data, although the precision of the estimates is not always enough to pin down the effects precisely.

3.7 Conclusions

This paper discusses to potential use of dynamic potential outcome models to analyze the effects of fertility on labor market interactions when large and informative data sets are available. The main advantages of such an approach are (i) its flexibility due to its non- or semiparametric nature, (ii) that it allows addressing the selection issues coming from the dynamic interaction between fertility and labor market decisions and realizations in a detailed way; and that (iii) it allows defining relevant parameters of interest in a more precise way than in any static approach. Based on artificial data, the approach is implemented and issues that come in any practical application of this approach are discussed.

Table 3.5 Labor market effects of birth sequences at age 45 (timing)

Outcome	Early vs. late births			Time between births		
	Employment	Earnings	001 (2879)	Employment	Earnings	2 (1076)
Sequences $s^1 - s^0$	1100000000-001100000	110000000-0000000110	0011000000-0000000110	200000-101000		
Target population s (obs. after common support, N_s)	1 (11589)	1 (4381)	001 (2879)			
Estimated outcome: $E(Y_t^{s^1} S = s)$	25 (1)	494 (25)	51 (2)	1064 (47)	57 (2)	21 (4)
Estimated outcome: $E(Y_t^{s^0} S = s)$	23 (2)	467 (44)	63 (6)	1258 (130)	64 (3)	21 (5)
Effect: $\theta_t^{s^0, s^1}(s)$	2 (2)	27 (51)	-12 (6)	-194 (138)	-7 (3)	1 (6)
$s(N_s)$	0 (56259)		0 (42337)		000 (39526)	1 (13914)
$E(Y_t^{s^1} S = s)$	56 (4)	1122 (98)	69 (5)	1386 (120)	80 (4)	17 (2)
$E(Y_t^{s^0} S = s)$	60 (3)	1243 (64)	81 (1)	1584 (36)	82 (1)	19 (2)
$\theta_t^{s^0, s^1}(s)$	-4 (4)	-120 (117)	-11 (5)	-198 (128)	-2 (4)	-2 (3)

Contd.

Table 3.5 *Contd.*

	Time between births		Early vs. late start				
	1100000000-1001000000 11 (7092)	1-01 1 (16739)	1-0001 1 (11954)	1-0000000001 1 (3766)			
$\frac{s^1 - s^0}{s(N_s)}$							
$E(Y_T^{s^1} S = s)$	21 (1)	18 (.3)	411 (8)	23 (.4)	472 (8)	50 (1)	1033 (18)
$E(Y_T^{s^0} S = s)$	24 (2)	19 (.7)	427 (20)	24 (6)	500 (145)	53 (5)	1058 (105)
$\theta_T^{s^0, s^1}(s)$	-3 (3)	-1 (1)	-17 (21)	-1 (6)	-28 (145)	-2 (5)	-25 (106)
$s(N_s)$		10 (4457)		0 (64821)			0 (46846)
$E(Y_T^{s^1} S = s)$	28 (2)	57 (1)	1180 (36)	59 (1)	1214 (37)	71 (2)	1462 (49)
$E(Y_T^{s^0} S = s)$	32 (2)	60 (1)	1275 (53)	63 (2)	1329 (46)	77 (1)	1515 (39)
$\theta_T^{s^0, s^1}(s)$	-4 (3)	-3 (2)	-95 (45)	-4 (2)	-114 (59)	-5 (2)	-52 (63)

See note below Table 3.3.

References

- Abbring, J.H. (2003). *Dynamic Econometric Program Evaluation*. IZA Discussion Paper 804.
- Adsera, A. (2005a). *Labor Market Performance and the Timing of Births*. University of Chicago, Department of Economics.
- Adsera, A. (2005b). *Where Are the Babies Gone? Labor Market Conditions and Fertility in Europe*. IZA Discussion paper 1576.
- Angrist, J.D. and W.N. Evans (1998). Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *The American Economic Review* 88: 450–477.
- Arroyo, C. and J. Zhang (1997). Dynamic microeconomic models of fertility choice: A survey. *Journal of Population Economics* 10: 23–65.
- Connelly, R. (1996). Comments on the Fertility / Employment Interaction. *Population and Development Review* 22, Supplement: 290–294.
- Dawid, A.P. (1979). Conditional Independence in Statistical Theory. *Journal of the Royal Statistical Society B* 41: 1–31.
- Del Boca, D. and R. Sauer (2005). *Life Cycle Employment and Fertility Across Institutional Environments*. mimeo.
- Di Tommaso, M.L. (1999). A trivariate model of participation, fertility and wages: the Italian case. *Cambridge Journal of Economics* 23: 623–640.
- Ding, W. and S.F. Lehrer (2003). *Estimating Dynamic Treatment Effects from Project STAR*. mimeo.
- Francesconi, M. (2002). A Joint Dynamic Model of Fertility and Work of Married Women. *Journal of Labor Economics* 20: 336–380.
- Geweke, J. (1984). Inference and Causality in Economic Time Series. In: *Handbook of Econometrics*, Vol. 2, eds. Z. Griliches and M. D. Intriligator. Amsterdam: North-Holland.
- Gill, R.D. and J.M. Robins (2001). Causal Inference for Complex Longitudinal Data: The Continuous Case. *The Annals of Statistics* 2001: 1–27.
- Granger, C.W.J. (1969). Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica* 37: 424–438.
- Granger, C.W.J. (1986). Comment. *Journal of the American Statistical Association* 81: 967–968.
- Heckman, J.J. (2000). Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective. *Quarterly Journal of Economics* 115: 45–97.
- Heckman, J.J. and J.J. Walker (1990). The Relationship between Wages and Income and the Timing and Spacing of Births: Evidence from Swedish Longitudinal Data. *Econometrica* 58: 1411–1441.
- Heckman, J.J., R.J. LaLonde and J.A. Smith (1999). The Economics and Econometrics of Active Labor Market Programs. In: *Handbook of Labor Economics*, Vol. 3A, eds. O. Ashenfelter and D. Card. Amsterdam: North-Holland.
- Holland, P.W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association* 81: 945–970.
- Hotz, J.V. and R.A. Miller (1988). An Empirical Analysis of Life Cycle Fertility and Female Labor Supply. *Econometrica* 56: 91–118.
- Hotz, J.V., J.A. Klermann and R.J. Willis (1997). The Econometrics of Fertility in Developed Countries. In: *Handbook of Population and Family Economics*, eds. M. R. Rosenzweig and O. Stark. Amsterdam: Elsevier.
- Imbens, G.W. (2000). The Role of the Propensity Score in Estimating Dose-Response Functions. *Biometrika* 87: 706–710.
- Imbens, G.W. (2004). Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review. *Review of Economics and Statistics* 86(1): 4–29.
- Klepinger, D., S. Lundberg and R. Plotnick (1999). How Does Adolescent Fertility Affect the Human Capital and Wages of Young Women? *The Journal of Human Resources* 34: 421–448.
- Lechner, M. (1998). Eine empirische Analyse der Geburtenentwicklung in den neuen Bundesländern aus der Sicht der neoklassischen Bevölkerungsökonomie. *Zeitschrift für Wirtschafts- und Sozialwissenschaften (ZWS)* 118: 463–488.

- Lechner, M. (2001a). The Empirical Analysis of East German Fertility after Unification: An Update. *European Journal of Population* 17: 61–74.
- Lechner, M. (2001b). Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption. In: *Econometric Evaluation of Active Labour Market Policies*, eds. M. Lechner and F. Pfeiffer. Heidelberg: Physica.
- Lechner, M. (2009). Sequential Causal Models for the Evaluation of Labor Market Programs. *Journal of Business & Economic Statistics* 27: 71–83.
- Lechner, M. (2008a). A Note on Endogenous Control Variables in Causal Studies. *Statistics and Probability Letters* 78: 190–195.
- Lechner, M. (2008b). Matching estimation of dynamic treatment models: Some practical issues, in: D. Millimet, J. Smith, and E. Vytlacil (eds.), *Advances in Econometrics: Volume 21, Modelling and Evaluating Treatment Effects in Econometrics* 289–333.
- Lechner, M. and Miquel, R. (2001). *A Potential Outcome Approach to Dynamic Programme Evaluation – Part I: Identification*. Discussion paper 2001–07, Department of Economics, University of St. Gallen; revised 2005.
- Lechner, M. and S. Wiehler (2007). *Does the Order and Timing of Active Labour Market Programmes Matter?* Discussion paper 2007–38, Department of Economics, University of St. Gallen.
- Neyman, J. (1923). On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9, translated in *Statistical Science* 5 (1990): 465–480.
- Miquel, R. (2002). *Identification of Dynamic Treatments Effects by Instrumental Variables*. University of St. Gallen, Discussion paper, 2002–11.
- Miquel, R. (2003). *Identification of Effects of Dynamic Treatments with a Difference-in-Differences Approach*. University of St. Gallen, Discussion paper, 2003–06.
- Moffitt, R. (1984). Profiles of Fertility, Labour Supply and Wages of Married Women: A Complete Life-Cycle Model. *Review of Economic Studies* 51: 263–278.
- Pötter, U. and H.-P. Blossfeld (2001). Causal Inference from Series of Events. *European Sociological Review* 17 (1): 21–32.
- Robins, J.M. (1986). A new approach to causal inference in mortality studies with sustained exposure periods – Application to control of the healthy worker survivor effect. *Mathematical Modeling* 7:1393–1512; with 1987 Errata to: A new approach to causal inference in mortality studies with sustained exposure periods – Application to control of the healthy worker survivor effect. *Computers and Mathematics with Applications* 14:917–921; 1987 Addendum to: A new approach to causal inference in mortality studies with sustained exposure periods – Application to control of the healthy worker survivor effect. *Computers and Mathematics with Applications* 14:923–945; and 1987 Errata to: Addendum to ‘A new approach to causal inference in mortality studies with sustained exposure periods – Application to control of the healthy worker survivor effect’. *Computers and Mathematics with Applications* 18: 477.
- Robins, J.M. (1989). The Analysis of Randomized and Nonrandomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies. In: *Health Service Research Methodology: A Focus on Aids*, eds. L. Sechrest, H. Freeman and A. Mulley. Washington, D.C.: Public Health Service, National Centre for Health Services Research.
- Robins, J.M. (1997). Causal Inference from Complex Longitudinal Data. Latent Variable Modelling and Applications to Causality. In: *Lecture Notes in Statistics*, ed. M. Berkane. New York: Springer-Verlag.
- Robins, J.M. (1999). Association, Causation, and Marginal Structural Models. *Synthese* 121: 151–179.
- Robins, J.M., S. Greenland and F. Hu, (1999). Estimation of the Causal Effect of a Time-varying Exposure on the Marginal Mean of a Repeated Binary Outcome. *Journal of the American Statistical Association* 94: 687–700.
- Rosenbaum, P.R. and D.B. Rubin (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70: 41–50.
- Rosenbaum, P.R. and D.B. Rubin, (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association* 79: 516–524.

- Rosenfeld, R.A. (1996). Women Work Histories. *Population and Development Review* 22, Supplement: 199–222.
- Rubin, D.B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology* 66: 688–701.
- Rubin, D.B. (2004). On principles for modeling propensity scores in medical research. *Pharmacoepidemiology and Drug Safety* 13: 855–857.
- Sims, C.A. (1972). Money, Income, and Causality. *American Economic Review* 62: 540–552.
- Troske, K.R. and A. Voicu (2004). *Joint estimation of sequential labour force participation and fertility decisions using Markov chain Monte Carlo techniques*. Columbia: University of Missouri, Department of Economics.
- Wiener, N. (1956). The Theory of Prediction. In: *Modern Mathematics for Engineers, Series 1*, ed. E.F. Beckenham, Chapter 8. New York: McGraw-Hill.
- Zhao, Z. (2004). Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metric and a Monte Carlo Study. *The Review of Economics and Statistics* 86: 91–107.

Chapter 4

Structural Modelling, Exogeneity, and Causality

Michel Mouchart, Federica Russo and Guillaume Wunsch

4.1 Causal Analysis in the Social Sciences

4.1.1 Goals of Causal Analysis

Whilst it might seem uncontroversial that the health sciences search for causes – that is, for causes of disease and for effective treatments – the causal perspective is less obvious in social science research, perhaps because it is apparently harder to glean general laws in the social sciences than in other sciences, due the probabilistic character of human behaviour. Thus the search for causes in the social sciences is often perceived to be a vain enterprise and it is often thought that social studies merely *describe* the phenomena.

On the one hand, an explicit causal perspective can already be found in pioneering works of Adolphe Quetelet (1869) and Emile Durkheim (1897) in demography and sociology respectively, and the social sciences have taken a significant step in *quantitative causal analysis* by following Sewall Wright's path analysis (1934), first applied in population genetics. Subsequent developments of path analysis – such as structural models, covariance structure models or multilevel analysis – have the merit of making the concept of cause operational by introducing causal relations into the framework of statistical modelling. However, these developments in causal modelling leave a number of issues at stake, for instance a deeper understanding of exogeneity and its causal importance.

On the other hand, an explicit causalist perspective still needs justification. Different social sciences study society and humans from different angles and perspectives. Sociology studies the structure and development of human society, demography attends to the vital statistics of populations, economics studies the management of goods and services, epidemiology studies the distribution of disease in human populations and the factors determining that distribution, etc. In spite of

Guillaume Wunsch (✉)

Institute of Demography, University of Louvain, Place Montesquieu 1/17,

Louvain-la-Neuve, Belgium

e-mail: Wunsch@demo.ucl.ac.be

these differences, social sciences share a common objective: to understand, predict and intervene on individuals and society. In these three moments of the scientific demarche, knowledge of causes becomes essential. The importance of causal knowledge is twofold. Firstly, we pursue a *cognitive goal* in detecting causes and thus in gaining general knowledge of the causal mechanisms that govern the development of society. Secondly, such general causal knowledge is meant to guide and inform social policies, that is we also pursue an *action-oriented goal*. If the social sciences merely *described* phenomena, it would not be possible to design efficient policies or prescribe treatments that rely on the results of research.

As stated above, the social sciences do not establish laws as physics does. Whether this is an intrinsic issue of these sciences, or merely a contingent issue due to the specificity of social problems, is still matter of debate and falls far beyond the scope of the present paper. In the following, we will rather reverse the perspective and try to tackle the issue: under what conditions can structural models give us causal knowledge?

4.1.2 Variation and Regularity in Causal Analysis

The first thing worth mentioning is that we need to abandon the paradigm of regularity as regular succession of events in time, a heritage of Hume, in favour of a more flexible framework. Hume believed that causality lies in the constant conjunction of causes and effects. In his *Treatise* Hume (1748) says that, in spite of the impossibility of providing rational foundations for the existence of objects, space, or causal relations, believing in the existence of causal relations is a “built in” habit of human nature. In particular, belief in causal relations is granted by experience. For Hume, simple impressions always precede simple ideas in our mind, and by introspective experience we also know that simple impressions are always associated with simple ideas. Simple ideas are then combined in order to form complex ideas. This is possible thanks to imagination, which is a normative principle that allows us to order complex ideas according to (i) resemblance, (ii) contiguity in space and time, and (iii) causality. Of the three, causation is the only principle that takes us beyond the evidence of our memory and senses. It establishes a link or connection between past and present experiences with events that we predict or explain, so that all reasoning concerning matters of fact seems to be founded on the relation of cause and effect. The causal connection is thus part of a principle of association that operates in our mind. Regular successions of impressions are followed by regular successions of simple ideas, and then imagination orders and conceptualizes successions of simple ideas into complex ideas, thus giving birth to causal relations. The famed problem is that regular successions so established by experience clearly lack the logical necessity we would require for causal successions. Hume’s solution is that if causal relations cannot be established a priori, then they must be grounded in our experience, in particular, in our *psychological habit* of witnessing effects that regularly follow causes in time and space.

If we want causality to be an empirical and testable matter rather than a psychological one, we need to replace the Humean paradigm of regularity with a paradigm of *variation*. In this framework structural models do not only aim at finding regular successions of events. Rather, causal models *model* causal relations by analysing suitable *variations* among variables of interest (see Russo 2006, 2008). Differently put, causal models are governed by a *rationale of variation*, not of regularity. A rationale is a principle of some opinion, action, hypothesis, phenomenon, model, reasoning, or the like. The quest for a rationale of causality is then the search for the principle that guides causal reasoning and thanks to which we can draw causal conclusions. This principle lies in the notion of *variation*.

The rationale of variation manifestly emerges, for instance, in the basic idea of probabilistic theories of causality and in the interpretation of structural equations. Probabilistic theories of causality, see Suppes (1970), focus on the difference between the conditional probability $P(E|C)$ and the marginal probability $P(E)$. To compare conditional and marginal probability means to analyse a statistical relevance relation, i.e. probabilistic independence. The underlying idea is that if C is a cause of E , then C must be statistically relevant for E . Hence, the *variation* hereby produced by C in the effect E will be detected because the conditional and the marginal probability differ. Analogously, quantitative probabilistic theories focus on the difference between the conditional distribution $P(Y \leq y|X \leq x)$ and the marginal distribution $P(Y \leq y)$. Again, to compare conditional distribution with marginal distribution means to measure the variation produced by the putative cause X on the putative effect Y .

In structural equation models, the basic idea is that, given a system of equations, we can test whether variables are interrelated through a set of linear relationships, by examining the *variances* and *covariances* of variables. Sewall Wright, as early as 1934, has taught us to write the covariance of any pair of observed variables in terms of path coefficients. The path coefficient quantifies the (direct) causal effect of X on Y ; given the numerical value of the path coefficient β , the equation $Y = \beta X + \varepsilon$ claims that a unit increase in X would result in a β unit increase of Y . In other words, β quantifies the *variation* of Y associated to a *variation* of X , provided that X doesn't have null variance. Another way to put it is that structural equations attempt to quantify the change in X that accompanies a unit change in Y . It is worth noting that the equality sign in structural equations does not state an algebraic equivalence. Jointly with the associated graph, the structural equation is meant to uncover a causal structure. That is, given a structural equation of the simple form $Y = \beta X + \varepsilon_1$, the reverse equation $X = \gamma Y + \varepsilon_2$ is not causally equivalent. Pearl (2000, pp. 159–160) makes a similar point.

4.1.3 Background Knowledge in Causal Analysis

Variation, however, is not itself a causal notion and consequently cannot guarantee, alone, the causal interpretations of probabilistic inequalities. Good epistemology ought to tell us under what conditions, i.e. what the constraints are, for variations

to be causal. A complete account of the guarantee of the causal interpretation should focus on the difference between purely associational models and causal models, pointing to the features proper to the richer apparatus of causal models (see Russo 2008; Russo 2006). For a model to be causal, we shall particularly focus on two types of constraints: background knowledge and structural stability. In a nutshell, concomitant variations will be deemed causal if they are structurally stable and if they are congruent with background knowledge; see also e.g. Engle et al. (1983), Florens and Mouchart (1985), Hendry and Richard (1983) or Thomas (1996). In this way regularity, which would be better understood here in terms of invariance of the model's structure (variables and relations), becomes a constraint that participates in the causal interpretation of variations.

On the one hand, background knowledge, both theoretical and empirical, serves three roles: (i) it provides a relevant causal context for the formulation of hypotheses, (ii) it guides the choice of variables and of the relations to be tested for structural stability, and (iii) it constitutes the sounding board for results as they have to be congruent with background knowledge. On the other hand, structural stability is a constraint we impose on a relation for being causal, in order to rule out accidental relations. Differently put, the crucial step in Hume's argument is significantly different from the rationale hereby proposed. We claim that we firstly look for variations. Once concomitant variations are detected, a condition of invariance or structural stability (among others) is imposed on them. What does structural stability give us? Not logical necessity, nor mere constant conjunction as Hume advocated. Invariance, which is an empirical feature, recalls Humean regularity but the scope of the former is wider than that of the latter. Structural stability is a condition required in order to ensure that the model correctly specifies the data generating process and that the model does not confuse accidental and/or spurious relations with causal ones. It is worth noting that, in the search for structurality, background knowledge and invariance play a complementary role. In particular, unexplained stable relations may lead to questioning background knowledge and eventually to modifying it.

It might be objected that if structural stability does not give us logical necessity either, it does not any better than regularity. Undoubtedly necessity is an essential feature for those who would like the social sciences to discover universal laws, or for those who question their scientific legitimacy on this ground. However, independently of whether it is a built-in impossibility of the social sciences to glean *laws*, this would be a too rigid framework, for society and individuals are too mutable objects of study to be fettered in immutable and even regular deterministic or probabilistic *laws*.

The philosophical gain of adopting this paradigm is twofold. Firstly, we go beyond the Humean tradition that somehow denies causation by reducing it to regularity. Secondly, we do not fall into untestable metaphysical positions either, because structural models stay at the level of *knowledge*. Let us clarify this last point. Structural modelling intends to represent an underlying causal structure, mathematically, by means of equations, and pictorially, by means of directed acyclic graphs. However, structural models don't pretend to attain the ontic level, i.e. to open the black box, so to speak. They stay at the level of field knowledge and theory: if concomitant

variations between, say, X and Y are structurally stable and are congruent with available field knowledge, then we have no reasons not to *believe* that X causes Y . In this sense structural models mediate epistemic access to causal relations without claiming that the *true* causes have been discovered. Differently put, structural modelling allows us to take a sensible causalist stance that guides actions and policies without overflowing into untestable metaphysical claims.

The practical gain of adopting this paradigm is having a clearer understanding of the causal import of background knowledge and of testing stability. Those aspects, in fact, turn out to be of fundamental importance for the interpretation of results.

4.1.4 Probabilistic Modelling in Causal Analysis

Structural models belong to the category of *probabilistic* models. This leads us to consider also the following issue. Is a *probabilistic* characterization of causation a symptom of indeterministic causality or rather of our incomplete and uncertain knowledge? In physics, quantum mechanics raised quite substantial issues about the possibility of indeterminism. However, *whether or not* the world is actually indeterministic, needs not to be decided once and for all. In fact, from an epistemological viewpoint, a probabilistic characterization of causal relations in structural models only commits us to state that our knowledge is incomplete and uncertain. Our endeavour to gain causal knowledge requires reducing, as far as possible, bias and confounding by building good structural models, that is models that pick up structurally stable relations consistent with background knowledge.

So far we have seen that the concept of variation plays a crucial role in the interpretation of structural equations. A simple form of a structural equation such as $Y = \beta X + \varepsilon$, can be interpreted as follows: variations in X lead to or are responsible for variations in Y . In other words, X is statistically relevant for Y , i.e. $P(Y|X) \neq P(Y)$. However, statistical relevance, and consequently also variation, are symmetrical notions. So how do we know that X causes Y and not the other way around? There are three different but nonetheless related elements that participate in determining the direction of the causal relations: background knowledge, invariance, and time. Let us focus on time. In the social sciences we need temporal direction. This is for several reasons.

Firstly, causal mechanisms – be they physiological, social or socio-physiological – are embedded in time. Smoking at time t causes cancer at t' ($t < t'$), but not the other way around. To give another example, use of contraceptives is followed by changes in the intensity and tempo of fertility. Secondly, although the two causal relations *marriage dissolution influences migration* and *migration influences marriage dissolution* both make sense, we need to know whether marriage dissolution or migration is the temporally prior cause for cognitive and/or policy reasons. One out of the two claims might be eventually disproved due to problems of observability or lack of theory. For instance, the causal chain *migration influences*

marriage dissolution might be incorrect: although marriage dissolution is observed *after* migration, there might exist a temporally prior process – marital problems and the subsequent decision to divorce – causing migration.

This oversimplified example clearly shows that causal modelling requires a constant interplay between observation, theory and testing. Indeed, this is the core of a hypothetico-deductive methodology of structural modelling (see Russo 2008; Russo 2006). Causal hypotheses need to be confirmed or disconfirmed (i.e. accepted or rejected in the statistical jargon) based on empirical testing: the model has to fit observations, but the causal hypothesis itself has to be formulated, along with the model building stage, in accordance with available well established theories and background knowledge. However, we also need structural models to be flexible enough to revise our theories in the light of new data disconfirming prior theories.

Following the H-D methodology, causal hypotheses are confirmed or disconfirmed depending on the results of empirical testing. Suppose, for the sake of the argument, that the causal hypothesis is rejected. Such a negative result can be nonetheless useful as it can suggest that improvement is needed in the theory backing the causal model, or that data may contain some source of bias. In other words, the rejection of a causal hypothesis can trigger further research. Suppose now, again for the sake of the argument, that the causal hypothesis is accepted. Such a positive result is not an immutable one, written on the stone, so to speak. Although the causal hypothesis is not rejected, this may be subject to revision (and even to rejection) in the future, due to new discoveries. It is worth stressing that the acceptance of the causal model is highly dependent on its structural stability. Unlike the traditional falsificationist account (see Popper 1959), hypothetico-deductivism in structural modelling allows and indeed encourages us to use at any stage of research all available information. Williamson (2005) also makes a similar point in putting forward a hybrid of inductive and hypothetico-deductive methodologies in which the hypothesising stage is always informed by previous results, whether positive or negative. This is indeed the advantage of handling structural models that are assumed to represent underlying causal structures without pretending to uncover immutable metaphysical causes. The following sections make more explicit and formal these ideas about causality and structural modelling.

4.2 Structural Modelling

4.2.1 *The Meaning of Structurality*

Inspired by the seminal works of Wright, Haavelmo, Blalock, Pearl and others, we will develop in this section a structural modelling approach to causation. In essence, a model is deemed structural if it uncovers a structure underlying the data generating process. As discussed in Section 4.1.3, this approach systematically blends two ingredients. First, the model must be congruent with background knowledge: modelling the data generating process must be operated in the light of the current

information on the relevant field. Second, the model must show stability in a wide sense: both the structure of the model and the parameters have to be stable or invariant with respect to a large class of interventions or of modifications of the environment. Often, but not always, structural models make use of latent variables. By integrating out the latent variables, the statistical model is thus obtained as the marginal distribution of the manifest or observable variables. It is crucial to note that this concept of structural modelling is wider than the framework of structural equations models, also known as covariance structure models or LISREL type models, widely used in psychology or in sociology, and of simultaneous equations models, widely used in econometrics.

A first consequence of this approach is that the notion of causality becomes relative to the model itself, rather than to the data, as is the case, for instance, in the Granger-type concept of causality. Also, this means that we do not aim at making metaphysical claims about causal relations, but rather at saying when we have enough reasons – specifically, reasons about background knowledge and about structural stability – to believe that we hit upon a causal relation. A second consequence of this model-based concept of causality, involving both background knowledge and stability, is that the model does not simply derive from theory as is often the case in the econometric tradition.

Therefore structural modelling is much more than a sophisticated statistical tool. Good structural modelling ought to be accompanied by a broad and sensible account of what a statistical model is and represents, of what statistical inference is, and of what rationale guides model building and testing. The last point has been dealt with in the previous section. The first and the second will be the object of the following sections. We first recall the formal nature of a statistical model and of the basic concepts of conditional modelling and of exogeneity, we then define the concept of causality in such a framework.

4.2.2 *The Statistical Model*

Formally, a statistical model \mathbf{M} is a set of probability distributions, explicitly:

$$\mathbf{M} = \{S, P^\omega: \omega \in \Omega\} \quad (4.1)$$

where S , called the sample space or observation space, is the set of all possible values of a given observable variable (or vector of variables) and for each $\omega \in \Omega$, P^ω is a probability distribution on the sample space, also called the sampling distribution; thus, ω is a characteristic, also called parameter, of the corresponding distribution and Ω describes the set of all possible sampling distributions belonging to the model. The basic idea is that the data can be analyzed as if they were a realization of one of those distributions. For example, in a univariate normal model, the sample space S is the real line and the normal distributions are characterized by a bivariate parameter, for instance the expectation (μ) and the variance (σ^2); in this case: $\omega = (\mu, \sigma^2)$.

A statistical model is based on a stochastic representation of the world. Its randomness delineates the frontier or the internal limitation of the statistical explanation, since the random component represents what is not explained by the model. A statistical model is made of a set of assumptions under which the data are to be analyzed. Typical assumptions of statistical models are: the observed random variables follow or not identical distributions; the observations are, or are not, independent; the basic sampling distributions are, or are not, continuous and may pertain, or not, to a family characterized by a finite number of parameters (e.g. the normal distributions).

If assumptions are satisfied, the statistical model correctly describes co-variations between variables, but no causal interpretation is allowed yet. In other words, it is not necessary that causal information be conveyed by the parameters, nor is it generally legitimate to give the regression coefficients a causal interpretation. It is worth noting that in specifying the assumptions typical of a statistical model, the problem is not to evaluate whether an assumption is true. A (frequentist) statistician may however want to test in due course whether a hypothesis is confirmed or not. If a model-builder could prove that an assumption were (exactly) true, this would not be an assumption anymore, but a description of the real world. Rather, the main issue is to evaluate whether an assumption is useful, in the sense of making possible a process of learning-by-observing on some aspects of interest of the real world.

4.2.3 Statistical Inference and Structural Models

Statistical inference is concerned with the problem of learning-by-observing and is inductive since it implies drawing conclusions about what has not been observed from what has been observed. Therefore, statistical inference is always uncertain and the calculus of probability is the natural, and in a sense logically necessary tool, (see e.g. de Finetti (1937), Savage (1954)), for expressing the conclusions of statistical inference. Therefore, the stochastic aspect of statistical models corresponds to a stochastic representation of the world and is a vehicle for the learning-by-observing.

Here, two aspects ought to be distinguished. On the one hand, learning-by-observing conveys the idea of learning about some features of interest, namely the characteristics of a distribution or the values of a future realization. On the other hand, learning-by-observing is also concerned with the problem of accumulating information as observations accumulate. These two aspects actually refer to the usefulness of the model. Structural models are precisely designed for making the process of statistical inference meaningful and operational.

To better understand the idea behind this last claim, it is worth distinguishing two families of models. In the first family we find purely statistical models, also called associational or descriptive models, and exploratory data analysis, also called data mining. In these approaches, the assumptions are either not made explicit or restricted to a minimum allowing us to interpret descriptive summaries of data. Interest may accordingly focus on the distributional characteristics of one variable at a

time, such as mean or variance, or on the associational characteristics among several variables, such as correlation or regression coefficients. It is worth noting that the absence or the reduced number of assumptions constituting the underlying model make these associational studies insufficient to infer a causal relation and leaves open a wide scope for interpreting the meaning of the results.

The second family consists in the so-called structural models. “Structural” conveys the idea of a representation of the world that is stable under a large class of interventions or of modifications of the environment. Structural models are also called “causal models”. Here, the concept of causality is internal to a model which is itself stable, in the sense of structurally stable. As a matter of fact, structural models incorporate not only observable, or manifest, variables but also, in many instances, unobservable, or latent, variables. The possible introduction of latent variables is motivated by the help they provide in making the observations understandable; for instance, the notion of “intelligence quotient” or of “associative imagination” might help to shape a model which explains how an agent succeeds in answering the questions of a test in mathematics. Thus a structural model aims at capturing an underlying structure; modelling this underlying structure requires taking into account the contextual knowledge of the field of application. The characteristics, or parameters, of a structural model are of interest because they correspond to relevant properties of the observed reality and can be safely used for accumulating statistical information, precisely because of their structural stability. In this context, a structural model is opposed to a “purely statistical model”, understood as a model that accounts for observable associations without linking those associations to stable properties of the world.

The invariance condition of a structural model is actually a complex issue. Two aspects have to be considered. A first one is a condition of stability of the causal relation. The idea is that each variable depends upon a set of other variables through a relationship that remains invariant when those other variables are subject to external influence. This condition allows us to predict the effects of changes in the environment or of interventions. A second condition is the stability of the distributions to ensure that the parameters will not be affected by changes in the environment or interventions.

4.3 Conditional Models, Exogeneity and Causality

4.3.1 Conditional Models

Originally, the concept of exogeneity appears with regression models. A first, and naive, approach was to consider an exogenous variable as a non-random variable, the endogenous variable being the only random one. That this approach was unsatisfactory became clear when considering complex models where the same variable could be exogenous in one equation and endogenous in another one. A first progress came through a proper recognition of the nature of a conditional model. Here, we

present a heuristic account of the basic concepts; for a more formal presentation, see Mouchart and Oulhaj (2000) and Oulhaj and Mouchart (2003).

Let us start with an (unconditional) parameterized statistical model \mathbf{M}_X^ω given in the following form:

$$\mathbf{M}_X^\omega = \{p_X(x|\omega) : \omega \in \Omega\} \quad (4.2)$$

where for each $\omega \in \Omega$, $p_X(x|\omega)$ is a (sampling) probability density on an underlying sample space corresponding to a (well-defined) random variable X and Ω is the parameter space, aimed at describing the set of sampling distributions considered to be of interest. A conditional model is constructed through embedding this concept into the usual concept of an unconditional statistical model (4.2). For expository purposes, this paper only considers the case where a random vector X of observations is decomposed into $X' = (Y', Z')$ (where \prime denotes transposition) and the model is conditional on Z .

The basic idea of a conditional model is the following: starting from a global model \mathbf{M}_X^ω as given in (4.2), each sampling density $p_X(x|\omega)$ is first decomposed through a marginal-conditional product:

$$p_X(x|\omega) = p_Z(z|\phi) p_{Y|Z}(y|z, \theta) \quad \omega = (\phi, \theta) \quad (4.3)$$

where $p_Z(z|\phi)$ is the marginal density of Z , parametrized by ϕ , and $p_{Y|Z}(y|z, \theta)$ is the conditional density of $(Y|Z)$, parametrized by θ . Next, one makes specific assumptions on the conditional component leaving virtually unspecified the marginal component. Thus a conditional model may be represented as follows:

$$\mathbf{M}_Y^{Z, \theta; \Phi} = \{p_X(x|\omega) = p_Z(z|\phi) p_{Y|Z}(y|z, \theta) \quad \omega = (\theta, \phi) \in \Omega = \Theta \times \Phi\} \quad (4.4)$$

where Φ parametrizes a typically large family of sampling probabilities on Z only and for each $\theta \in \Theta$, $p_{Y|Z}(y|z, \theta)$ represents a conditional density of $(Y|Z)$. The essential features of a conditional model are therefore:

1. θ indexes a well specified family of conditional distributions. This family constitutes the kernel of the concept of a conditional model. The concept of conditional model relates, however, to a family of joint distributions $p_X(x|\omega)$ obtained by crossing the family of conditional densities $p_{Y|Z}(y|z, \theta)$ with a family of marginal distributions $p_Z(z|\phi)$.
2. ϕ is a nuisance parameter which is identified by definition (because Φ is a set of distributions of Z). Furthermore θ and ϕ are variation free. The notation $\mathbf{M}_Y^{Z, \theta; \Phi}$ conveys the idea that θ is the only parameter of actual interest, leaving to ϕ no explicit role.
3. The modelling restrictions are concentrated on the conditional component, i.e. the set $P_Y^{Z, \theta} : \theta \in \Theta$ embodies the main hypotheses of the model, whereas in most cases, the set Φ embodies a minimal amount of restrictions, typically only the hypotheses necessary to guarantee essential properties for the inference on

θ , such as identifiability or convergence of estimators. For instance, in a linear regression model, suitable asymptotic properties of the Ordinary Least Squares estimators require conditions such as stationarity or ergodicity of the process generating the explanatory variables. Consequently, in most situations, but not in all, Φ represents a “thick” subset of the set of all probability distributions of Z . The role of Φ is to stress the random character of Z at the same time as the vague specification of its data generating process; Φ may nevertheless play an important role because its specification may determine desirable properties of the estimators of θ , the parameter of interest. Oulhaj and Mouchart (2003) provides more information on conditional models.

Let us give an example. Consider four variables: tabacism (T), cancer of the respiratory system (C), asbestos exposure (A) and socio-economic status (SES). A global (unconditional) model would consider a family of distributions on the four variables (T, C, A, SES) parametrized by, say, ω , as in (4.2). A conditional modelling approach would run as follows. Suppose we are interested in the impact of T, A and SES on C . Attention would therefore focus on a particular component of the global model, namely the conditional distribution of C given T, A and SES , leaving the marginal distribution of T, A and SES with a minimum amount of specification. In other words, for each distribution indexed by ω in the global model (4.2), we have in mind a marginal-conditional decomposition as in (4.3):

$$p_{C,T,A,SES}(c, t, a, ses | \omega) = p_{C|T,A,SES}(c | t, a, ses, \theta) p_{T,A,SES}(t, a, ses | \phi)$$

$$\omega = (\phi, \theta) \tag{4.5}$$

The basic idea of the conditional model, as in (4.4), is to endow the global model (4.2) with two properties. Firstly, the parameters characterizing the marginal (ϕ) and the conditional (θ) components are independent. Here, “independence” means “variation-free” in a sampling theory framework, i.e. $\omega = (\theta, \phi) \in \Omega = \Theta \times \Phi$, or independent in the (prior) probability in a Bayesian framework, i.e. $\phi \perp\!\!\!\perp \theta$ in Bayesian terms. Secondly the marginal component is left almost unspecified, i.e. the set Φ represents a “very large” set of possible distributions for (T, A, SES).

4.3.2 Conditional Model and Exogeneity

Suppose we analyze data set $X = (Y, Z)$. A challenging issue is to decide whether it is admissible, in the sense of losing no relevant information, to only specify a conditional model $\mathbf{M}_Y^{Z, \theta; \Phi}$ rather than specifying the model \mathbf{M}_X^ω . This is the issue of exogeneity.

The motivation for specifying a conditional model rather than a model on the complete data set X is parsimony: some specifications on the marginal process may not be avoided for ensuring suitable properties of the inference on the parameters of the conditional process but by specifying less stringently the marginal process,

generating Z , one looks for protection against specification error. The cost could however be substantial if the marginal process generating Z contains relevant information, an example of which is given in Section 4.5.1.

Formally, the condition of exogeneity is therefore: the parameter of interest should only depend on the parameters identified by the conditional model and the parameters identified by the marginal process should be “independent” of the parameters identified by the conditional process. It should be stressed that the independence among parameters has no bearing on a (sampling) independence among the corresponding variables.

In order to make the argument more transparent, we slightly modify the notation. In Section 4.3.1 we constructed a model on the X -space, where $X = (Y', Z)'$, by crossing a family of distributions on Z , indexed by ϕ , and a family of conditional distributions on $(Y | Z)$, indexed by θ , and eventually obtained a joint model, parametrized by $\omega = (\phi, \theta)$. We now start from a joint model on X , parametrized by ω , and deduce from the decomposition (4.3) the parameters characterizing the family of marginal distributions of Z , denoted by θ_Z , and the parameters characterizing the family of conditional distributions of $(Y | Z)$, denoted by $\theta_{Y|Z}$. Equation (4.3) is accordingly rewritten as follows:

$$p_X(x | \omega) = p_Z(z | \theta_Z) p_{Y|Z}(y | z, \theta_{Y|Z}) \quad (4.6)$$

where θ_Z , respectively $\theta_{Y|Z}$, represents the parameter identified by the marginal, respectively conditional, process. The condition of independence, namely:

$$(\theta_Z, \theta_{Y|Z}) \in \Theta_Z \times \Theta_{Y|Z} \quad \text{or} \quad \theta_Z \perp\!\!\!\perp \theta_{Y|Z} \quad (4.7)$$

is a *condition of (Bayesian) cut* (see Barndorff-Nielsen (1978) in a sampling theory framework, and Florens et al. (1990) in a Bayesian framework), and is deemed to allow for a separation between the inference on the parameters of the marginal process and the inference on the parameters of the conditional process. More explicitly, condition (4.7) implies that any inference on θ_Z , respectively, $\theta_{Y|Z}$, be based only on the marginal, respectively conditional, model characterized by the marginal distributions $p_Z(z | \theta_Z)$, respectively conditional distributions $p_{Y|Z}(y | z, \theta_{Y|Z})$.

This condition, along with the condition that the parameter of interest, say λ , depends only on the parameters identified by the conditional process, i.e. $\lambda = f(\theta_{Y|Z})$, formalizes the concept of “losing no relevant information” when basing the inference on the conditional model rather than on the complete model, characterized by the distributions $p_X(x | \omega)$. In this setting, the concept of exogeneity appears as a binary relation between a function of the data, namely Z , and a function of the parameters, namely λ . Thus, Florens et al. (1990) suggests the expression “ Z and λ are mutually exogenous” (or Z is exogenous for λ), to stress the idea that a variable is not exogenous by itself but is exogenous in a particular inference problem. Treating Z as exogenous means therefore that the (marginal) process generating Z is minimally specified (and may be heuristically qualified as “left unspecified”) and that

the inference on the parameter of interest, although based on the joint distribution of all the variables in X , is nevertheless invariant with respect to any specific choice of the marginal distribution of Z . Summarizing: exogeneity is the condition that makes admissible the use of the conditional model as a reduction of the complete model.

The consequences of a failure of exogeneity may be twofold. There may be a loss of efficiency in the inference if the failure comes from a restriction (equality or inequality), or a lack of independence in a Bayesian framework, between the parameters of the marginal model and those of the conditional model. There may also be an impossibility of finding a suitable, e.g. unbiased or consistent, estimator if the parameter of interest is not a function of $\theta_{Y|Z}$ only. A typical example, well known in the field of simultaneous equations in econometrics, is that the parameter of interest in a structural equation may not be a function of the parameters identified by the conditional model corresponding to a specific equation.

4.3.3 Exogeneity and Causality

In general, the specification of a parameter of interest is a contextual rather than a statistical issue. A most usual rationale for specifying the parameter of interest is based on the notion of a *structural model*. In this framework, Russo (2006) approach causality as *exogeneity in a structural conditional model*. In the very simple case of two variables Y and Z , this concept may be paraphrased as follows: if the conditional distribution of Y given Z is structurally stable and reflects a good scientific knowledge of the field, there is no reason not to believe that Z causes Y . This approach might be considered empirical because the observations providing the ground for a causal interpretation are not only the data under immediate scrutiny but also the whole body of observations underlying the “field knowledge” and leading accordingly to the present state of scientific knowledge. In this sense, causal attribution “ Z causes Y ” is an issue of structural modelling, namely this is the question whether the conditional model characterized by $p_{Y|Z}(y|z, \theta_{Y|Z})$ is actually *structural*.

4.4 Confounding, Complex Systems and Completely Recursive Systems

4.4.1 Confounders and Confounding

In many circumstances, the same effect can be produced by several causes or the same cause can produce several effects. We may however focus our interest on a particular cause, say X and a particular effect, say Y . In this case, the causal relation $X \rightarrow Y$ can be subject to confounding. In epidemiology and in demography, for example, when one examines the impact of a treatment/exposure on a response/outcome, a confounding variable – or confounder – is often defined as a

variable associated both with the putative cause and with its effect, see e.g. Jenicek and Cl  roux (1982), Elwood (1988). Sometimes the definition is more precise, such as in Anderson et al. (1980) or in Leridon and Toulemon (1997). According to these authors, a variable is a confounder whenever two conditions simultaneously hold:

1. The risk groups differ on this variable;
2. The variable itself influences the outcome.

Some authors gloss condition 1 adding that the variable, as a background factor, should not be a consequence of the putative cause, see e.g. Schlesselman (1982).

For instance, if we examine the impact of cigarette smoking on the incidence of cancer of the respiratory system, a variable such as exposure to asbestos dust confounds the relation between smoking and this type of cancer. Indeed, exposure to asbestos dust and smoking are associated, i.e. proportionally there are more persons exposed to asbestos in the smoking group than in the non-smoking group. Condition 1 is therefore satisfied. In addition, inhalation of asbestos dust is a strong cause of cancer of the pleura; condition 2 is thus also satisfied. Cancer is the outcome variable in this example, smoking a potential cause, and exposure to asbestos a confounder. Vice-versa if one were to examine the impact of asbestos exposure on the incidence of cancer of the respiratory system, smoking this time would be the confounding factor, as it is associated with asbestos exposure and is a cause of lung cancer. This simplified example is discussed in Russo (2006) but a real study would also consider other causal factors and paths, and the synergy between smoking and asbestos exposure.

Condition 1 needs to be clarified however; on this subject, see also McNamee (2003). Why are smoking and asbestos exposure associated? In demography and in epidemiology, one knows that both smoking and asbestos exposure are dependent upon one's socio-economic status (*SES*): those with a lower *SES* tend more to smoke and work in unhealthy environments than those with a higher *SES*. The causal graph can therefore be drawn as in Fig. 4.1, where *A* represents exposure to asbestos, *T* tabacism, and *C* cancer incidence. It is worth noting that Fig. 4.1 incorporates two assumptions, namely: $A \perp\!\!\!\perp T \mid SES$ and $C \perp\!\!\!\perp SES \mid A, T$.

This graph shows that tabacism and asbestos exposure are in fact not independent from one another as they are both related to one's *SES*, i.e. they have a common cause. Note that *SES* is also a common cause of *T* and *C* as it has an impact on cancer through the intervening or intermediate variable *A*. However an association between two variables such as smoking and asbestos exposure could also be due

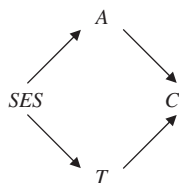


Fig. 4.1 Socio-economic status, smoking, asbestos exposure and cancer of the respiratory system

Fig. 4.2 The relation between T and A , A being an intervening variable between T and C

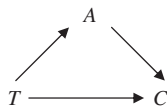
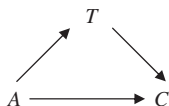


Fig. 4.3 The relation between T and A , A being a common cause of T and C



to a causal relation between them. T could be a cause of A or vice-versa. The two corresponding causal graphs are given in Fig. 4.2 and 4.3 respectively.

This distinction leads to a more precise definition of a confounder: a confounding variable, or confounder, is a variable which is a common cause of both the putative cause and its outcome (Bollen 1989; Pearl 2000; Wunsch 2007). In graphical representations, a common cause is a common ancestor to both putative cause and effect. For example, A is a confounder in Fig. 4.3 because in this model it is a common cause of both T and C . For the same reason, SES is a confounder in Fig. 4.1, as it is a common cause of both T and C (the latter via A). In Fig. 4.2, A is not a common cause of T and C ; therefore A is not a confounder. Notice that confounding is always relative to a particular cause and a particular effect. The confounder can be either latent (i.e. unobserved) or observed; the issue of latent confounding is considered in Section 4.5. This definition avoids taking an intervening (intermediate) variable between the putative cause and the outcome such as in Fig. 4.2 as a confounder, even though it is associated with the putative cause (as the latter has a causal influence on the former) and it has an impact on the outcome.

Judea Pearl (2000) proposes two criteria for controlling confounding bias: the back-door and the front-door. The *back-door criterion* tackles the problem of which variables to control for in cases of possible confounding of a cause (C) and effect (E) relation. A variable or a set of variables Z should be controlled for, according to the back-door criterion, if (i) Z is not a descendant of the cause C and (ii) Z blocks every path between C and E that contains an arrow into C . For example, in Fig. 4.4 taken from Pearl (2000), the sets (X_4, X_3) and (X_3, X_5) meet the back-door criterion by blocking every path between C and E containing an arrow into C , while (X_3) alone does not. The variable X_3 is a collider depending upon the inverted fork X_1 and X_2 . If we condition on X_3 , the variables X_1 and X_2 become dependent

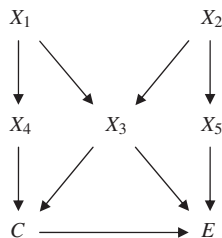


Fig. 4.4 An example of Pearl’s back-door criterion

(Pearl 2000; Wunsch 2007) and thus controlling for the sole variable X_3 does not block the path $(C, X_4, X_1, X_3, X_2, X_5, E)$.

The *front-door criterion* uses the presence of an intervening variable between cause and effect to estimate the causal relation. As an example of Pearl's front-door criterion, consider the relation between smoking and lung cancer. If the impact of smoking on lung cancer is mediated by the amount of tar in the lungs, one can estimate on the one hand, the impact of smoking on the amount of tar and on the other hand, the impact of the amount of tar on lung cancer. If these relations are not confounded by other variables, one can then combine the two effects in order to obtain an estimate of the impact of smoking on lung cancer. If the relations between smoking and tar and between tar and lung cancer are confounded, it is sometimes possible to assess the two relations in the absence of confounding if one can control for another variable causing tar accumulation (such as environmental pollution) which blocks the back-door paths from smoking to tar and from tar to lung cancer. An example is given in Pearl (2000, pp. 67 and 83). An application of the front-door criterion to the more complex problem of the causal effect of Catholic schooling on learning is given in Morgan and Winship (2007, p. 183).

4.4.2 Complex Systems and Completely Recursive Systems

In the previous sections, only small systems of a few variables have been discussed. Let us now consider a decomposition of X into p components: $X = (X_1, X_2, \dots, X_p)$. Once p increases, the analysis sketched above requires more structure because unrestricted systems become quickly unmanageable. In this section, we show how to use field knowledge with the purpose of obtaining a recursive decomposition of complex systems, giving space to further contextually meaningful restrictions.

Suppose that the components of X have been ordered in such a way that in the complete marginal-conditional decomposition:

$$p_X(x|\omega) = p_{X_p|X_1, X_2, \dots, X_{p-1}}(x_p|x_1, x_2, \dots, x_{p-1}, \theta_{p|1, \dots, p-1}) \cdot p_{X_{p-1}|X_1, X_2, \dots, X_{p-2}}(x_{p-1}|x_1, x_2, \dots, x_{p-2}, \theta_{p-1|1, \dots, p-2}) \dots p_{X_1}(x_1|\theta_1) \quad (4.8)$$

each component of the right hand side may be considered as a structural model with mutually independent parameters, i.e. in a sampling theory framework:

$$\omega = (\theta_{p|1, \dots, p-1}, \theta_{p-1|1, \dots, p-2} \dots, \theta_1) \in \Theta_{p|1, \dots, p-1} \times \Theta_{p-1|1, \dots, p-2} \dots \times \Theta_1 \quad (4.9)$$

Equations (4.8) and (4.9) characterize a *completely recursive system*. For $p = 3$, Equation (4.8) may be represented by Fig. 4.5, for $p = 4$ by Fig. 4.6. Once the

Fig. 4.5 First 3 components of a completely recursive system

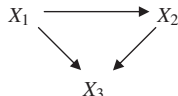
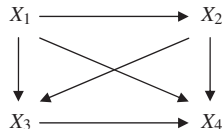


Fig. 4.6 First 4 components of a completely recursive system



value of p increases, graphical representations become quickly unmanageable *unless* some assumptions, in the form of conditional independences, operate simplifications on the system. This is indeed a main issue in structural modelling: field knowledge aims not only at ordering the components of X to obtain (4.8), but also at bringing in more structure than in the complete system (4.8).

More specifically, statistical modelling of complex systems raises several issues:

1. Given a p -dimensional vector of variables to be modelled, is field knowledge sufficient for ordering the variables in such a way that one may obtain a completely recursive system as in (4.8), i.e. in such a way that each component X_j is univariate? It often happens, in particular in econometrics, that it is not possible to disentangle recursively the process generating a vector of variables, in other words that some components X_j are subvectors of X rather than univariate random variables. For instance, Mouchart and Vandresse (2005) handles a case where the data are made of vectors, the components of which are price and attribute of a set a contracts concluded through a bargaining process. The data and the contextual information do not allow to know whether the prices have been bargained after or before the attributes have been agreed upon. This is a case of *simultaneity* where the model describes a process generating a vector of (so-called “endogenous”) variables conditionally on a vector of exogenous variables, in such a way that the equations of the model do not correspond to a marginal-conditional decomposition. The econometric literature, particularly between the Sixties and the Eighties, is rich in developing this class of models, called “simultaneous equation models”.
2. Endowing each distribution of (4.8) with a structural interpretation amounts to saying that each of these distributions represents a contextually relevant data generating process. Parsimony recommends focusing the attention on the processes of actual interest and is made operational by selecting a subvector $(X_{r+s}, X_{r+s-1}, \dots, X_r)$ of X such that the joint distribution of $(X_{r+s}, X_{r+s-1}, \dots, X_r | X_1, \dots, X_{r-1})$ gathers all data generating processes of actual interest. In such a case the subvector (X_1, \dots, X_{r-1}) becomes globally exogenous for the system of interest.

4.5 Partial Observability and Latent Variables

4.5.1 A Three-Component System

In this paper, the concept of causality is not rooted in latent variables, as in the literature on counterfactuals (see for instance Morgan and Winship 2007). However, this section shows that when latent variables are present in a structural model, causal attribution becomes substantially more complex.

Historically, latent variables have been object of interest since at least the Forties and early Fifties, see e.g. Reiersøl (1950), Neyman and Scott (1948, 1951). Latent variables appear in measurement error models and in factor analytic and LISREL type models, among others. Also those models and simultaneous equation models have been shown to be mathematically equivalent as they are all based on the idea that mathematical expectations are required to lie in a linear space (Florens et al. 1976, 1979). The last years have seen a voluminous amount of publications on the large role of latent variables in statistical modelling. Thus Chapter 1 of Skrondal and Rabe-Hesketh (2004) speaks of “the omni-presence of latent variables”, and the book presents an interesting account of methodological issues and of applications. Rabe-Hesketh et al. (2004) suggest how to use a latent variable framework as a unifying device for a large class of models including multilevel and structural equation models.

We begin by considering a three-variate case and next extend the analysis to a p -dimensional vector. Consider a three-variate completely recursive system, represented in Fig. 4.7, for data in the form $X = (Y, Z, U)$:

$$p_X(x|\theta) = p_{Y|Z,U}(y|z, u, \theta_{Y|Z,U}) p_{Z|U}(z|u, \theta_{Z|U}) p_U(u|\theta_U) \quad (4.10)$$

where each of the three components of the right hand side may be considered as structural models with mutually independent parameters, i.e. in a sampling theory framework:

$$\theta = (\theta_{Y|Z,U}, \theta_{Z|U}, \theta_U) \in \Theta_{Y|Z,U} \times \Theta_{Z|U} \times \Theta_U \quad (4.11)$$

This diagram suggests that U causes Z and (U, Z) cause Y . Thus, according to the definition offered above, U is a confounding variable for the effect of Z on Y . Also, Equations (4.10) and (4.11) say that U is exogenous for $\theta_{Z|U}$ and that (U, Z) are jointly exogenous for $\theta_{Y|Z,U}$.

Now suppose that U is not observable. It might be tempting to collapse the diagram in Fig. 4.7 into that of Fig. 4.8. Formally, Fig. 4.8 may be obtained by integrating the latent variable U out of (4.10):

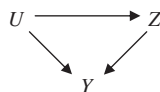


Fig. 4.7 3-component completely recursive system

Fig. 4.8 2-component system



$$p_{Y|Z} (y |z, \theta_{Y|Z}) = \frac{\int p_{Y|Z,U} (y |z, u, \theta_{Y|Z,U}) p_{Z|U} (z |u, \theta_{Z,U}) p_U (u | \theta_U) du}{\int \int p_{Y|Z,U} (y |z, u, \theta_{Y|Z,U}) p_{Z|U} (z |u, \theta_{Z,U}) p_U (u | \theta_U) du dy} \tag{4.12}$$

$$p_Z (z | \theta_Z) = \int p_{Z|U} (z |u, \theta_{Z,U}) p_U (u | \theta_U) du \tag{4.13}$$

Therefore:

$$\theta_{Y|Z} = f_1 (\theta_{Y|Z,U}, \theta_{Z,U}, \theta_U) \quad \theta_Z = f_2 (\theta_{Z|U}, \theta_U) \tag{4.14}$$

Two remarks are in order:

1. In general, Z is not exogenous anymore because (4.14) shows that the parameter $\theta_{Y|Z}$ and θ_Z are, in general, not independent; indeed some components of $\theta_{Z|U}$ and of θ_U may be common to $\theta_{Y|Z}$ and θ_Z . Therefore, Fig. 4.8 is an inadequate simplification of Fig. 4.7 (see however next remark);
2. the non-observability of U typically implies a loss of identification: the functions f_1 and f_2 are *not* one-to-one; thus Z might still be exogenous because potentially common parameters in $\theta_{Y|Z}$ and θ_Z might not be identified;

One might also look for further conditions deemed to recover the exogeneity of Z . A simplifying assumption frequently used is the sampling independence between Z and U :

$$Z \perp\!\!\!\perp U | \theta \tag{4.15}$$

This assumption implies that $\theta_{Z|U}$ is now written as θ_Z and Fig. 4.7 becomes Fig. 4.9 suggesting that U and Z both cause Y (without U causing Z).

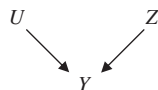
Under condition (4.15), when U is not observable Fig. 4.8 is again obtained under the following integration of U :

$$p_{Y|Z} (y |z, \theta_{Y|Z}) = \int p_{Y|Z,U} (y |z, u, \theta_{Y|Z,U}) p_U (u | \theta_U) du \tag{4.16}$$

Therefore:

$$\theta_{Y|Z} = f_3 (\theta_{Y|Z,U}, \theta_U) \tag{4.17}$$

Fig. 4.9 3-component completely recursive system with marginal independence



is independent of θ_Z and the exogeneity between Z and $\theta_{Y|Z}$ may be recovered. In particular, under condition (4.15), U is not a common cause of Z and Y anymore, but, from (4.17), the meaning of $\theta_{Y|Z}$ comes from a combination of the causal action of U along with that of Z , represented by $\theta_{Y|Z,U}$, and of the distribution of U , represented by θ_U .

An example may be useful to better grasp some difficulties. Suppose, for simplifying the argument, that the joint distribution of X in (4.10) is multivariate normal; thus the regression functions are linear and the conditional variances are homoscedastic, i.e. they do not depend on the value of the conditioning variables. Let us compare the following two regression functions:

$$E[Y | Z, U, \theta_{Y|Z,U}] = \alpha_0 + Z\alpha_1 + U\alpha_2 \quad (4.18)$$

$$\begin{aligned} \alpha_1 &= [\text{cov}(Y, Z | U)] [V(Z | U)]^{-1} \\ &= [\text{cov}(Y, Z) - \text{cov}(Y, U) [V(U)]^{-1} \text{cov}(U, Z)] \\ &\quad \times [V(Z) - \text{cov}(Z, U) [V(U)]^{-1} \text{cov}(U, Z)]^{-1} \end{aligned} \quad (4.19)$$

$$E[Y | Z, \theta_{Y|Z}] = \beta_0 + Z\beta_1 \quad \beta_1 = [\text{cov}(Y, Z)] [V(Z)]^{-1} \quad (4.20)$$

Therefore, if the effect on Y of the cause Z is measured by the regression coefficient, the correct measure would be α_1 rather than β_1 , once the conditional model generating $(Y | Z, U)$ is structural. Note that, in this particular case, $\alpha_1 = \beta_1$ when $Z \perp\!\!\!\perp U$, but this is a particular feature of the normal distribution for which $Z \perp\!\!\!\perp U$ implies that $\text{cov}(Y, Z | U) = \text{cov}(Y, Z)$, and $\text{cov}(Y, U | Z) = \text{cov}(Y, U)$, which is in general not true. Moreover, $\alpha_1 = \beta_1$ is also true when $\alpha_2 = 0$, i.e. when $Y \perp\!\!\!\perp U | Z$, which is contextually different from $Z \perp\!\!\!\perp U$.

This example makes two issues explicit:

- (i) measuring the effect of a cause should be operated relatively to a completely specified structural model; failing to properly recognize this issue may lead to fallacious conclusions because in general: $\alpha_1 \neq \beta_1$
- (ii) prima facie ancillary specifications, such as a normality assumption, may be more restrictive than first thought; indeed, under a normality assumption, the hypotheses $Z \perp\!\!\!\perp U$ and $Y \perp\!\!\!\perp U | Z$ each imply that $\alpha_1 = \beta_1$, although they are contextually different once the normality assumption is not retained. This happens because, in the normal case, independence is equivalent to uncorrelatedness, and because the regression functions are linear.

4.5.2 The General Case

A difficult issue in structural modelling is bound to the fact that many theories in the social sciences involve latent or nonobservable variables. These are introduced in order to help structuring a theoretical framework; think, for instance, of the concept of “anomy” in sociology or of “permanent income” in economy. In such a case, the

initial model includes both latent and manifest or observable variables, from which a statistical model is obtained by integrating out all the latent variables. A typical benefit of such an approach is to obtain a statistical model with more structure, i.e. more restrictions, than a “saturated” statistical model constructed independently of a structural approach. A well-known case is provided by the LISREL type model, or covariance structure model. However this structural approach has also a cost, sometimes difficult to handle. Indeed, the analysis performed around the simplest case of one unobservable variable along with two observable variables, given through Equations (4.12) and (4.13), suggests that the analysis of exogeneity *at the level of the statistical model bearing on the manifest variables only* soon becomes intractable, jeopardizing most exogeneity properties and making the interpretation of the identifiable parameters difficult.

4.6 Discussion and Conclusion

Philosophers have wandered for long time in search of the ultimate *concept* of causality, i.e. in search of what causality in fact *is*. Hume (1748), unable to find what gives logical necessity to causal relations, came to the conclusion that causality is nothing more than a regular succession of events deemed to be causal only thanks to our psychological habit to experience such regular sequences. In his *System of Logic*, John Stuart Mill, as early as 1843, put forward an experimentalist notion of cause. Causes are *physical*, i.e. one physical fact is said to be the cause of another. In the *System of Logic* the experimental approach is seen as the privileged way for ascertaining what phenomena are related to each other as causes and effects. We have, says Mill, to follow the Baconian rule of varying the circumstances, and for this purpose we may have recourse to observation and experiment. Mill believed that his four methods – Method of Agreement, Method of Difference, Method of Residues, and Method of Concomitant Variation – were particularly well suited to natural science contexts but not at all to social sciences. The inapplicability of the experimental method to the social sciences ruled them out straight away from the realm of the sciences and still nowadays leads to a skeptical despair about the very possibility of establishing causal relations in social contexts.

Causal analysis has indeed proved to be a challenging enterprise in the social sciences. There are at least two difficulties in establishing causal relations. A first difficulty is, as just mentioned, that a pure randomized experimentation is rarely possible. A second one, already discussed in the Introduction, is that society and individuals are too mutable to generate “laws of social physics” *à la* Quetelet. However, is this reason enough to give up causal analysis? Should we then content ourselves with Humean regular successions?

Interestingly enough, Durkheim (1895, Chapter VI) strongly argued against the Millian attempt to dismiss social sciences as *sciences* and therefore against any attempt to dismiss causal analysis. In particular, he maintained that the method of concomitant variation is fruitfully used in sociology and indeed this is what makes

sociology scientific. Although an explicit causalist perspective has been adopted by the forefathers of quantitative causal analysis, in more recent times practising scientists have, mistakenly, hardly ever taken a clear stance in this respect. As we have suggested in the opening of this paper, a cognitive goal and an action-oriented goal justify our effort in making causality an empirical and testable, i.e. scientific, matter.

We have argued that structural modelling tries to make causality meaningful and operational and we have seen that this objective can be achieved if two fundamental ingredients are incorporated. The first one is an epistemological element – viz. the rationale of variation, and the second is a methodological element – viz. the concept of *structural* model. Structural modelling aims at uncovering a structure underlying the actual data generating process. Clearly there is an infinity of conceivable structural models leading to a same statistical model “explaining” the data under scrutiny. A main issue for the model builder is selecting one of those structures, taking into account the knowledge of the field and desirable properties of invariance/stability. Thus the practical implication of this paper is twofold. Firstly, causation may be attributed only within a structural model reflecting the state of knowledge of the domain considered. Secondly, the structural stability of the relationships *and* of the parameters of the distributions should be thoroughly checked. This approach is therefore at variance with purely statistical ones where causation is supposedly tested from correlations without making explicit a suitable structural model. Furthermore, causation should not be attributed from a model only based on purely theoretical considerations. Finally, the search for agreement with background knowledge and for structural stability leaves a lesser role to the goodness of fit.

However, although the development of a more adequate rationale of causality and of an accurate concept of structural model give a meaningful framework for causal analysis, we claimed that specific issues still needed to be addressed, e.g. exogeneity and confounding. In this causal framework, the concepts of exogeneity and of confounding have been explicitly defined. On the one hand, exogeneity is a condition of separability of inference that allows us to concentrate on the conditional distribution leaving aside the marginal one. On the other hand, we have adopted a definition of confounders as common ancestors of both cause and effect. However, we have shown that the impact of confounders complicates substantially the analysis and the operational interpretation of exogeneity, because a variable may lose its exogenous status under the impact of a latent confounder. Furthermore, if a latent variable U is a determinant of an outcome Y but is independent of another cause Z of this outcome, Z remains exogenous but, at the level of the manifest variables, the measure of the effect of Z on Y depends upon the original causal effect of Z and upon the distribution of the latent variable U .

Let us now give some general conclusions. In the framework of structural modelling what is the meaning of the claim X causes Y ? Not metaphysical: by means of structural modelling we do not pretend to attain the ontic level and to discover the true and ultimate causes. If causal claims cease to have metaphysical meaning, then they must have an epistemic one: we have reasons to *believe* that X causes Y . Causality thus becomes a matter of *knowledge* generated by the sensible use

of structural modelling. A major task of epistemology and methodology is then to make explicit the conditions under which our causal beliefs are justified and to inform us correctly about causal relations in the world. The net advantage of spousing an epistemic view is to avoid committing to the discovery of the “true” causes or of the “true” model. Instead, causal beliefs are part of our knowledge of the world, and thus are naturally subject to change and improvement.

Acknowledgements The research underlying this paper is part of a research project conducted by the three authors on Causality and Statistical Modelling in the Social Sciences. Parts of this paper have been prepared for the workshop *Causality, Exogeneity and Explanation*, Causality Study Circle – Evidence Project, UCL, London, May 5, 2006, and for the conference *Causal Analysis in Population Studies: Concepts, Methods and Applications* held in Vienna November 30 – December 1, 2007, at the Vienna Institute of Demography, Austrian Academy of Sciences. Financial support to M. Mouchart from the IAP research network nr P5/24 of the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs) is gratefully acknowledged. F. Russo wishes to thank the FSR (Fonds Spécial de Recherche, Université catholique de Louvain), the British Academy, and the FNRS (Belgian National Science Foundation) for financial support. G. Wunsch thanks the Austrian Academy of Sciences for financial assistance. Comments from an anonymous referee are also gratefully acknowledged.

References

- Anderson, S., A. Auquier, W.W. Hauck, D. Oakes, W. Vandaele and H.I. Weisberg (1980). *Statistical Methods for Comparative Studies*. New York: John Wiley & Sons.
- Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*. New York: John Wiley & Sons.
- Bollen, K.A. (1989). *Structural Equations with Latent Variables*. New York: John Wiley & Sons.
- de Finetti, B. (1937). La prévision, ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré* 7: 1–68.
- Durkheim, E. (1895–1912). *Les règles de la méthode sociologique*, 6th edition. Paris: Librairie Félix Arcan.
- Durkheim, E. (1897–1960). *Le suicide*. Paris: Presses Universitaires de France.
- Elwood, J.M. (1988). *Causal Relationships in Medicine*. Oxford: Oxford University Press.
- Engle, R.F., D.F. Hendry and J.-F. Richard (1983). Exogeneity. *Econometrica* 51(2): 277–304.
- Florens, J.-P. and M. Mouchart (1985). Conditioning in Dynamic Models. *Journal of Time Series Analysis* 53(1): 15–35.
- Florens, J.-P., M. Mouchart and J.-F. Richard (1976). *Likelihood Analysis of Linear Models*. CORE Discussion Paper 7619. Université catholique de Louvain, Belgium.
- Florens, J.-P., M. Mouchart and J.-F. Richard (1979). *Specification and Inference in Linear Models*. CORE Discussion Paper 7943. Université catholique de Louvain, Belgium.
- Florens, J.-P., M. Mouchart and J.-M. Rolin (1990). *Elements of Bayesian Statistics*. New York: Marcel Dekker.
- Hendry, D.F. and J.-F. Richard (1983). The Econometric Analysis of Economic Time Series. *International Statistical Review* 51: 111–163.
- Hume, D. (1748). *An Enquiry Concerning Human Understanding*. Indianapolis: Bobbs-Merrill, 1955.
- Jenicek, M. and R. Cléroux (1982). *Epidémiologie*. Paris: Maloine.
- Leridon, H. and L. Toulemon (1997). *Démographie. Approche statistique et dynamique des populations*. Paris: Economica.
- McNamee, R. (2003). Confounding and Confounders. *Occupational and Environmental Medicine* 60(3): 227–234.

- Mill, J.S. (1843). *A System of Logic*. London: Longmans, Green and Co., 1889.
- Morgan, S.L. and C. Winship (2007). *Counterfactuals and Causal Inference*. New York: Cambridge University Press.
- Mouchart, M. and A. Oulhaj (2000). *On Identification in Conditional Models*. Discussion paper DP0015. Institut de statistique, UCL, Louvain-la-Neuve (B).
- Mouchart, M. and M. Vandresse (2007). Bargaining Power and Market Segmentation in Freight Transport. *Journal of Applied Econometrics* 22: 1295–1313.
- Neyman, J. and E. Scott (1948). Consistent Estimates Based on Partially Consistent Observations. *Econometrica* 16: 1–2.
- Neyman, J. and E. Scott (1951). On Certain Methods of Estimating the Linear Structural Relationship. *The Annals of Mathematical Statistics* 22: 352–361. (Corrections: 23 (1952): 135.).
- Oulhaj, A. and M. Mouchart (2003). The Role of the Exogenous Randomness in the Identification of Conditional Models. *Metron* LXI(2): 267–283.
- Pearl, J. (2000). *Causality*. Cambridge: Cambridge University Press.
- Popper, K. (1959). *The Logic of Scientific Discovery*. London: Hutchinson.
- Quetelet, A. (1869). *Physique sociale. Ou Essai sur le développement des facultés de l'homme*. Bruxelles: Muquardt.
- Rabe-Hesketh, S., A. Skrondal and A. Pickels (2004). Generalized Multilevel Structural Equation Modelling. *Psychometrika* 69: 167–190.
- Reiersøl, O. (1950). Identifiability of a Linear Relation Between Variables Which Are Subject to Error. *Econometrica* 18: 375–389.
- Russo, F. (2006). The Rationale of Variation in Methodological and Evidential Pluralism, *Philosophica*, 77, Special Issue on Causal Pluralism, pp. 97–124.
- Russo, F. (2008). *Causality and Causal Modelling in the Social Sciences. Measuring Variations*. New York: Springer.
- Russo, F., M. Mouchart, M. Ghins and G. Wunsch (2006). *Statistical Modelling and Causality in Social Sciences*. Discussion Paper 0601. Institut de Statistique, Université catholique de Louvain.
- Savage, L.J. (1954). *The Foundations of Statistics*. New York: John Wiley.
- Schlesselman, J.J. (1982). *Case-Control Studies – Design, Conduct, Analysis*. New York: Oxford University Press.
- Skrondal, A. and S. Rabe-Hesketh (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Modeling*. Boca Raton, FL: Chapman & Hall/CRC.
- Suppes, P. (1970). *A Probabilistic Theory of Causality*. Amsterdam: North Holland Publishing Company.
- Thomas, R.L. (1996). *Modern Econometrics*. Harlow: Addison-Wesley.
- Wright, S. (1934). The Method of Path Coefficients. *Annals of Mathematical Statistics* 5(3): 161–215.
- Williamson, J. (2005). *Bayesian Nets and Causality*. Oxford: Oxford University Press.
- Wunsch, G. (2007). Confounding and Control. *Demographic Research* 16: 15–35.

Chapter 5

Causation as a Generative Process. The Elaboration of an Idea for the Social Sciences and an Application to an Analysis of an Interdependent Dynamic Social System

Hans-Peter Blossfeld

5.1 Introduction

The empirical investigation of causal relationships is an important but difficult scientific endeavor. In the social sciences, two understandings of causation have guided the empirical analysis of causal relationships: (1) *Causation as robust dependence* and (2) *causation as consequential manipulation*. Both approaches clearly have strengths and weaknesses for the social sciences which will be described in detail in this chapter. Based on this discussion, a third understanding of *causation as generative process*, proposed by David Cox, is then further developed. This idea seems to be particularly valuable for modern social sciences because it leads to a longitudinal analysis of social processes and can easily be combined with a narrative in terms of an actor's objectives, knowledge, reasoning, and decisions (methodological individualism). Using event history models, this approach will then be applied to the causal analysis of an interdependent dynamic social system. In doing so, we first describe parallel processes and time-dependent covariates, the latter of which are often used to include the sample path of parallel processes in transition rate models. The widely used "system" and "causal" approach are contrasted, with the latter proposed as a more appropriate method from an analytical point of view and that it provides straightforward solutions to simultaneity problems, time lags and varying temporal shapes of effects. Based on separate applications in West and East Germany, Canada, Latvia, and the Netherlands, the usefulness of the approach of "causation as generative process" is demonstrated by analyzing two highly interdependent family processes: entry into marriage (for individuals in a consensual union) as the dependent process and first pregnancy/childbirth as the explaining one. After

H.-P. Blossfeld (✉)

Faculty of Social and Economic Sciences, University of Bamberg, Lichtenhaidestr.
11, Bamberg, Germany
e-mail: hpb3007@t-online.de

potential statistical reasons for the time-dependent effects are described, we move to more substantive explanations, including the importance of actors, probabilistic causal relations, preferences and negotiation, observed and unobserved decisions and the problem of conditioning on future events.

5.2 Models of Causal Inference

The goal to find scientifically based evidence for causal relationships leads to design questions, such as which inference model is appropriate to specify the relationship between cause and effect and which statistical procedures can be used to determine the strength of that relationship (Schneider et al. 2007). Two different models of causal inference have dominated the work of practitioners in the social sciences over the last decades: (1) *Causation as robust dependence* and (2) *causation as consequential manipulation*. The former approach – which in multiple regression or path analysis is known as the “control variable” or “partialling” approach (Duncan 1966; Kerlinger and Pedhazzer 1973; Blalock 1970) and in the econometric analysis of time-series as *Granger causation* (Granger 1969; Johnston 1972) – starts from the presumption that correlation does not necessarily imply causation but causation must in some way or the other imply correlation. In this view, the key problem of causal inference is to determine whether an observed correlation of variable X with variable Y , where X is temporally prior to Y , can be established as a “genuine causal relationship.”

The advocates of the *causation as robust dependence* approach call X a “genuine” cause of Y in so far as the dependence of Y on X cannot be eliminated through additional variables being introduced into the statistical analysis. Thus, in this approach causation is established essentially through the elimination of spurious (or non-causal) influences. Although this approach has dominated the social sciences for several decades, sociologists consider it as a too limited approach. First, they think that causal inference should not be limited entirely to a matter of statistical predictability but should include predictability in accordance with theory (Goldthorpe 2001: 3). Second, since scientists rarely know all of the causes of observed effects or how they relate to one another, it is impossible to be sure that all other important variables have in fact been controlled for (Shadish et al. 2002). A variable X can therefore never be regarded as having causal significance for Y in anything more than a provisional sense: “At any point, further information might be produced that would show that the dependence of Y on X is not robust after all or, in other words, that the apparent causal force of X is, at least to some extent, spurious” (Goldthorpe 2001: 5).

The second understanding of *causation as consequential manipulation* seems to have emerged as a reaction to the limitations of *causation as robust dependence*. Instead of “establishing the causes of effects,” Holland (1986, 1988) and Rubin (1974, 1978, 1980) are concerned with “establishing the effects of causes.” They make clear that it is more to the point to take causes simply as given, and to con-

centrate on the question of how their effects can be securely measured. According to this approach, causes can only be those factors that could serve as treatments or interventions in well-designed controlled experiments or quasi-experiments. Thus, given appropriate experimental controls, if a causal factor X is manipulated, then a systematic effect is produced on the response variable Y . The particular strength of this design is that “. . .while statements in the form ‘ X is a cause of Y ’ are always likely to be proved wrong as knowledge advances, statements in the form ‘ Y is an effect of X ’, once they have been experimentally verified, do not subsequently become false: ‘Old, replicable experiments never die, they just get reinterpreted’” (Goldthorpe 2001: 5).

Understood in this way, causation is always relative in the sense that the specific treatment of X_{tr} and its observed outcome Y_{tr} are compared with what would have happened to the same unit if it had not been exposed to this treatment (counterfactual account of causality). Since it is not possible in the same experiment for a unit to be both exposed and not exposed to the treatment, the conception of *causation as consequential manipulation* leads to what Holland (1986) has called the “fundamental problem of causal inference”. For example, a student who completes one mathematics program cannot go back in time and complete a different program so that we can compare the two outcomes. Thus, the question arises of how we make sure that one gets convincing measurements for something that is in fact impossible to measure, i.e., the outcome Y_{con} for a unit in the control group which had not been exposed to the treatment (X_{con}) in the same experiment?

In the hard sciences, such as physics or chemistry, it is often relatively easy to conduct strictly controlled experiments and to demonstrate, based on the qualities of the objects under study (e.g., physical entities), what would have happened (Y_{con}) to the same unit (u) of analysis if it had not been exposed to the treatment (X_{con}). In other words, it is often plausible to assume that these objects have a constancy of response over time (temporal stability) and that the effect of the first treatment is transient and does not affect the object’s response to the second treatment (causal transience). Or at least, that the physical entities or chemical substances respond in a similar way under certain conditions. In these cases, the causal effect for u , $CauEff_u$, is easily defined as $CauEff_u = Y_{tr} - Y_{con}$. In fact, the model of *causation as consequential manipulation* based on the well-designed controlled experiment has been quite successful in the hard sciences.

In other disciplines such as biology, medicine or psychology, it is often not possible to assume temporal stability and causal transience at the level of the individual unit, and it is normally impossible to eliminate the impact of confounding influences at the individual level. For these sciences, Rubin and Holland suggest a statistical approach to the fundamental problem of causal inference: rather than focusing on specific units, this approach estimates an average causal effect for a population of units: $CauEff = E(Y_{tr} | X_{tr}) - E(Y_{con} | X_{con})$, where $E(Y_{tr} | X_{tr})$ is the expected value for participants in the treatment group, and $E(Y_{con} | X_{con})$ is the expected value for participants in the control group. For this solution to work, however, participants in the treatment and control groups should differ only in terms of treatment group assignment, not on any other variables that might potentially affect their re-

sponses. The approach to make sure that this is indeed the case is the randomized experiment, where participants are randomly assigned to the treatment and control conditions, so that one can expect that treatment group assignment would, on average, over repeated experiments, be independent of any measured or unmeasured pretreatment characteristics (Fisher 1935). In randomized experiments treatment assignment and unit response are therefore statistically independent of each other and any kind of selection bias is eliminated. However, it must be noted that the average causal effect of randomized experiments in populations with different distributions might be quite different, so that the effect of a randomized controlled experiment is strongly context-dependent, too (see Rohwer, 2007, unpublished). For example, when experimenters use convenience samples (e.g. if they use university students as experimental units), the outcome might differ from the outcome of an experiment based on a random sample from the larger population (Agresti and Franklin 2007: 170 pp).

In sociology, economics, and demography, however, the situation under which causal inferences have to be drawn is often even more complex and complicated than in the disciplines mentioned above. In particular, randomization is often practically or socially unacceptable (e.g., it is morally and legally impossible to assign twins at birth randomly to different social classes in order to study the impact of various social environments on school success). In addition, strict experimental controls are hard to apply. Thus, well-designed randomized controlled experiments or quasi-experiments are rarely applied by practitioners in the social sciences and most demographic and sociological causal inference is based on non-experimental observations of social processes.

Since these observational data are often highly selective, Rubin, Holland and others subscribing to the approach of *causation as consequential manipulation* recommend that in their empirical work social scientists should make the process of unit assignment itself a prime concern of the inquiry. In particular, social scientists should attempt to identify, and then to represent through covariates in their data analyses all unobserved and observed influences on the response variable that could conceivably be involved in, or follow from, this unit assignment process (Goldthorpe 2001). A difficulty at once obvious here is that of how one demonstrates that given a constellation of covariables, treatment assignment and unit response are indeed independent of each other. Thus the question arises: Have all relevant variables been included and adequately measured and controlled?

A whole battery of statistical techniques has been developed to help to approximate randomized controlled experiments with observational data (Schneider et al. 2007). These methods include fixed effect models (i.e., the adjustment for fixed, unobserved individual characteristics), instrumental variables (i.e., a method to correct for omitted variables bias due to unobserved characteristics), propensity score matching (an approach where individuals are matched on the basis of their observed aggregate characteristics), and regression discontinuity designs (where samples and comparisons between groups be restricted to individuals who fall just above or below a specific cut-off point and, at the same time, are likely to be similar on a set of unobserved variables). Yet, however valuable these techniques might be, "...it

is still difficult to avoid the conclusion that, in non-experimental social research, attempts to determine the effects of causes will lead not to results that ‘never die’ but only to ones that have differing degrees of plausibility. . . . (Or in other words), such results will have to be provisional in just the same way and for just the same reasons as those of attempts to determine the causes of effects via the ‘partialling’ approach” (Goldthorpe 2001: 6). It therefore seems that the benefits of the approach of *causation as consequential manipulation* in the social sciences is quite limited.

Another and even more serious issue for social scientists arises from the insistence of the exponents of the *causation as consequential manipulation approach* that causes must be manipulable (by an experimenter or intervener – at least in principle) (e.g., Holland 1986). Here it is not important, whether one requires a situation of strong manipulation, as Holland (1986) does, or only a situation where it is possible to conceive of a world where units of analysis receive treatment rather than control for different reasons (e.g. choice, force, happenstance etc.) as Heckman and Vytlačil (2005) do. The basic idea is that once the treatment or intervention is introduced, it will quasi automatically lead to an outcome: $X_{tr} \rightarrow Y_{tr}$. There is no explicit idea of how treatment and control are translated in a time-dependent way by acting individuals. The units of analysis in the social sciences, the individuals, are therefore assumed to be passive subjects whose behavior is explained only by causal factors and their “. . . objectives, knowledge, reasoning and decisions have no further relevance” (Goldthorpe 2001: 8). This understanding of causation clearly reduces the testability of relevant theories and models in the social sciences. In particular, it seems not to be compatible with the micro-foundation of modern sociological theory where actors are considered to have agency, where individuals have objectives and knowledge and, when faced with a choice between different courses of action, will make decisions. Thus, the *causation as consequential manipulation approach* has a limited bearing for social scientists who have conceptionally moved from so-called factor-based to so-called actor-based models (Macy 1991; Macy and Willer 2002).

This limiting understanding of *causation as consequential manipulation* is particularly obvious, if dynamic social systems are studied over longer time-spans. Life course researchers have demonstrated that by studying lives over substantial periods of time they increase their opportunities to understand and explain the lives within their changing social context, including relationships, workplaces, schools, and communities (Elder et al. 2004). Individuals and their purposeful actions are embedded and shaped by the historical times so that the same event may differ in substance and meaning across different birth cohorts. The same events may also affect individuals in different ways depending on when they occur in the life course. Lives are also lived interdependently so that events in one person’s life often entail events for other people as well. Thus, lives cannot be adequately represented when removed from relationships with significant others. It is well known that individuals’ objectives, knowledge and beliefs are influenced by the interactions with others over time (Hedström 2005). It is therefore theoretically important to study dynamic social systems as processes over substantial periods of time.

These issues lead us to the third understanding of *causation as generative process*. According to Cox (1990, 1992) it is crucial to the claim of a causal link that

there is an elaboration of an underlying, generative process existing in time and space. A causal association between X and Y must be considered as being produced by a process and is created by some (substantive) mechanism. A major shortcoming of the approaches of *causation as robust dependence* and *causation as consequential manipulation* is that there is no explicit notion of an underlying generative process present in these models. Thus, *causation as generative process* seems to be a necessary expansion of these two understandings of causation (Goldthorpe 2001).

In summary, causal inference clearly should not be limited entirely to a matter of statistical predictability as in the *causation as robust dependence* approach. Well-designed controlled experiments or quasi-experiments would be a great study design for causal inference, but since in the social sciences randomization is often practically or socially unacceptable, they are rarely applied by their practitioners. Thus, most demographic and sociological causal inference has to be based on non-experimental observations of social processes. Under these conditions, both approaches, *causation as consequential manipulation* and *causation as generative process*, need also to eliminate spurious (or non-causal) influences and will therefore never lead to results that “never die” but only to ones that have differing degrees of plausibility. Finally, the *causation as generative process approach* has the comparative advantage that it focuses our thoughtful consideration on the theoretical and statistical elaboration of an underlying, generative causal process existing in time and space, including also actors who make decisions within social contexts.

In the present contribution, I would like to explore what the approach of *causation as generative process* has to offer to empirically working social scientists who wish to engage in the causal analysis of dynamic systems using event history data. Event history models are linked very naturally to an understanding of *causation as generative process* because the transition rate provides a local, time-related description of how the process evolves in time (Blossfeld et al. 2007: 33). For each point in time, these models try to predict future changes of the transition rate of the dependent process on the basis of events of independent processes in the past. Of course, this concentration on event history analysis does not imply that there might also be other tools in the statistical arsenal of longitudinal data analysis (from simple growth curve models to full-blown generalized equation models) which allow to apply the *causation as generative process* approach.

5.3 Parallel and Interdependent Processes

The study of parallel or interdependent processes with transition rate models is one of the most important advances of event history analysis (Willekens 1991; Courgeau and Lelièvre 1992; Blossfeld and Rohwer 2002; Blossfeld et al. 2007). Parallel or interdependent processes can operate at a variety of different levels. There may be interdependent or parallel processes at the level of:

- *different domains of an individual's life*. For instance, one may ask how upward and downward moves in an individual's job career influences her/his family trajectory (e.g. Blossfeld and Huinink 1991).

- *individuals interacting with each other*, termed “interdependent or linked lives” (Elder 1987). One might study the effect of the career of the husband on his wife’s labour force participation (Blossfeld and Drobnič 2001) or how the death or migration of the head of the household impacts other family members (Courgeau and Lelièvre 1992).
- *intermediate organizations*, such as how the changing household structure determines women’s labour force participation.
- *macro processes*, where the researcher may be interested, for instance, in the effect of changes in the business cycle on family formation (e.g. Blossfeld and Huinink 1991).
- *any combination of the aforementioned processes*. For example, in the study of life course, cohort, and period effects, time-dependent covariates measured at different levels must be included simultaneously (Blossfeld 1986; Mayer and Huinink 1990). Such an analysis combines processes at the individual level (life course change) with two kinds of processes at the macro level: (1) variations in structural conditions across successive (birth, marriage, etc.) cohorts; and, (2) changes in particular historical conditions affecting all cohorts in the same way.

In event history analysis, time-dependent covariates are often used to include the sample path of parallel processes in transition rate models. In the literature, however, only two types of time-dependent covariates have been described as not being subject to reverse causation (see e.g. Kalbfleisch and Prentice 1980; Tuma and Hannan 1984; Blossfeld et al. 1989; Yamaguchi 1991; Courgeau and Lelièvre 1992). The first are *defined* time-dependent covariates whose total time path (or functional form of change over time) is determined in advance in the same way for all subjects under study. For example, process time like age or duration in a state (e.g., duration of marriage in divorce studies), is a defined time-dependent covariate because its values are predetermined for all subjects. It is the predefined onset of the process when the individual becomes “at risk” in the event history model. Thus, by definition, the values of these time-dependent covariates cannot be affected by the dependent process under study. The second type are *ancillary* time-dependent covariates whose time path is the output of a stochastic process that is external to the units under study. Again, by definition, the values of these time-dependent covariates are not influenced by the dependent process itself. Examples of time-dependent covariates that are approximately external in the analysis of individual life courses are variables that reflect changes at the macro level of society (unemployment rates, occupational structure, etc.) or the population level (composition of the population in terms of age, sex, race, etc.), provided that the contribution of each unit is small and does not really affect the structure in the population (Yamaguchi 1991).

In contrast to defined or ancillary time-dependent covariates are *internal* time-dependent covariates, which are often referred to as being problematic for causal analysis in event history models (e.g. Kalbfleisch and Prentice 1980; Tuma and Hannan 1984; Blossfeld et al. 1989; Yamaguchi 1991; Courgeau and Lelièvre 1992). An internal time-dependent covariate Y_t^B describes a stochastic process, considered

in a causal model as being the cause, that in turn is affected by another stochastic process Y_t^A , considered in the causal model as being the effect. Thus, there are direct effects in which the processes autonomously affect each other (Y_t^B affects Y_t^A and Y_t^A affects Y_t^B), and there are “feedback” effects, in which these processes are affected by themselves via the respective other processes (Y_t^B affects Y_t^B via Y_t^A and Y_t^A affects Y_t^A via Y_t^B). In other words, such processes are interdependent and form what has been called a dynamic system (Tuma and Hannan 1984). Interdependence is typical at the individual level for processes in different domains of life and at the level of a few individuals interacting with each other (e.g., career trajectories of partners) (see Blossfeld and Drobnič 2001). For example, the empirical literature suggests that the employment trajectory of an individual is influenced by his/her marital history and marital history is dependent on the employment trajectory. In the literature, there are two central approaches to modelling these processes, what we term here as the “system approach” and the “causal approach,” with the former often used to deal with such dynamic systems.

5.3.1 Interdependent Processes: The System Approach

The system approach in the analysis of interdependent processes (Tuma and Hannan 1984; Courgeau and Lelièvre 1992) defines change in the system of interdependent processes as a new “dependent variable.” Thus, instead of analyzing one of the interdependent processes with respect to its dependence on the respective others, the focus is on the modelling of a system of state variables. In other words, the interdependence between the various processes is taken into account only implicitly.

Suppose that there are J interrelated qualitative time-dependent variables (i.e., processes): $Y_t^A, Y_t^B, Y_t^C, \dots, Y_t^J$. A new time-dependent variable (or process) Y_t , representing the system of these J variables, is then defined by associating each discrete state of the ordered J -tuple with a particular discrete state of Y_t . As shown by Tuma and Hannan (1984), as long as change in the entire system only depends on the various states of the J qualitative variables and on exogenous variables, this model is identical to modelling change in a single qualitative variable. Thus, the idea of this approach is to simply define a new joint state space, based on the various states spaces of the coupled qualitative processes, and then to proceed as in the case of a single dependent process.

Although the system approach provides insights into the behaviour of the dynamic system as a whole, it has several disadvantages. First, from a causal analytical point of view, the approach presented by Courgeau and Lelièvre (1992) does not provide direct estimates of effects of coupled processes on the process under study. In other words, when using the system approach, one normally does not know to what extent one or more of other coupled processes affect the process of interest, controlling for other exogenous variables and the history of the dependent process. Since the effects can only be identified in simple models via a comparison of the constant terms of hazard rate equations, it is only possible to compare transition rates

for general models without covariates (see Courgeau and Lelièvre 1992; Blossfeld and Rohwer 2002). Second, a mixture of qualitative and quantitative processes, in which the transition rate of a qualitative process depends on the levels of one or more metric variables, presents a problem in this approach. Tuma and Hannan (1984) suggest that in these situations it is not very useful. Third, this approach is unable to handle interdependencies between coupled processes occurring in specific phases of the process (e.g., processes might be interdependent only in specific phases of the life course) or interdependencies that are dynamic over time (e.g., an interdependence might be reversed in later life phases, see Courgeau and Lelièvre 1992), what Tuma and Hannan (1984) term “cross-state dependence”. Finally, the number of origin and destination states of the combined process Y_t , representing the system of J variables, may lead to practical problems. Even when the number of variables and their distinct values is small, the state space of the system is large. Therefore, in light of the increase in the number of parameters with the system approach, the event history data sets must contain a large number of events, even if only the most general models of change (i.e., models without covariates) are to be estimated. Considering these limitations, Blossfeld and Rohwer (2002) therefore suggested a different perspective in modelling dynamic systems, which they call the “causal approach”.

5.3.2 *Interdependent Processes: The Causal Approach*

The underlying idea of the causal approach for analyzing interdependent processes can be outlined as follows (Blossfeld and Rohwer 2002). Based on theoretical reasons, the researcher focuses on one of the interdependent processes and considers it as the dependent one. The future changes of this process are linked to the present state and history of the entire dynamic system as well as to other exogenous variables (see Blossfeld 1986; Blossfeld and Huinink 1991). Thus, in this approach the variable Y_t , representing the system of joint processes at time t , is not used as a multivariate dependent variable. Instead, the history and the present state of the system are seen as a condition for change in (any) one of its processes. The question of how to give a more precise formulation for the causal approach remains. The following ideas may be helpful.

Causes and time-dependent covariates. As discussed above, Holland (1986) developed the idea that causal statements imply counterfactual reasoning: If the cause had been different, there would have been another outcome, at least with a certain probability. However, the consequences of conditions that could be different from their actual state are obviously not empirically observable. This means that it is simply impossible to observe the effect that would have happened on the same unit of analysis, if it were exposed to another condition at the same time.

To find an empirical approach to examine longitudinal causal relations, Blossfeld and Rohwer (2002) suggested the examination of conditions which actually do change in time, controlling for other factors. These changes are characterized as events or transitions. More formally, an event is specified as a change in a variable, and this change must happen at a specific point in time. The most obvious

empirical representation of causes is therefore in terms of quantitative or qualitative variables that can change their states over time. These variables are easily included as *time-dependent covariates* in event history analysis. The role of a time-dependent covariate in this approach is to indicate that a (qualitative or metric) causal factor has changed its state at a specific time and that the unit under study is exposed to another causal condition. From this point of view, it seems somewhat *misleading to regard whole processes as causes*. Rather, *only events, or changes in state space can sensibly be viewed as possible causes*.

Time and casual effects. Consequently, we do not suggest that process Y_t^A is a cause of process Y_t^B , but that a *change* in Y_t^A could be a cause (or provide a new condition) of a *change* in $Y_{t'}^B$. Or, more formally: $\Delta Y_t^A \rightarrow \Delta Y_{t'}^B, t < t'$, meaning that a change in variable Y_t^A at an earlier time t is a cause of a change in variable $Y_{t'}^B$ at a later point in time, t' . Of course, it is not implied that Y_t^A is the only cause which might affect $Y_{t'}^B$. We speak of causal conditions to stress that there might be, and normally is, a quite complex set of causes (see Marini and Singer 1988). Thus, if causal statements are studied empirically, they must intrinsically be related to time, which relates to three important aspects of *causation as generative process*:

First, to speak of a change in variables necessarily implies reference to a time axis. We need at least two points in time to observe that a variable has changed its value. Of course, approximately, we can say that a variable has changed its value *at a specific point in time*. Therefore, we use new symbols to refer to changes in the values of the time-dependent variable ΔY_t^A and the state variable $\Delta Y_{t'}^B$ at time t and t' . This leads to the important point that *causal statements relate changes in two (or more) variables*, if we think in terms of *causation as generative process*.

Second, we must consider *time ordering, time intervals and apparent simultaneity*. Time ordering assumes that cause must precede the effect in time: $t < t'$, in the formal representation given above, an assumption which is generally accepted (Eells 1991: Chapter 5). As an implication, the *causation as generative process approach* must specify a temporal interval between the change in the variable representing a cause and the corresponding effect (Kelly and McGrath 1988). The finite time interval may be very short or very long, but can never be zero or infinity (Kelly and McGrath 1988). In other words, in *time-continuous event history models there can never be simultaneity of the causal event and its effect event*.

Some effects take place almost instantaneously. However, some effects may occur in a time interval that requires small time units (e.g., microseconds) or are too small to be measured by any given methods, so that cause and effect seem to occur at the same point in time. Apparent simultaneity is often the case where temporal intervals are relatively crude such as, for example, yearly data. For example, the events “first marriage” and “first childbirth” may be interdependent, but whether these two events are observed simultaneously or successively depends on the degree of temporal refinement of the scale used in making the observations. Other effects need a long time until they start to occur. Marini and Singer (1988), for example, discuss the gap between mental causal priority and observed temporal sequences of behaviour. Thus, there is a *delay or lag between cause and effect* (see Fig. 5.1) that must be specified in an appropriate model of *causation as generative process*.

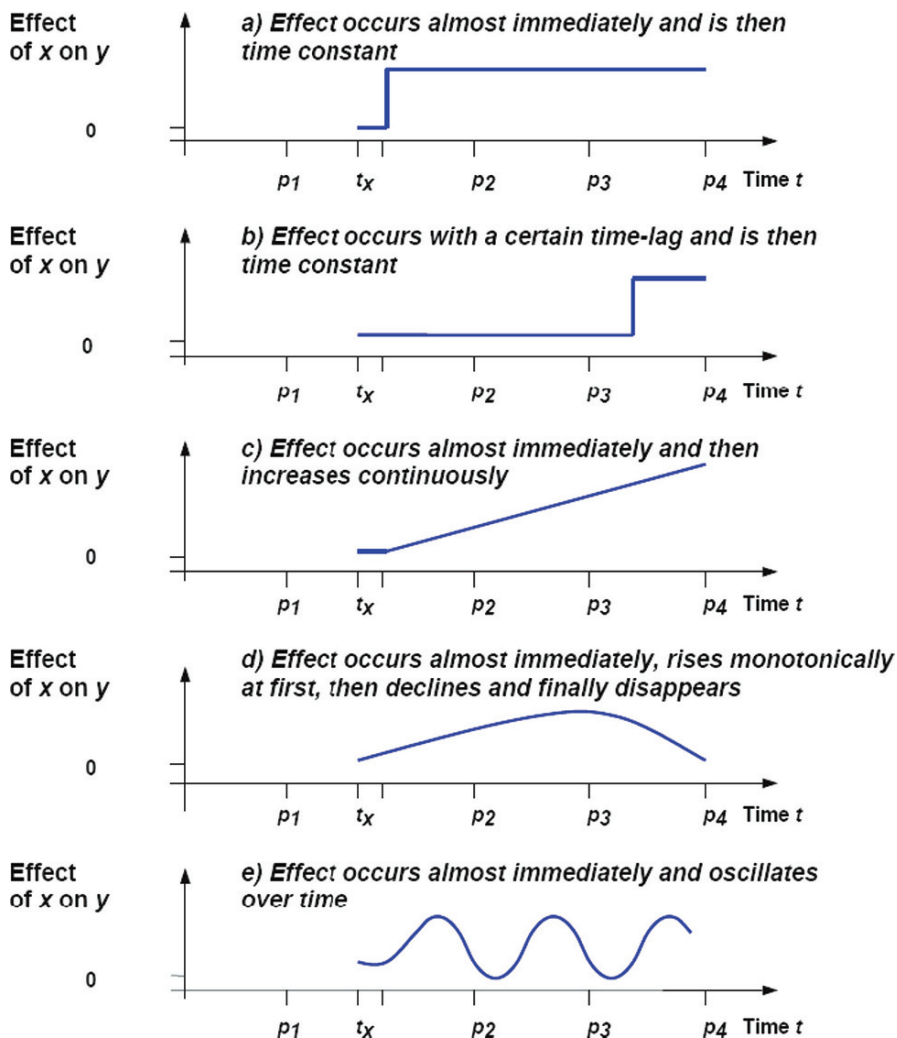


Fig. 5.1 Hypothetical temporal lags and effect shapes

Unfortunately, in most of the current social science theories and interpretations of research findings, this interval is left conceptionally unspecified.

This leads to the *third* point of *causation as generative process*: *temporal shapes of the unfolding effect*. This means that there might be different shapes of how the causal effect Y_t , unfolds over time (see Fig. 5.1). While the problem of time-lags is widely recognized in the social science literature, little attention has been given to the temporal shapes of effects in the social sciences (Kelly and McGrath 1988). Researchers (using experimental or observational data) often seem to either ignore or be ignorant about the fact that causal effects could be highly time-dependent, which, of course, is an important aspect of *causation as generative process*. For instance in

Fig. 5.1a, there may be an immediate impact of change that is then maintained (this obviously is the idea underlying the approaches of *causation as robust dependence* and *causation of consequential manipulation* because there is no explicit notion of an underlying generative process present in these models). Or, the effect could occur with a lengthy time-lag and then become time-invariant (see Fig. 5.1b). The effect could start almost immediately and then gradually increase (see Fig. 5.1c) or there may be an almost all-at-once increase which reaches a maximum after some time and then decreases (see Fig. 5.1d). Finally, there could exist a cyclical effect pattern over time (see Fig. 5.1e). Thus, based on these examples it is clear that we cannot rely on the assumption of eternal, time-less laws but have to recognize that the causal effect may change during the development of social processes. Since the approaches of *causation as robust dependence* and *causation of consequential manipulation* do not have an explicit idea of an underlying generative process in time and space, it might happen that the timing of observations in observational or experimental studies (see for example the arbitrary chosen observation times p_2 , p_3 , or p_4 in Fig. 5.1) lead to completely different empirical evidences for causal relationships.

The principle of conditional independence. Suppose we consider only interdependent processes that are not just an expression of another underlying process so that it is meaningful to assess the properties of the two processes without regarding the underlying one (control variable approach). This means, for instance, that what happens next to Y_t^A should not be directly related to what happens to Y_t^B , at the same point in time, and vice versa. This condition, which we call “local autonomy” (see Pötter and Blossfeld 2001), can be formulated in terms of the uncorrelatedness of the prediction errors of both processes, Y_t^A and Y_t^B , and excludes stochastic processes that are functionally related.

Combining the ideas above, a causal view of parallel and interdependent processes becomes easy, at least in principle. Given two parallel processes, Y_t^A and Y_t^B , a change in Y_t^A at any (specific) point in time t' may depend on the history of both processes up to, but not including t' . Or stated in another way: what happens with Y_t^A at any point in time t' is conditionally independent of what happens with Y_t^B at t' , conditional on the history of the joint process $Y_t = (Y_t^A, Y_t^B)$ up to, but not including, t' . Of course, the same reasoning can be applied if one focuses on Y_t^A instead of Y_t^B as the “dependent variable”. This is the principle of conditional independence for parallel and interdependent processes.

The same idea can be developed more formally. Beginning with a transition rate model for the joint process, $Y_t = (Y_t^A, Y_t^B)$ and assuming the principle of conditional independence, the likelihood for this model can be factorized into a product of the likelihoods for two separate models: a transition rate model for Y_t^A which is dependent on Y_t^B as a time-dependent covariate, and a transition rate model for Y_t^B which is dependent on Y_t^A as a time-dependent covariate. Estimating the effects of time-dependent (qualitative and metric) processes on the transition rate can be easily achieved by applying the method of episode-splitting (Blossfeld et al. 1989; Blossfeld and Rohwer 2002; for a detailed explanation in relation to this analysis see also Mills (2000)).

The conditional independence assumption has important implications for the modelling of event histories. From a technical point of view there is no need to distinguish between defined, ancillary, and internal covariates because all of these time-dependent covariate types can be treated equally in the estimation procedure. A distinction between defined and ancillary covariates on the one hand and internal covariates on the other is however sensible from a theoretical perspective, because only in the case of internal covariates does it make sense to examine whether parallel processes are independent, whether one of the parallel processes is endogenous and the other ones are exogenous, or whether parallel processes form an interdependent system (i.e., they are all endogenous). We will now present empirical applications that illustrate the viability of the approach of causation as generative process to interdependent dynamic systems.

Joint determination of interdependent processes. The principle of conditional independence implies that the prediction errors (or residuals) of the correlated processes Y_t^A and Y_t^B are uncorrelated, given the history of each process up to t and the covariates. In practice, however, there may be time-invariant unmeasured characteristics that affect both Y_t^A and Y_t^B leading to a residual correlation between the processes. In that case, we say that the two processes are jointly determined by some unmeasured influences. Suppose, for example, that we are interested in studying the relationships between employment transitions and fertility among women. We might expect that a woman's chance of making an employment transition at t would depend on her childbearing history up to t (e.g. the presence and age of children), and that her decision on whether to have a(nother) child at t would depend on her employment history up to t . There may be unobserved individual characteristics, fixed over time, that affect the chances of both an employment and a fertility transition at t . For example, more "career-minded" women may delay childbearing and have fewer children than less "career-minded" women. In the absence of suitable measures of "career-mindedness", this variable would be absorbed into the residual terms of both processes, leading to a cross-process residual correlation. If the residual correlation cannot be explained by time-dependent and time-invariant covariates, the two processes should be modelled simultaneously and multiprocess models (Lillard and Waite 1993) have been developed for this purpose.

Unobserved heterogeneity. Often we are not able to include all important factors into the event history analysis. One reason is the limitation of available data; we would like to include some important variables, but we simply do not have the information. Furthermore, we often do not know what is important. So what are the consequences of this situation? Basically, there are two aspects to be taken into consideration. The first one is well-known from *causation as robust dependence*. Because our covariates are often correlated, the parameter estimates depend on the specific set of covariates included in the model. Every change in this set is likely to change the parameter estimates of the variables already included in previous models. Thus, as in the *causation as robust dependence* approach the only way to proceed is to estimate a series of models with different specifications and then to check whether the estimation results are stable or not. Since our theoretical models are normally

weak, this procedure can provide additional insights into what may be called context sensitivity of causal effects in the social world.

Second, changing the set of covariates in a transition rate model will very often also lead to changes in the time-dependent shape of the transition rate. A similar effect occurs in traditional regression models: Depending on the set of covariates, the empirical distribution of the residuals changes. But, as opposed to regression models, where the residuals are normally only used for checking model assumptions, in transition rate models the residuals become the focus of modelling. In fact, if transition rate models are reformulated as regression models, the transition rate becomes a description of the residuals, and any change in the distribution of the residuals becomes a change in the time-dependent shape of the transition rate (see Blossfeld et al. 2007). Consequently, the empirical insight that a transition rate model provides for the time-dependent shape of the transition rate more or less depends on the set of covariates used to estimate the model. So the question is whether a transition rate model can provide at least some reliable insights into a time-dependent transition rate.

The transition rate that is estimated for a population can be the result (a mixture) of quite different transition rates in the subpopulations. What are the consequences? First, this result means that one can “explain” an observed transition rate at the population level as the result of different transition rates in subpopulations. Of course, this will only be a sensible strategy if we are able to identify important subpopulations. To follow this strategy one obviously needs observable characteristics to partition a population into subpopulations. Although there might be unobserved heterogeneity (and we can usually be sure that we were not able to include all important variables), just to make more or less arbitrary distributional assumptions about unobserved heterogeneity will not lead to better models. On the contrary, the estimation results will be more dependent on assumptions than would be the case otherwise (Lieberson 1985). Therefore, we would like to stress our view that the most important basis for any progress in model building is sufficient and appropriate data.

There remains the problem of how to interpret a time-dependent transition rate from a causal view. The question is: Can time be considered as a proxy for an unmeasured variable producing a time-dependent rate, or is it simply an expression of unobserved heterogeneity, which does not allow for any substantive interpretation? There have been several proposals to deal with unobserved heterogeneity in transition rate models, which cannot be developed here (see e.g. Tuma and Hannan 1984; Blossfeld et al. 2007). Furthermore, fixed-effects methods have become increasingly popular in the analysis of event history data in which repeated events are observed for each individual. They make it possible to control for all stable characteristics of the individual, even if those characteristics cannot be measured (Yamaguchi 1986; Allison 1996; Steele 2003; Zhang and Steele 2004). All these models broadly enrich the spectrum of models and can be quite helpful in separating robust estimation results (i.e., estimation results that are to a large degree independent of a specific model specification) and “spurious” results, which might be defined by the fact that they heavily depend on a specific type of model.

5.4 An Application Example

In order to demonstrate the utility of the *causation as generative process* approach to interdependent dynamic systems, we report the results of three cross-national comparative studies about the effect of first pregnancy/first birth on entry into first marriage for couples living in consensual unions. The earliest investigation was conducted by Blossfeld et al. (1993), followed by Blossfeld et al. (1996, 1999) and finally, Mills and Trovato (2001).

The basic research problem underpinning these studies can be defined as follows. Historically, marriage has preceded the birth of a child in many countries. However, in the last two decades, the link between marriage and childbirth has become more complex, a phenomenon that has occurred in conjunction with a rapid rise in consensual unions. The three studies explored this relationship by examining how the experience of a pregnancy within a consensual union conditioned the likelihood of transition to a formal marriage with the same partner. In the later investigations, the process was modeled as explicitly time-dependent, with entry into first marriage as the dependent and first pregnancy/childbirth as the explaining process. The theoretical framework used by the authors to guide a substantive explanation of the time-dependent process was the rational actor model, which proposes that norm-guided and rational self-centered behaviour co-exist.

5.4.1 *The Blossfeld-Manting-Rohwer Study*

The purpose of the earlier study by Blossfeld et al. (1993) was to gain insight into the process of how consensual unions were transformed into marriages in the former West Germany and the Netherlands. It focused on the effect of fertility on the rate of entry into marriage, controlling for other important covariates in a transition rate model. Nationally representative longitudinal data were used: the German Socio-economic Panel (West Germany) and the Fertility Survey (Netherlands) were applied. Both data sets provide information about the dynamics of consensual unions in the 1980s. Attention was limited to cohorts born between 1950–1969 that started a consensual union between 1984–1989 (West Germany) and 1980–1988 (Netherlands).

Recall that a change in the marriage process at any point in time during a consensual union may depend on the history of both processes up to, but not including t' .¹ Thus, given appropriate statistical controls, a change in the marriage process at time t' is conditionally independent of what happens with the fertility process at t' , conditional on the history of the joint process up to, but not including t' . The likelihood for the joint process of first marriage and birth can therefore be factorized into a product of the likelihoods for two separate transition rate models for: (1) first

¹ We are viewing each of these two processes as having various states in their histories. For example, the partnership process could consist of the states of never married, consensual union, married and the pregnancy/birth process may consist of the states of not pregnant, pregnant and first child.

pregnancy/first birth, dependent on first marriage as a time-dependent covariate; and, (2) first marriage, dependent on first pregnancy/first birth as a time-dependent covariate.

We will discuss only the fertility effects of one transition model from this study, which utilized a piecewise constant exponential model to estimate transitions from consensual unions to both marriage and dissolution (results not shown here, see Blossfeld et al. 1993). The change in the fertility process was included as a series of time-dependent dummy variables with the states: “not pregnant”, “pregnant”, “first childbirth”, and “6 months after birth”. The effects of the fertility variables on the marriage rate were significant for both countries and worked in the same direction. As long as women were not pregnant, they observed a significant and comparatively low rate of entry into marriage. But, as soon as a woman in both countries became pregnant (and in West Germany also around the time when the woman gets her child), the rate of entry into marriage increased strongly. If the couple did not get married within six months after the child was born, the rate of entry into marriage again dropped to a comparatively low level in West Germany. In the Netherlands, this level is even below the “not pregnant” level (see Manting 1994).

5.4.2 The Blossfeld-Klijzing-Pohl-Rohwer Study

About a year after this comparative study was conducted, Blossfeld et al. (1996, 1999) wanted to examine whether these results could be replicated with other data from the German Fertility and Family Survey. These data were collected retrospectively from respondents aged 20–39 years in West and East Germany in 1992. They started with a simple model of the process of entry into first marriage for couples living in consensual unions using only one time-dependent dummy variable for the event of first birth. However, the effect of this covariate was – surprisingly – not significant. What happened to the fertility effect? After much theoretical discussion, a hypothesis was put forward that could explain the seemingly contradictory results of the estimated models: the effect of changes in fertility on entry into marriage *must be strongly time-dependent in a very specific way*. According to the first study, the rate is low as long as women are not pregnant, then starts to rise at some time shortly after conception, increases during pregnancy to a maximum and finally drops again a few months after birth has taken place. Thus, when a time-dependent covariate was switched at the time of childbirth, a period with a low marriage rate up to the time of discovery of conception and a period with a high marriage rate during pregnancy was confounded and compared with a relatively low rate after the birth had taken place. Thus, the aggregated average tendency to marry before the child is born could equal the aggregate average tendency to marry after the child is born, therefore making the estimated coefficient of the time-dependent covariate “childbirth” not significantly different from zero.

To deal with this problem, a series of 14 time-dependent pregnancy/birth binary variables were created using information from the reported date of first birth (see Table 5.1). These variables were grouped into categories ranging from “marriage

Table 5.1 Partial likelihood estimates of the transition from consensual union to marriage (final model), West and East Germany, Canada, Latvia, the Netherlands

Covariates	Final model results by country				
	West Germany	East Germany	Canada	Latvia	The Netherlands
Pregnancy/birth process (1)					
[time before pregnancy]	-1.2595	-0.6179	-1.0768	-1.3918	-1.0909
month of pregnancy	0.1131	0.1729	-0.1157	0.3822	-0.2217
1 month since pregnancy	0.4783	0.2715	0.7107	0.2009	0.3769
2 months since pregnancy	0.8837*	0.4225	1.0851*	1.0109*	0.9374*
3 months since pregnancy	1.0260*	0.7723*	0.5849	1.2959*	1.3229*
4 months since pregnancy	0.8578*	1.3903*	0.6563	1.0817*	1.5587*
5 months since pregnancy	0.9905*	0.7938*	0.2480	0.9328*	1.0743*
6 months since pregnancy	0.8701*	0.1510	-0.8948	0.7525*	0.0227
7 months since pregnancy	0.8158*	-0.5166	-0.0365	0.4793	0.1028
8 months since pregnancy	-0.8121*	-2.5449*	-0.5693	-0.4727	-0.2350
Month of birth	-1.4709	-0.6254	-0.1115	-1.6669	-1.2711
1-3 months after birth	-0.7513	0.2875	0.0096	-0.0136	-0.4595
4-6 months after birth	-0.7638	0.1351	0.0363	-1.3576*	-0.4404
More than 7 months after birth	-0.9877*	-0.0921	-0.5263*	-1.2336*	-1.6771*
Birth cohort (2)					
1965-69	-0.3094	-0.6001*	-0.4341*	-1.3096*	-2.2829*
1960-64	-0.1700	-0.0536	-0.3589*	-0.8563*	-1.4258*
1955-59	-0.1486	0.0920	-0.4324*	-0.6154	-0.8228*
[1950-54]	0.0	0.0	0.0	0.0	0.0
Historical period					
[Before 1974]	0.0	0.0	0.0	0.0	0.0
1974-83	0.0882	0.3521	-0.3027	0.0010	-0.2488
After 1983	-0.1554	0.0363	-0.2905	-0.3164	-1.7642*
Highest education level					
Low	0.1722*	-0.0189	0.1563	-0.0164	0.2490*
[Medium]	0.0	0.0	0.0	0.0	0.0
High	-0.0354	0.0941	-0.1092	-0.0763	-0.1962*
Educational enrollment					
In school	-0.3575*	0.0061	-0.3187	0.2700	-0.1856
[Out of school]	0.0	0.0	0.0	0.0	0.0

* = significant at the 0.05 level. Results are shown for the final model. (1) First covariate coded as centered effects, all others as cornered effects. Reference groups denoted by brackets. (2) Birth cohorts for West and East Germany are represented by 1968-72, 1963-67, 1958-62 and 1953-57. *Source*: Blossfeld et al. (1999) for West and East Germany and Mills and Trovato (2001) for Canada, Latvia and the Netherlands. Both the pregnancy/birth and educational enrollment variables are time-dependent.

before the month of pregnancy”, “month of the pregnancy”, “one month since pregnancy”, and so on, to “more than seven months after birth”. To be clear, since no information on the timing of pregnancy and only on the timing of successful births was available, we were looking backward in time from the first birth and thus estimated the date of pregnancy as nine months before the date of birth. As we discuss in greater detail shortly, this presents two potential problems: neglecting abortions and miscarriages, and conditioning past on future events.

5.4.3 *The Mills-Trovato Study*

Building on the previous two studies, Mills and Trovato (2001) wanted to see if the findings would hold in other diverse contexts such as North America or Eastern Europe or during a more recent time period within Western Europe. For this reason, we selected Canada and Latvia and more recent data from the Netherlands. Replication using diverse contexts provides a harsher and more useful validation than statistical testing of many models on only one data set. Normally, there is less chance of an artefact, more kinds of variation can be explored, and alternative explanations can be ruled out (Freedman 1991).

A further impetus for this study centered on the fact that consensual unions and non-marital births in Eastern Europe and the Baltic States have skyrocketed since the 1980s (Katus 1992). Yet, these countries are rarely included in comparative analyses. Similarly, we questioned whether this type of behaviour would still hold in the North American context in a country such as Canada. Using data from the Fertility and Family Surveys (FFS) for Canada (1995), Latvia (1995) and the Netherlands (1993), we selected a comparative sample of women born between 1950 and 1969.

Table 5.1 summarizes the results of the partial likelihood estimates from the Cox models for the transition from consensual union to marriage and for the final models from the Blossfeld et al. (1999) and Mills and Trovato (2001) studies. Figure 5.1 plots the final partial likelihood estimates (coefficients) for the time-dependent pregnancy/birth process variable. Overall, the findings suggest a high degree of uniformity, though the levels and significance of effects tend to vary slightly across countries. Notwithstanding these similarities, we acknowledge that the Canadian and East German case show a few unexpected effects on the transition rate. In Canada, the likelihood appears to drop earlier, at approximately three months before birth, with fluctuations after that point. We attribute this largely to methodological factors since some of the monthly data had to be partially estimated (see Mills and Trovato 2001). In East Germany, there is a large drop one month before birth as opposed to the month of birth. Difference in the significance level of results by country (especially Canada and East Germany) may also be related to smaller sample sizes and less events. The theoretical reasons behind the generally comparable effects that we observe across the five areas are central to understanding these investigations.

5.5 Substantial Explanations

We just speculated about these time-dependent fertility effects in *statistical* terms, which does not, however, explain why we should expect these time-dependent effects in *substantive* terms at all. How can this effect found across a variety of countries be explained? Before we give a more detailed answer to this question, some more general remarks about actors and probabilistic causal relations in the *causation as generative process* approach are in order.

5.5.1 *Actors, Probabilistic Causal Relations and the Hazard Rate*

“When an analysis becomes causal, social regularities represent the effects for which causes have to be discovered. And this task, contrary to what proponents of the idea of causation as robust dependence would seem to have supposed, cannot be a purely statistical one but requires a crucial subject-matter input” (Goldthorpe 2001: 11). Today, there is a general consensus that demographic and sociological phenomena are always directly or indirectly based on actions and interactions of individuals (methodological individualism). We do not deal with associations among variables per se, but with variables that are associated via acting people (see Blossfeld and Prein 1998; Blossfeld et al. 2007).

There are at least three consequences for explanations of causal relations. First, if individuals relate causes and effects through actions and interactions, then explanation of demographic processes should be related to *individuals*. This is why life history data on individuals, and not aggregated longitudinal data, provide the most appropriate empirical evidence for causal relationships. Second, explaining or understanding of demographic processes requires: (1) a *time-related* specification of structural constraints which cut down the set of abstractly possible courses of action to a vastly smaller subset of feasible actions; and, (2) a *mechanism* that singles out which of the feasible courses of action shall be realized (see Elster 1979). Because this is done by individuals, this mechanism must rest on the beliefs, expectations, and motivations of the agents. (3) Since individuals are the actors, causal inference must also take into account their *free will*.

This introduces an essential element of indeterminacy into causal inferences. Hence, in demography and sociology we can only reasonably account for and model the generality but not the determinacy of behaviour. The aim of substantive (and statistical) causal models in the social sciences must therefore be to capture common elements in the behaviour of people, or patterns of action that recur in many cases (Goldthorpe 1998, 2000). A narrative of action must be provided that captures the main tendencies that arise in similar situations. This theoretical model must not seek to explain the behaviour of single individuals, but abstract ideal-typical actors (Hedström 2005: 38). As Stinchcombe (1968) has shown, the behaviour of large aggregates can be reasonably well comprehended, even when the individual components of the aggregate are poorly understood. Given this macro-level focus, small idiosyncratic deviations from the postulated model are not damaging (Hedström 1995). The consequence, however, is that in demographic applications, randomness has to enter as a defining characteristic of causal models.

We can only hope to make sensible causal statements about how a given or (hypothesized) change in variable Y_t^A (e.g., pregnancy/birth) in the past affects the probability of a change in variable $Y_{t'}^B$ (e.g., marriage) in the future. Correspondingly, the basic causal relation becomes: $\Delta Y_t^A \rightarrow \Delta \Pr(\Delta Y_{t'}^B)$, $t < t'$. In other words, a change in the time-dependent covariate Y_t^A will change the probability that the dependent variable $Y_{t'}^B$ will change in the future ($t < t'$).

In demography and sociology, this interpretation seems more appropriate than the traditional deterministic approach. The essential difference is not that our knowledge about causes is insufficient allowing only probabilistic statements, but that the causal effect to be explained is a probability. Thus, probability in this context is not just a technical term anymore, but is considered a theoretical one: it is the propensity of social agents to change their behaviour intentionally.

Using continuous event history data and hazard rate models, the causal reasoning underlying our approach can therefore be restated in a somewhat more precise form as: $\Delta Y_t^A \rightarrow \Delta r(t'), t < t'$. As a causal effect, the changes in covariates Y_t^A in the past may lead to changes in the time-dependent transition rate $r(t')$ in the future, which in turn describes the propensity that the actors under study will change their course of action. This causal interpretation requires that we take the temporal order in which structural constraints and the actors' beliefs and motivations evolve in time very seriously.

5.5.2 Diffuse Marriage Preferences and the Negotiation Process

With regard to the marriage decision in our example study, it seems important to distinguish two completely different situations at the time of the discovery of the pregnancy: (1) the preferences of the partners to marry are vague and diffuse; and, (2) the couple has already reached a decision to marry or not to marry in the case of child. In the first instance, the occurrence of a pregnancy may initiate a process of preference formation and persuasion. *Formation* means that initially rather vague preferences with regard to marriage are formed, resulting in more clear-cut preferences in a step-wise negotiation process. *Persuasion* means that an individual is led by a sequence of short-term improvements into preferring marriage over non-marriage, even if he or she has initially vaguely preferred non-marriage over marriage. In such cases the discovery of a pregnancy engenders a time-structured process of reasoning and interactions which results in a change in preferences. On the one hand, the opportunity to legalize the birth of the child tends to decrease with the duration of pregnancy. At the same time, the likelihood of possible medical complications connected with the pregnancy and the visibility of pregnancy to others increases.

With these contradicting factors in mind, the optimal time for marriage is at a relatively early pregnancy phase. On the other hand, the optimum in the sense of a safe, well thought-out decision based on a negotiation process between the partners, is often at a relatively later phase of the pregnancy. Thus, there is constant tension between these opposing forces that may often (but not necessarily) be connected to a considerable shift in preferences with regard to marriage. Based on these contradictory forces, one would expect that the rate of entry into marriage after the discovery of pregnancy at first increases with the duration of pregnancy and then, after reaching some maximum, decreases again as the time of birth comes

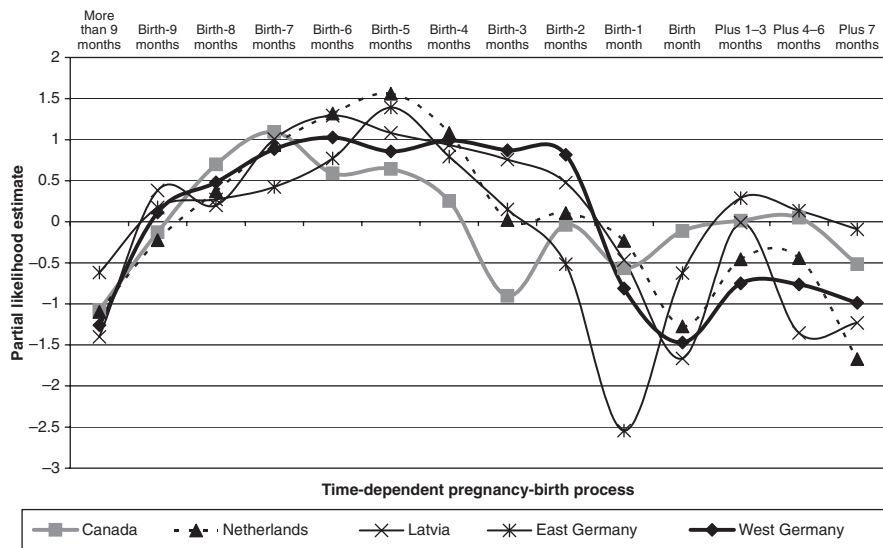


Fig. 5.2 Comparison of partial likelihood estimates (coefficients) of the transition from consensual union to marriage, West and East Germany, Canada, Latvia and the Netherlands

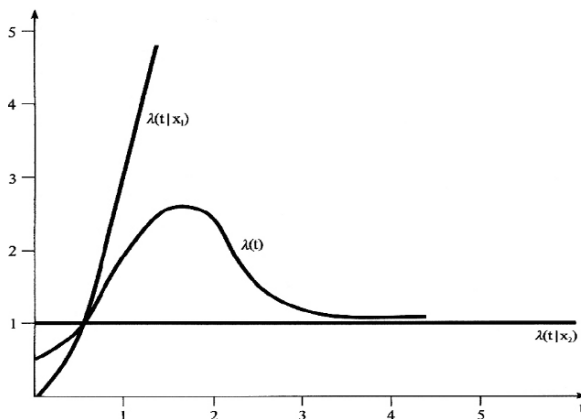
closer. Shortly before and after the birth, one would expect a very low marriage rate. Finally, after the birth has already taken place out of wedlock, the decision of whether or not to marry has a different social meaning. The child is then already “illegitimate”, and the normative time pressure to marry has disappeared, thus resulting in a relatively low marriage rate after some time since the birth of the child.

Table 5.1 and Fig. 5.2 illustrate that after controlling for several important covariates, women in consensual unions do indeed seem to follow this pattern with respect to the rate of entry into marriage: the marriage rate is very low before pregnancy across all countries; it generally increases strongly up to about 5 months before birth, then falls deeply around the time of birth, and is finally at a relatively low level more than 7 months after the birth. Therefore, our substantive interpretation of the time-dependence in Table 5.1 is derived from a theoretically supposed underlying negotiation process with the time-dependent dummy-variables serving as proxies for a theoretically important process that is hard (or even impossible) to measure.

5.5.3 Unobserved Marriage Decisions and the Observed Rate of Entry into Marriage

Of course, one could also argue that many couples have already reached a decision to marry or not to marry in the case of a child at the time of the discovery of the

Fig. 5.3 Marriage rate for couples who had already decided to marry before the event of pregnancy $\lambda(t | x_1)$, marriage rate for couples who had decided that they would never marry before the event of pregnancy $\lambda(t | x_2)$, and observed marriage rate $\lambda(t)$, if these two subpopulations are not controlled for in the model



pregnancy. Thus, couples are in fact extremely *heterogeneous* with regard to their baseline rate to enter into marriage when the pregnancy is observed. Consider the example where the consensual union population consists of two groups – one with a constantly low marriage rate $\lambda(t | x_2)$ and the other with an increasing rate as pregnancy progresses $\lambda(t | x_1)$ (see Fig. 5.3). This neglected heterogeneity would result in a bell-shaped marriage rate $\lambda(t)$ (see Fig. 5.3).

This is due to the fact that when pregnancy progresses, the composition of the unmarried couples shifts towards couples being “less” or “not” ready for marriage which, at first, increases and then decreases the observed effect pattern. Thus, if we do not know whether the couples have already reached a decision to marry in the case of a child at the time of pregnancy, we are unable to say whether the effects of the dummy variables must be considered as proxies for the formation of couples’ decisions during pregnancy, or for the heterogeneity of couples’ marriage decisions at the beginning of pregnancy. Obviously, in reality both interpretations may be valid. The important conclusion is, however, that the discovery of a pregnancy leads to a changing marriage rate for most couples.

5.5.4 Abortion, Miscarriage and the Problem of Conditioning on Future Events

Another methodological problem is that we have not considered abortion and miscarriage. Couples can avoid the birth of children (and therefore marriage) by abortion, and they can experience a miscarriage. Both groups present a problem for our causal analysis because we do not have any information about abortion and miscarriages in our data sets and have constructed the fertility variables on the basis of successful births. In other words, there is the danger that we have committed one of the most serious methodological errors in causal analysis: We have conditioned past events on future events, reversing the temporal order of cause and effect. As long as conditions are random and concern only a small proportion of couples, as

is the case with miscarriages, this objection is not exceedingly important. We get biased estimates only if specific couples sort themselves out by choice in greater numbers, as is probably the case with abortion. In particular, we overestimate the size of the pregnancy/birth effect because we systematically underrepresented pregnant couples that would not have wanted to marry because of a child in our “risk set of pregnant couples” effect (i.e., if we overestimate, then the effect is negative on the rate which gives a downward bias). In former East Germany and Latvia, abortion was easier and more socially accepted than in the other countries. In Latvia, abortion is a widespread method of fertility control with 111 terminated pregnancies per 100 live births and stillbirths in 1991 (Government of Latvia 1999: 125).

5.6 Summary and Concluding Remarks

Two understandings of causation have guided the empirical analysis of causal relationships in the social sciences: (1) *Causation as robust dependence* and (2) *causation as consequential manipulation*. On the one hand, our discussion of both approaches made clear, that the idea of *causation as robust dependence* is too limited because causal inference cannot be a purely statistical consideration. Rather it requires a crucial subject-matter input. On the other hand, the idea of *causation as consequential manipulation* requires well-designed randomized controlled experiments or quasi-experiments. Since such designs can only rarely applied in the social sciences, most demographic and sociological causal inference is based on non-experimental observations of social processes. These data are often highly selective. A whole battery of statistical techniques has therefore been developed to help to approximate randomized controlled experiments with observational data. However, it is still difficult to avoid the conclusion that, non-experimental social research, will lead to results that “never die” but only to ones that have differing degrees of plausibility. Thus such results will have to be provisional in just the same sense and for just the same reasons as those of attempts to determine the causes of effects via the *causation as robust dependence* approach.” Furthermore, the approach of *causation as consequential manipulation* is still too restrictive for modern social sciences because the idea is that once the treatment or intervention is introduced, it will automatically lead to an outcome. The units of analysis in the social sciences, the individuals, are therefore assumed to be passive subjects whose behavior is explained only by causal factors. This restricted understanding of *causation as consequential manipulation* is particularly problematic, if dynamic social systems are studied over longer time-spans. A necessary augmentation of the two understandings of causation is therefore the idea of *causation as generative process*, proposed by David Cox. According to this view, it is crucial to the claim of a causal link that there is an elaboration of an underlying (substantive) generative process existing in time and space.

The main aim of this paper was to further develop the idea of *causation as generative process* and to demonstrate the viability of this understanding in a cross-national empirical investigation of interrelated family events. The story these

empirical studies tell is persuasive. In substantive terms, the investigations confirm the existence of a highly time-dependent causal process between pregnancy and marriage for individuals in consensual unions across five different national contexts. In particular, it shows that the force of an empirical analysis results from the clarity of the prior substantive reasoning and the bringing together of seemingly contradictory evidence. All studies have been instructive in methodological terms because: (1) they *analyzed two highly interdependent processes* from a causal point of view, (2) the interdependence occurs mainly in a very *specific phase of individuals' lives* (i.e., family formation), (3) the relationship between cause and its effect involves *time lags* (e.g., time until detection of pregnancy); and, (4) the unfolding *effect is highly dynamic over time*.

These applications illustrate the substantive importance and methodological pitfalls of the identification of time-dependent causes and their time-dependent effect patterns. A central contribution is that we have been able to demonstrate that one process is influencing or causing a change in the other – even if they are interdependent. In cross-sectional data, we often have interdependent systems with feedback mechanisms, but are unable to discern how one process influences the other. We witness associations that describe what has happened, but cannot separate the effect. Associations are quite different from causal statements designed to say something about how events are produced or conditioned by other events. With the event history approach, however, it becomes possible to separate correlation and causation (Blossfeld and Rohwer 2002).

One shortcoming is that our applications are only based on observed behaviour. It could happen that a couple first decides to marry, the woman becomes pregnant, and then the couple marries. In this case, we would observe only pregnancy occurring before marriage and assume that it increases the likelihood of marriage. Yet, the time order is exactly the other way around. Courgeau and Lelièvre (1992) have introduced the notion of “fuzzy time” to represent this time span between decisions and behaviour. Since the time between decisions and behaviour is probably not random and differs per couple, examining observed behaviour could lead to false causal inferences. This does not alter the key temporal issues embedded within the causal logic. However, we must admit that using the time order of only behavioural events without taking into account the timing of decisions could lead to serious misspecification. Thus, for studies aiming to model *causation as a generative process* through the relationship between individuals' objectives, knowledge, reasoning and decisions over time, prospective panel observations of objectives and decisions and retrospective information on behavioural events appear to be a very desirable design.

References

- Agresti, A. and C. Franklin (2007). *Statistics. The Art and Science of Learning from Data*. Upper Saddle River, NJ: Pearson Prentice Hall.
- Allison, P.D. (1996). Fixed-effects partial likelihood for repeated events. *Sociological Methods & Research* 25: 207–222.

- Blalock, H.M. (ed.) (1970). *Causal Models in the Social Sciences*. Chicago, IL: Aldine.
- Blossfeld, H.-P. (1986). Career opportunities in the Federal Republic of Germany: A dynamic approach to the study of life-course, cohort, and period effects. *European Sociological Review* 2: 208–225.
- Blossfeld, H.-P., A. Hamerle and K.U. Mayer (1989). *Event History Analysis*. Hillsdale: Erlbaum.
- Blossfeld, H.-P. and J. Huinink (1991). Human capital investments or norms of role transition? How women's schooling and career affect the process of family formation. *American Journal of Sociology* 97: 143–168.
- Blossfeld, H.-P., D. Manting and G. Rohwer (1993). Patterns of change in family formation in the Federal Republic of Germany and the Netherlands: Some consequences for the solidarity between generations. In: *Solidarity of Generations*, eds. H.A. Becker and P.L.J. Hermkens. Amsterdam: Thesis Publishers.
- Blossfeld, H.-P., K. Golsch and G. Rohwer (2007). *Event History Analysis with Stata*. Mahwah: Erlbaum.
- Blossfeld, H.-P. and G. Rohwer (2002). *Techniques of Event History Modeling*. Mahwah: Erlbaum.
- Blossfeld, H.-P., E. Klijzing, K. Pohl and G. Rohwer (1996). Die Modellierung interdependenter Prozesse in der demographischen Forschung: Konzepte, Methoden und Anwendung auf nichteheliche Lebensgemeinschaften. *Zeitschrift für Bevölkerungswissenschaft* 22: 29–56.
- Blossfeld, H.-P. and G. Prein (eds.) (1998). *Rational Choice Theory and Large-scale Data Analysis*. Boulder: Westview Press.
- Blossfeld, H.-P., E. Klijzing, K. Pohl and G. Rohwer (1999). Why do cohabiting couples marry? An example of a causal event history approach to interdependent systems. *Quality and Quantity* 33(3): 229–42.
- Blossfeld, H.-P. and S. Drobnič (eds.) (2001). *Careers of Couples in Contemporary Societies*. Oxford: Oxford University Press.
- Courgeau, D. and E. Lelièvre (1992). *Event History Analysis in Demography*. Oxford: Clarendon Press.
- Cox, D.R. (1990). Role of models in statistical analysis. *Statistical Science* 5: 169–174.
- Cox, D.R. (1992). Causality: Some statistical aspects. *Journal of the Royal Statistical Society* 155 (Series A): 291–301.
- Duncan, O.D. (1966). Path analysis: Sociological examples. *American Journal of Sociology* 72: 1–16.
- Eells, E. (1991). *Probabilistic Causality*. Cambridge: Cambridge University Press.
- Elder, G.H. Jr. (1987). Families and lives: Some developments in life-course studies. *Journal of Family History* 12(1–3): 179–199.
- Elder, G.H. Jr., M. Kirkpatrick Johnson and R. Crosnoe (2004). The emergence and development of life course theory. In: *Handbook of the Life Course*, eds. J.T. Mortimer and M.J. Shanahan. New York: Springer.
- Elster, J. (1979). *Ulysses and the Sirens*. Cambridge: Cambridge University Press.
- Fisher, R.A. (1935). *The Design of Experiments*. Edinburgh: Oliver & Boyd.
- Freedman, R.A. (1991). Statistical analysis and shoe leather. *Sociological Methodology* 21: 291–313.
- Goldthorpe, J.H. (1998). The quantitative analysis of large-scale data-sets and rational action theory: For a sociological alliance. In: *Rational Choice Theory and Large-scale Data Analysis*, eds. H.-P. Blossfeld and G. Prein. Boulder: Westview Press.
- Goldthorpe, J.H. (2000). *On Sociology*. Oxford: Oxford University Press.
- Goldthorpe, J.H. (2001). Causation, statistics, and sociology. *European Sociological Review* 17: 1–20.
- Government of Latvia (1999). National report submitted by the government of Latvia. *Population in Europe and North America on the Eve of the Millennium*. Geneva: UN-ECE: 123–129.
- Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross-special methods. *Econometrica* 37: 424–438.
- Heckman, J.J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica* 73: 669–738.

- Hedström, P. (1995). Rational choice and social structure: On rational-choice theorizing in sociology. In: *Social Theory and Human Agency*, ed. B. Wittrock. London: Sage.
- Hedström, P. (2005). *Dissecting the Social. On the Principles of Analytical Sociology*. Cambridge: Cambridge University Press.
- Holland, P.W. (1986). Statistics and causal inference. *Journal of the American Statistical Association* 81: 945–960.
- Holland, P.W. (1988). Causal inference, path analysis, and recursive structural equation models. *Sociological Methodology* 18: 449–484.
- Johnston, J. (1972). *Econometric Methods*, 2nd ed. New York: McGraw-Hill.
- Kalbfleisch, J.D. and R.L. Prentice (1980). *The Statistical Analysis of Failure Data*. New York: John Wiley & Sons.
- Katus, K. (1992). Fertility transition in Estonia, Latvia and Lithuania. In: *Demographic Trends and Patterns in the Soviet Union Before 1991*, eds. W. Lutz, S. Scherbov and A. Volkov. London: Routledge.
- Kelly, J.R. and J.E. McGrath (1988). *On Time and Method*. Newbury Park: Sage.
- Kerlinger, F.N. and E. Pedhazér (1973). *Multiple Regression in Behavioral Sciences*. New York: Holt, Rinehart and Winston.
- Liebertson, S. (1985). *Making it Count. The Improvement of Social Research and Theory*. Berkeley: University of California Press.
- Lillard, L.A. and L.J. Waite (1993). A joint model of marital childbearing and marital disruption. *Demography* 30: 653–681.
- Macy, M.W. (1991). Chains of co-operation: threshold effects in collective action. *American Sociological Review* 56: 730–747.
- Macy, M.W. and R. Willer (2002). From factors to actors: computational sociology and agent-based modeling. *Annual Review of Sociology* 28: 143–66.
- Manting, D. (1994). *Dynamics in Marriage and Cohabitation*. Amsterdam: Thesis Publishers.
- Marini, M.M. and B. Singer (1988). Causality in the social sciences. In: *Sociological Methodology*, ed. C.C. Clogg. Washington, D.C.: American Sociological Association.
- Mayer, K.U. and J. Huinink (1990). Age, period, and cohort in the study of the life course: A comparison of classical A-P-C-analysis with event history analysis or farewell to Lexis? In: *Data Quality in Longitudinal Research*, eds. D. Magnusson and L.R. Bergmann. Cambridge: Cambridge University Press.
- Mills, M. (2000). *The Transformation of Partnerships. Canada, the Netherlands and the Russian Federation in the Age of Modernity*. Amsterdam: Thela Thesis Population Studies Series.
- Mills, M. and F. Trovato (2001). The effect of pregnancy in cohabiting unions on marriage in Canada, the Netherlands, and Latvia. *Statistical Journal of the United Nations ECE* 18: 103–118.
- Pötter, U. and H.-P. Blossfeld (2001). Causal inference from series of events. *European Sociological Review* 17(1): 21–32.
- Stinchcombe, A.L. (1968). *Constructing Social Theories*. New York: Harcourt, Brace, and World.
- Rubin, D.B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66: 688–701.
- Rubin, D.B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics* 6: 34–58.
- Rubin, D.B. (1980). Discussion of ‘randomization analysis of experimental data in the Fisher randomization test’ by Basu. *Journal of the American Statistical Association* 81: 961–962.
- Schneider, B., M. Carnoy, J. Kilpatrick, W.H. Schmidt and R.J. Shavelson (2007). *Estimating Causal Effects. Using Experimental and Observational Designs*. Washington, D.C.: American Educational Research Association.
- Steele, F. (2003). A multilevel mixture model for event history data with long-term survivors: An application to an analysis of contraceptive sterilisation in Bangladesh. *Lifetime Data Analysis* 9: 155–174.
- Tuma, N.B. and M.T. Hannan (1984). *Social Dynamics. Models and Methods*. Orlando, FL: Academic Press.

- Willekens, F.J. (1991). Understanding the interdependence between parallel careers. In: *Female Labour Market Behaviour and Fertility*, eds. J.J. Siegers, J. de Jong-Gierveld and E. van Imhoff. Berlin: Springer-Verlag.
- Shadish, W.R., T.D. Cook and D.T. Campbell (2002). *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.
- Yamaguchi, K. (1986). Alternative approaches to unobserved heterogeneity in the analysis of repeatable events. In: *Sociological Methodology*, ed. N. Brandon Tuma. Washington, D.C.: American Sociological Association.
- Yamaguchi, K. (1991). *Event History Analysis*. Newbury Park: Sage.
- Zhang, W. and F. Steele (2004). A semiparametric multilevel survival model. *Journal of the Royal Statistical Society* 53 (Series C): 387–404.

Chapter 6

Instrumental Variable Estimation for Duration Data

Govert E. Bijwaard

6.1 Introduction

Social scientists have a long tradition of exploring the substantive implications of endogeneity in both methodological work and empirical work. Endogeneity is troublesome because it precludes the usual causal kinds of statements social scientists like to make. A canonical example is the evaluation of the effect of training programs of unemployment individuals on earnings and employment status. In general, the indicator for those who were trained is endogenous, because those individuals who choose to get training perceive the training as beneficial for earning or employment status. Other examples include the effect of union status and child-bearing on labor market outcomes. All these problems have a treatment-control flavor. The notion that treatment status is endogenous reflects the fact that simple comparisons of treated and untreated individuals are unlikely to have a causal interpretation.

In recent years, social experiments have gained popularity as a method for evaluating social and labor market programs (see e.g. Meyer 1995; Heckman et al. 1999; Angrist and Krueger 1999). In experiments the assignment of individuals to the treatment can be manipulated. If assignment is random, the average impact of the treatment can be estimated. However, a randomized assignment may be compromised, if the individuals can refuse to participate, either by dropping out, if they are to receive the treatment, or by obtaining the treatment, if they are in the control group. If this non-compliance to the assigned treatment is correlated with the outcomes in the treatment or control regimes, the observed effect of the treatment is a biased estimate of the treatment effect. Thus, even with random assignment the actual treatment status can be endogenous.

Most of the evaluation literature has focused on static treatments, i.e. treatment that is administered at a particular point in time or in a particular time interval. If the outcome is a duration the treatment or its effect can be dynamic,

G.E. Bijwaard (✉)
Netherlands Interdisciplinary Demographic Institute (NIDI), P.O.Box 11650, The Hague,
The Netherlands
e-mail: bijwaard@nidi.nl

i.e. it can be switched on and off over time. Examples are the unemployment insurance experiments (see Meyer (1995) for a survey) in which the unemployed receive a cash bonus if they find a job in a specified period. Another example is a temporary cut in unemployment benefits of unemployed individuals who do not expend sufficient effort to find a job (e.g. see Van den Berg (2004) for The Netherlands, Lalive et al. (2005) for Switzerland and Ashenfelter et al. (2005) for the U.S.).

The problem of endogeneity in duration models is similar to other statistical models: when endogeneity is present the standard interpretations given by any statistical model generally do not hold. If the training is perceived beneficial those individuals who choose to get training differ *ex ante* from those who choose not to get training. Similarly, unemployed who choose to be eligible for a cash bonus if they find a job in time, differ both in observed and unobserved characteristics that may influence their job finding probability. For linear models the problem of endogeneity can be solved if an instrument is available. The only requirement is that such an instrument affects the endogenous variable but is not correlated with the errors of the regression. We extend that notion to duration models that are inherently non-linear and propose an estimation technique.

In this article we assume the durations follow a Generalized Accelerated Failure Time (GAFT) model, a model introduced by Ridder (1990). The GAFT model is based on transforming the duration and assuming some distribution for this transformed duration. The transformation is related to the integrated hazard of a PH model. The AFT model is obtained by restricting the transformation. The AFT does not restrict the distribution of the transformed duration, while the MPH model restricts this distribution to a mixture of exponentials. The regression coefficients in a GAFT model can be interpreted in terms of the effect of regressing on the quantiles of the distribution of the transformed duration for the reference individual. In an AFT model the relation between the quantile of a individual with observed characteristics X and the quantile of the reference individual is the acceleration factor. In a GAFT this acceleration factor is multiplied by the ratio of the 'duration dependence' at the two quantile durations.

The basis of the proposed Instrumental Variable Linear Rank estimator (IVLR) is that for the true GAFT model the instrument is independent of the transformed duration. The intuition behind this idea can be clarified by considering the simple example of a re-employment experiment with random assignment to treatment and a selective compliance. Assume that both the assignment and the compliance decision are made at the start of the study. If the treatment has no impact on the re-employment hazard, then the probability of observing an individual from the treatment group among those still unemployed at a given unemployment duration should be equal to the treatment assignment probability at the start. However, if the treatment has a positive effect on the hazard the probability of observing an individual from the treatment group among those still unemployed declines with the duration, because the treated individuals find a job faster. A GAFT model transforms the duration and for the true transformed durations the hazard of these transformed durations does not depend on the treatment group. This implies that the proportion

of people in the treatment group on the *transformed duration time* remains the same, and is equal to the treatment assignment probability.

The IVLR estimation method uses the inverse of the rank test to obtain the parameters of the GAFT model, including the effect of the endogenous variable. The rank test is a commonly applied method to test the significance of a covariate on the hazard. The test is based on (possibly weighted) comparisons of the estimated non-parametric hazard rates. It is also equivalent to the score test for significance of a (vector of) coefficient(s) that arises from the Cox partial likelihood. The test rejects the influence of the covariate(s) on the hazard when it is 'close' to zero. Tsiatis (1990) shows that the inverse of the rank test can be used as an estimation equation for AFT models. The inverse of the rank test is the value of the (vector of) coefficient(s) that makes the rank-test equal to zero. Here we extend the inverse rank estimation to a GAFT model, which also includes the parameters of the transformation.

A common feature of duration data is that the durations are (right)-censored, in the sense that we only know that their realisation exceeds the censoring time. The existence of endogenous covariates implies (possible) dependence between the transformed duration and the censoring time. This implies that the IVLR estimator, which exploits the independence between the transformed durations and the instruments, may give biased results. We can often make the assumption that the (potential) censoring time is known at the start of the study. In the re-employment bonus data, for example, we can only observe the unemployed while receiving UI benefits. In this case the potential censoring time for all individuals is at 26 weeks, the maximum duration of UI benefits in Illinois at the time of the experiment. With known (potential) censoring time we can modify the GAFT transformation by introducing additional censoring such that this modified transformation and the instruments become independent for the uncensored observations. Then, the IVLR estimator on this modified transformation leads to consistent estimators.

The IVLR estimation is based on a vector of mean restrictions on weight functions of the covariates, instrument and the transformed durations. Thus the IVLR is also related to GMM estimation. In GMM estimation it is feasible to get the most efficient GMM estimator in just two steps. In the first step directly observed weighting matrices lead to a consistent, but not necessary efficient estimator. From this consistent estimator we can consistently estimate the efficient weighting matrices. It is then possible to obtain an efficient estimate of the parameters involved in just one additional step. A similar reasoning applies to the IVLR-estimator. In the first step we use simple weighting functions to obtain consistent estimates of the parameters of the GAFT model. From these parameters we can estimate the distribution of the transformed durations, which are needed to calculate the most efficient weighting functions. Then, in just one additional step the efficient IVLR is obtained.

For our empirical application we use data from the Illinois unemployment bonus experiment. These data have been analysed before with increasing sophistication by Woodbury and Spiegelman (1987), Meyer (1996) and Bijwaard and Ridder (2005). In this experiment a group of individuals who became unemployed during four months in 1984 were divided at random in three groups of about equal size: two

bonus groups and a control group. The unemployed in the claimant bonus group qualified for a cash bonus if they found a job within 11 weeks and would hold this job for at least four months. In the employer bonus group, the bonus was paid to their employer. The members of the two bonus groups were asked whether they were prepared to participate in the experiment. About 15% of the claimant bonus and 35% of the employer bonus groups refused participation. It is very likely that the decision to be eligible for a bonus is related to the unemployment duration. This makes the participation indicator an endogenous variable in relation to the unemployment duration.

The outline of the article is as follows. Section 6.2 discusses the problems associated with endogenous variables in duration models. We introduce the GAFT model and discuss the interpretation of the parameters of a GAFT model. We also give the intuition for the idea that transforming of the durations, inherent in the GAFT model, provide the basis for estimating the effect of endogenous covariates. In Section 6.3 we introduce the IVLR estimator, derive its asymptotic properties and discuss the efficiency and the practical implementation of the estimator. Section 6.4 discusses the empirical application of the IVLR estimator to the re-employment bonus experiment. We conclude with a summary and discuss possible avenues for further research in Section 6.5.

6.2 Endogenous Covariates in Duration Models

For many economic and demographic phenomena the timing of a transition from one state into another state is important. Examples include the time till re-employment of an unemployed individual, the time till marriage and the time till death. Two important features of such transition data are that relevant characteristics of the individual may change over time and that, due to a limited observation window, we do not observe the completed duration for all individuals. In a duration model the timing of a particular event is modeled and it is straightforward to incorporate time-varying variables and allow for (right)-censoring.

The key variables in duration analysis are the duration till the next event, T , and the indicator of censoring, δ . The observed durations may be right-censored, i.e. we observe $\tilde{T} = \min(T, C)$ with C the censoring time. The possible time-varying covariates are given by the vector $X_i(t)$ where i refers to a member of the population. The path of the covariates are predetermined. Thus $\bar{X}(t) = \{X(s); 0 \leq s \leq t\}$ does not depend on future events.

Two competing approaches for the analysis of duration data has been the (Mixed) Proportional Hazard (MPH) model and the Accelerated Failure Time (AFT) model. The MPH model assumes that the covariates and the unobserved heterogeneity affect the baseline hazard proportionally (see Van den Berg (2001) for a recent overview). The AFT model assumes that the covariates affect the duration proportionally. An AFT model implies that the distribution of the duration of an individual with covariate vector X and the transformed duration distribution of $e^{-\beta'X}T$ are the

same (see a.o. Brännäs 1992, Kalbfleisch and Prentice 2002). Thus the covariate accelerates the duration, when the coefficient β is smaller than zero, or decelerates the duration, when the coefficient is greater than zero. This is equivalent to a linear regression model for the log-duration.

6.2.1 The Generalized Accelerated Failure Time Model

A class of duration models that generalizes the AFT models in such a way that it also includes the MPH models is the Generalized Accelerated Failure Time (GAFT) model. The GAFT model, introduced by Ridder (1990), is not specified by the distribution of the log-duration. Instead, we transform the duration, and assume that this transformed duration has some distribution, either known or unknown. The transformation of the duration is related to the integrated hazard in a PH-model. The GAFT model is also related to the generalized regression model proposed by Han (1987).

The GAFT model assumes that the relation between the duration T and the covariates is specified as

$$\int_0^T \lambda(s; \alpha) e^{\beta'X(s)} ds = U \quad (6.1)$$

where $\lambda(t; \alpha)$ is a non-negative ‘baseline’ function on $[0, \infty)$. In the sequel we assume that λ is the piecewise constant function, i.e.

$$\lambda(t, \alpha) = \sum_{j=0}^J e^{\alpha_j} I(t_j < t \leq t_{j+1}) \quad (6.2)$$

with $t_0 = 0$ and $t_{J+1} = \infty$ and the hazard on the last interval is normalized to 1, thus $\alpha_J = 0$. Other λ -functions are also possible. The non-negative regression function $e^{\beta'X(s)}$ captures the effect of the covariates.

The GAFT model is characterized by these baseline and regression functions and by the distribution of the non-negative random variable U . We denote the survivor function of U_0 , the transformation in the true population parameters α_0 and β_0 , by $\bar{G}_0(u)$ and its hazard function by $\kappa_0(u)$. We assume that the distribution of U_0 is absolutely continuous and independent of X . The semi-parametric estimators considered in this article avoid assumptions on the distribution of U_0 .¹

As mentioned, the GAFT model contains as special cases the AFT, the PH and the MPH models. The AFT model restricts the transformation to $\lambda(t; \alpha) \equiv 1$, but leaves the distribution of U_0 unrestricted (with the exception of that U_0 should be non-negative, see e.g. Cox and Oakes (1984)). The (M)PH model restricts the distribution of U_0 , but leaves the λ unrestricted (non-negative). The distribution of U_0

¹ In appendix A we show when the parameters of GAFT model are identified.

is an unit exponential distribution (PH) or a mixture of exponential distributions (MPH).

We can interpret the GAFT model in terms of the effect of regressing on baseline quantiles, the quantiles for the reference individual. To illustrate this let $t_q(\bar{X})$ be the q -th quantile of the distribution of duration with covariate history \bar{X} . Let t_q be the q -th quantile for the reference individual (i.e. with $X(t)$ identically equal to zero). Then the ratio of the change in quantiles is

$$\frac{d t_q(\bar{X})}{d t_q} = e^{-\beta'_0 X(t_q(\bar{X}))} \frac{\lambda(t_q; \alpha_0)}{\lambda(t_q(\bar{X}); \alpha_0)} \quad (6.3)$$

In an AFT model the ratio of the quantiles is the acceleration factor $e^{-\beta'_0 X}$. Thus, in the GAFT model the ratio of the change in the quantiles is the acceleration factor multiplied by the ratio of the values of the baseline $\lambda(t)$ evaluated at the q -th quantile of the reference duration and the q -th quantile of the duration with covariate X .

In the MPH model we can interpret $\lambda(t)$ as the baseline hazard, i.e. the factor in the proportional hazard that captures the (duration) time variation in the hazard function. Thus, in the MPH model the ratio in (6.3) can be interpreted as the ratio of baseline hazards and the regression parameter, β , is the proportional change in the hazard rate due to a unit change in $X(t)$ given the unobserved heterogeneity.

6.2.2 Endogenous Covariates in GAFT Models

It can rarely be defended that a study on unemployment durations includes all the relevant characteristics of the individuals looking for a job. For example, consider our application of analysing the effect of a cash-bonus on the re-employment probability. Because such a bonus increases the reward of leaving unemployment it gives an incentive to search more intensively and therefore it increases the re-employment hazard. However, the search intensity of the unemployed individuals is usually not observed. Suppose that the unemployed have to choose at the start of their unemployment spell whether they want to be eligible for a bonus. If they choose to be eligible they have to fill in some forms, notify their new employer and provide a proof that they held that new job for at least four months. Thus, joining the bonus program implies some administrative duties for the unemployed and cooperation with their new employer. This might refrain some individuals from joining the bonus program. It is very likely that the unobserved motivation to return to work has an impact on both the decision to join the bonus program and the search intensity. This implies that the indicator of joining the bonus program is an endogenous variable for the analysis of the unemployed duration. Without adjusting for this (self)-selection standard duration analysis give biased results of the effect of the bonus on unemployment duration.

A way of adjusting for an endogenous variable is the conventional instrumental variable method that assumes instrument-error independence and an exclu-

sion restriction. A familiar example of an instrumental variable is the treatment assignment-indicator of a randomly assigned treatment in which the actual treatment still depends on a decision by the agents (or on decisions made by those who execute the program). For instance, long-term unemployed can be randomly assigned to a training program, but for many programs they can still decide not to join, or the training manager can decide to withhold some training from some people. Then, the assignment indicator is an instrument for the actual indicator of training received.

The method of instrumental variables (IV) is widely used in econometrics. For illustration consider the simple linear model

$$Y = \beta'X + \gamma D + \varepsilon$$

where Y is observed outcome, X is a vector of exogenous variables, D is an endogenous variable, and ε is a disturbance with mean 0. If D and ε are correlated OLS gives biased estimates of $\theta = (\beta, \gamma)$. The conventional IV method uses an instrument R that affects D but is uncorrelated with ε , like the assignment indicator in a random but compromised experiment. If we denote $Z = (X, R)$ and $\tilde{X} = (X, D)$ the IV estimator is

$$\hat{\theta}_{IV} = (Z' \tilde{X})^{-1} Z' Y$$

Complications arise if the outcome variable of interest is a duration variable, like the unemployment duration. Models for duration data are usually non-linear in the mean. Then the standard IV-methods can not be applied. An important issue in duration models is that the value of the endogenous variable may depend on information that accumulates during the evolution of the duration. The common approach to accommodate such time-varying variables is to relate them to the hazard rate. Another issue is that duration data are usually (right)-censored, due to a limited observation window. The hazard rate is invariant to censoring and is therefore the natural choice for the analysis of duration data.

In this paper we provide an instrumental variable method for duration data based on inference on the hazard rate. Let $D(t)$ be the value of the endogenous variable at duration t . The GAFT model with endogenous variables is

$$\int_0^T \lambda(s; \alpha) \exp(\beta'X(s) + \psi(s, D(s), \gamma)) ds = U = h(T, \bar{X}(T), \bar{D}(T), \theta) \quad (6.4)$$

where $\psi(t, D(t), \gamma)$ captures the effect of the endogenous variable and $\theta = (\beta', \alpha', \gamma')$ is the whole parameter vector. Without loss of generality we assume that the endogenous variable is binary and only changes at prediscrbed durations.

We also assume that the effect of the endogenous variable may change over the duration. Then a flexible functional form for the ‘treatment’ function is

$$\psi(t, D(t), \gamma) = \sum_{j=0}^J \gamma_j \cdot D_j \cdot I_j(t) \quad (6.5)$$

where $I_j(t) = I(t_j < t \leq t_{j+1})$ are interval indicators with $t_0 = 0$, $t_{J+1} = \infty$ and D_j is the value of D during I_j .

If D were exogenous, standard techniques for the analysis of survival time data could be used to estimate the γ 's. For example, we can use a Mixed Proportional Hazards model and estimate γ using (semi-parametric) Maximum Likelihood procedures, depending on the assumptions we make about the distribution of the unobserved heterogeneity, and the baseline hazard. If the model is correctly specified the MLE yields a consistent estimate. However, we will get biased estimation results for the parameters if the covariate is endogenous. The problem is that those who comply with their assigned treatment differ in observed and unobserved characteristics from those who do not comply.

Since physical randomization implies that at time zero all attributes of the two treatment groups are (in expectation) identical, a commonly used solution to this problem is to ignore the post-randomization compliance and rely on the analysis of the treatment assignment groups. This intention-to-treat (ITT) analysis replaces the actual value of the endogenous variable, D by the instrument, R in the estimation procedure. Further, if the model is correctly specified the estimated γ 's effect will correspond to the overall effect that would be realized in the whole population, under the assumption that the compliance rate and the factors influencing compliance in the sample are identical to those that would occur in the whole population.

The major drawback of the intention-to-treat analysis is that the estimated effect is a mixture of the population effect and the effect on the compliance. Hence, if the treatment effectively raises the re-employment hazard, the intention-to-treat measure of this effect will diminish as non-compliance increases. Another disadvantage is that compliance is very likely to depend on the perceived effects of the treatment. If, for example, the unemployed know that being eligible for a re-employment bonus does not stigmatize them, they will be more prone to participate. Thus, when the pattern of compliance is a function of the perceived efficacy of the treatment the estimated intention-to-treat will not represent the overall effect of the treatment had it been adopted in the whole population.

6.2.3 Intuition for Instrumental Variable Estimation

The basis of the proposed Instrumental Variable Linear Rank estimator (IVLR) is that for the true GAFT model the instrument is independent of the transformed duration. This implies that the proportion of people in the treatment group, $R = 1$ on the (true) transformed duration time remains the same, and is equal to the treatment assignment probability. Thus, for the true transformed duration $U_0 = h(T, \bar{X}(T), \bar{D}(T), \theta_0)$ we have

$$Pr(R = 1 | U_0 \geq u) = Pr(R = 1 | U_0 = 0), \quad (6.6)$$

This implies that the hazard of the *true* transformed duration is independent of the instrument. This independence only holds for the true parameters and we can therefore build an estimation procedure that exploits this conditional independence. In the next section we introduce our proposed method based on this condition independence assumption.² First we discuss the implications of right-censoring on these independence assumption.

A common feature of duration data is that some of the observations are censored. Assume the censoring time, C , is (potentially) known. For example, in the analysis of unemployment duration based on administrative data the duration is often only observed while the individual receives unemployment benefits. Usually, the maximum duration of receiving benefits is based on the labor market history of the individual and is recorded in the data. Then, the potential censoring time is known and the observed durations are $\tilde{T} = \min(T, C)$ and $\Delta = I(T \leq C)$, where Δ is one if T is observed.

One is tempted to define the censored transformed durations by the minimum of the transformed time till (potential) censoring and the transformed time till the event occurs, $\tilde{U}(\theta) = \min(h(T; \theta), h(C; \theta)) = h(\tilde{T}; \theta)$. However, the existence of endogenous covariates and censoring makes some of the orthogonality conditions fail to hold. This can be illustrated by a simple example: Consider a fixed censoring time, all individuals have the same maximum duration of receiving benefits. Then for all individuals, irrespective of their value of the endogenous variable, censoring occurs at time C . Suppose the binary endogenous variable, D , and other covariates all be determined at the start of the study and have a constant effect on the hazard. Finally, we assume that except for γ the effect of the endogenous variable, all parameters, β_0 and α_0 , are known. Then, the transformation is

$$U_0 = e^{\gamma_0 D + \beta_0' X} \Lambda_0(T) \quad (6.7)$$

with $\Lambda_0(t) = \int_0^t \lambda(s, \alpha_0) ds$. Hence, if $D = 0$ censoring in the transformed time occurs at $e^{\beta_0' X} \Lambda_0(C)$, but if $D = 1$ censoring occurs at $e^{\beta_0' X + \gamma_0} \Lambda_0(C)$. Thus, if $\gamma_0 > 0$, then all transformed durations in the interval $[e^{\beta_0' X} \Lambda_0(C), e^{\beta_0' X + \gamma_0} \Lambda_0(C)]$ have $D = 1$ (for $\gamma_0 < 0$ the boundaries are reversed). The hazard of U_0 on this interval clearly depends on D and hence on R . The independence of the hazard of U_0 and R only holds up to the lower bound of the interval. This implies that in the IVLR, which exploits this independence, the transformed durations that fall in the problematic interval have to be censored. In Appendix 2 we derive the additional

² Here we only concentrate on a static binary instrument and a discrete, but possible time-varying according to a prescribed protocol, endogenous variable. It is not difficult to extend the analysis to more, discrete, levels of both the instrument and the endogenous variable and to have a sequential instrument.

censoring required in a more general setting. This additional censoring, $C^U(\theta)$, depends on the (unknown) parameters. The IVLR estimation method is then based on the (transformed) durations $\tilde{U}(\theta) = \min(U(\theta), C^U(\theta))$, with $U(\theta)$ given in (6.4) and the censoring indicator $\Delta^U(\theta) = I(U(\theta) < C^U(\theta))$. Then for the ‘uncensored’ observations, that is for $\Delta^U(\theta) = 1$, the transformed duration $\tilde{U}(\theta)$ is independent of the instrument. This is explained in more detail in Appendix 2.

6.3 Instrumental Variable Linear Rank Estimation

In this section we introduce an Instrumental Variable method for duration models that adjusts for the possible endogeneity of the intervention, without suffering the problems of the intention-to-treat method. The basis of this IVLR estimator is that for the true GAFT model the instrument does not influence the hazard of the transformed duration. A typical way to test the significance of a covariate is the rank-test, see Prentice (1987). The test is based on (possibly weighted) comparisons of the estimated non-parametric hazard rates. It is also equivalent to the score test for significance of a (vector of) coefficient(s) that arises from the Cox partial likelihood. The test rejects the influence of the covariate(s) on the hazard when it is ‘close’ to zero. Tsiatis (1990) shows that the inverse of the rank test can be used as an estimation equation for AFT models. The inverse of the rank test is the value of the (vector of) coefficient(s) that puts the rank-test equal to zero. Here we extend the inverse rank estimation to a GAFT model, which also includes the parameters of the duration dependence.

6.3.1 The IVLR Estimator

Before we turn to the general model we discuss a simple AFT example to provide more insight into the inverse rank estimation approach. Suppose we would like to test whether a covariate X influences the hazard. If the covariate does not influence the hazard, the mean of the covariate among the survivors does not change with the survival time, i.e. $E[X | T \geq t] = E[X]$. Define the observation indicator, that is the indicator that individual j is still alive (unemployed) at time t , by $Y_j(t) = I(t_j \geq t)$. Then the rank test-statistic is (assuming no censoring)

$$\sum_i^n \left[X_i - \frac{\sum_j Y_j(t_i) X_j}{\sum_j Y_j(t_i)} \right]$$

where the second term is the mean of the covariate among those individuals still alive at t_i . Thus for each observation of the covariate we compare the observed value with its expected value among those still alive (under the hypothesis of no effect of the covariate) and sum over all observations. If this sum is significantly different from zero, we reject the null of no influence.

Now assume that the true model is an AFT-model with $U = e^{\beta X} T$. Then, for the true parameter $\beta = \beta_0$ the hazard of U does not depend on the covariate X . This implies that the rank statistic for the true parameter on the transformed U -time is zero. However the β_0 is unknown and an inverse rank estimate $\hat{\beta}$ of β_0 is the value of β for which

$$\sum_i^n \left[X_i - \frac{\sum_j Y_j^U(U_i) X_j}{\sum_j Y_j^U(U_i)} \right] = 0$$

with $U_i = e^{\hat{\beta} X_i} t_i$ and $Y_j^U(u) = I(U_j \geq u)$, the observation indicator on the (transformed) U -time. Tsiatis (1990) derives the asymptotic properties of this estimator. Robins and Tsiatis (1991) discuss how the rank estimator can be used to estimate the effect of an endogenous variable in an AFT-model.

We extend the method of Robins and Tsiatis (1991) to GAFT models. We use the transformed GAFT durations in (6.4) and adjust them for censoring, see Appendix 2. Just as in the example above, we have that for the population parameter vector θ_0 the hazard of the implied transformed duration U_0 , which is $\kappa_0(u)$, is independent of the covariate and instrument history up to $h_0^{-1}(u)$. Because this is true only for $\theta = \theta_0$, we can use the inverse of the rank statistic to get an estimate of θ_0 . Note that for notational convenience we suppress the dependence on θ in censored durations $\tilde{U}(\theta)$.

The estimating equations that defines the IVLR estimator contain a left-continuous vector weight function W . The weight function may depend on $\tilde{U}_i(\theta) = \tilde{U}_i, \bar{X}_i^U(u)$, the path of the covariate on the transformed time scale (see Appendix 2) and R . Typical examples are $W = (W_\beta, W_\gamma, W_\alpha)$ with $W_\beta = X$ for the coefficient vector β of the exogenous variables and $W_\gamma = R$, the instrument, for a dummy endogenous variable D and $W_{\alpha_j} = I_j(u) = I(h(t_j) < u \leq h(t_{j+1}))$ for a piecewise constant baseline hazard on intervals $(t_j, t_{j+1}]$. The variance of the IVLR estimator depends on the choice of the weight-function and in Section 6.3.2 we discuss the optimal choice of this function. For a given choice of the weight-function and possible additional censoring the IVLR estimator is defined by the estimating equations

$$S_n(\theta; W) = \sum_{i=1}^n \Delta_i^U \{ W(\tilde{U}_i, \bar{X}_i^U(\tilde{U}_i), R_i; \theta) - \bar{W}(\tilde{U}_i; \theta) \} \quad (6.8)$$

where

$$\bar{W}(\tilde{U}_i; \theta) = \frac{\sum_{j=1}^n Y_j^U(\tilde{U}_i) W(\tilde{U}_i, \bar{X}_j^U(\tilde{U}_i), R_j; \theta)}{\sum_{j=1}^n Y_j^U(\tilde{U}_i)},$$

is the average weight function among the individuals still at risk at the transformed duration $\tilde{U}_i(\theta)$. Note that we use Δ_i^U instead of Δ_i to assure independence of the instruments and the transformed durations for all uncensored observations.

The interpretation of the estimation equations is that it compares the value of the weight function at a transformed duration $\tilde{U}_i(\theta)$ to the average of the weight functions for those individuals that are still at risk at that particular transformed duration. For the true parameter vector $\theta_0 = (\beta_0, \alpha_0, \gamma_0)$ the expected difference of the weight function and its average for those still at risk is zero. Thus, the statistic $S_n(\theta; W)$ has mean zero for the true parameters. We therefore base our estimator on the roots of $S_n(\theta; W) = 0$, which is the inverse of the extended rank statistic. However, the estimating functions are discontinuous, piecewise constant, functions of θ and a solution may not exist. For that reason we define the Instrumental Linear Rank estimator (IVLR) $\hat{\theta}_n(W)$ as the minimizer of the quadratic form, i.e.

$$\hat{\theta}_n(W) = \inf \{ \theta \mid S_n(\theta; W)' S_n(\theta; W) \} \quad (6.9)$$

To ensure weak consistency and asymptotic normality of the IVLR estimator we make the following assumptions. The random variable R is an instrument that is determined at the start. We restrict both the instrument, R , and the endogenous variable D , to be binary. The other assumptions can be found in Appendix 3.

If $S_n(\theta; W)$ were differentiable with respect to θ , then asymptotic normality can be proved using Taylor series expansion in a neighborhood of θ_0 . Tsiatis (1990) showed that, if $S_n(\theta; W)$ is not differentiable, as in the current problem, we can still use a linear approximation of $n^{-1/2} S_n(\theta; W)$. Using this approximation and the asymptotic normality of $S_n(\theta_0; W)$, we can show that $\sqrt{n}(\hat{\theta}_n(W) - \theta_0)$ is asymptotically normal. For the derivation of the asymptotic properties we use counting process theory (see Appendix 2). Let $a(u; \theta_0)$ be the probability limit of the average weight function (see assumption 7 in appendix 3), C_0 the transformed censoring time for $\theta = \theta_0$. Let $d_{i0}(u)$ the derivative of the hazard of $U(\theta)$ w.r.t. θ , i.e.

$$d_{i0}(u) = \left. \frac{\partial \kappa_i^U(u; \theta)}{\partial \theta} \right|_{\theta=\theta_0}$$

and $V(u, \theta)$ is the probability limit of

$$\frac{1}{n} \sum_{i=1}^n [W(u, \bar{X}_i^u(u), R_i) - \bar{W}(u; \theta)] \times d_{i0}(u)' Y_i^U(u)$$

The asymptotic properties of the IVLR estimator are summarized in the following two theorems.

Theorem 1. (Consistency) *If assumptions 1 to 7 (in appendix 3) hold $\hat{\theta}_n(W)$ converges in probability to θ_0 .*

Theorem 2. (Asymptotic Normality) *If assumptions 1 to 9 (in appendix 3) hold and $Q(W)$ has full rank, then*

$$\sqrt{n}(\hat{\theta}_n(W) - \theta_0) \overset{D}{\rightarrow} N\left(0, Q^{-1}(W) \Omega(W) Q^{-1}(W)\right) \quad (6.10)$$

where

$$\Omega(W) = \int_0^{C_0} a(u; \theta_0) \kappa_0(u) du \quad (6.11)$$

is the asymptotic variance of $n^{-1/2}S_n(\theta_0; W)$ and,

$$Q(W) = \int_0^{C_0} V(u, \theta_0) du \quad (6.12)$$

the limiting covariance matrix of the processes $W(u, \bar{X}_{i0}(u), R_i)$ and $d_{i0}(u)/\kappa_0(u)$.

Proof. See Appendix 3.

6.3.2 Efficiency of the IVLR Estimator

Many different choices of the weight functions lead to consistent estimates of the parameters. By properly choosing the weight function the asymptotic variance of the IVLR can be minimized. Tsiatis (1990) has shown that for the AFT model with exogenous covariates weight functions proportional to $u\kappa'_0(u)/\kappa_0(u)X$, with $\kappa_0(u)$ is the hazard of the true transformed durations U_0 , minimize the asymptotic variance of the estimated regression parameters. In general the distribution of the true transformed duration, U_0 , is unknown. This distribution can consistently be estimated from the implied transformed durations induced by IVLR-estimation with a weight function that does not depend on the transformed durations.

The IVLR estimation is based on a vector of mean restrictions on weight functions of the covariates, the instrument and the transformed durations. GMM estimation is also based on moment conditions and in GMM estimation it is feasible to get the most efficient GMM estimator in just two steps. A similar reasoning applies to the IVLR-estimator. This justifies an adaptive construction of an efficient estimator. In the next section we address the practical implementation of an adaptive estimation procedure. First, we introduce the optimal weight function.

Theorem 3. (Optimal weight function in IVLR) *The weight-function that gives the smallest asymptotic variance for $\hat{\theta}_n(W)$ is*

$$W_{opt}(u, \bar{X}(u), R) \propto \left. \frac{\partial \ln \kappa^U(u; \theta)}{\partial \theta} \right|_{\theta=\theta_0} = \frac{d_{i0}(u)}{\kappa_0(u)} \quad (6.13)$$

The asymptotic covariance matrix of the optimal IVLR estimator reduces to

$$\Omega^{-1}(W_{opt}) = Q^{-1}(W_{opt}). \quad (6.14)$$

Proof of Theorem 3. From

$$\frac{1}{\sqrt{n}} \begin{pmatrix} S_n(\vartheta_0; W) \\ S_n(\vartheta_0; W_{opt}) \end{pmatrix} \xrightarrow{D} N \left(0, \begin{pmatrix} \Omega(W) & Q(W)' \\ Q(W) & \Omega(W_{opt}) \end{pmatrix} \right)$$

follows that the matrix

$$Z = \begin{pmatrix} \Omega(W) & Q(W)' \\ Q(W) & \Omega(W_{opt}) \end{pmatrix}$$

is non-negative definite, the same is true for its inverse. In particular, the submatrices on the main diagonal of the inverse are non-negative definite. Hence the matrix

$$Q^{-1}(W) \Omega(W) Q'^{-1}(W) - \Omega^{-1}(W_{opt})$$

is a non-negative definite matrix Q.E.D.

Consider, for example, a GAFT model with a piecewise constant λ function as defined in (6.2). Assume that the model has a constant coefficient for the endogenous variable then by (6.13) the optimal weight functions are

$$W_{opt,\beta} = X(u) \left[1 + u \frac{\kappa'_0(u)}{\kappa_0(u)} \right] \quad (6.15)$$

$$W_{opt,\alpha_j} = \left(1 + u \frac{\kappa'_0(u)}{\kappa_0(u)} \right) \cdot (R I_j^1(u) + (1 - R) I_j^0(u)) + \quad (6.16)$$

$$+ R \left[(1 + u \kappa_0(u)) \frac{f_0(u|1, R) - f_0(u)}{f_0(u)} + u \frac{f'_0(u|1, R) - f'_0(u)}{f_0(u)} \right] I_j^1(u)$$

$$+ (1 - R) \left[(1 + u \kappa_0(u)) \frac{f_0(u|0, R) - f_0(u)}{f_0(u)} + u \frac{f'_0(u|0, R) - f'_0(u)}{f_0(u)} \right] I_j^0(u)$$

$$W_{opt,\gamma} = R \left[1 + u \frac{\kappa'_0(u)}{\kappa_0(u)} \right] + \quad (6.17)$$

$$+ R \left[(1 + u \kappa_0(u)) \frac{f_0(u|1, R) - f_0(u)}{f_0(u)} + u \frac{f'_0(u|1, R) - f'_0(u)}{f_0(u)} \right]$$

where $f_0(u|D, R)$ is the density of U_0 given D and R , $f'_0(\cdot)$ is the derivative of the density and $I_j^D(u) = I(m_j(X, D) < u \leq m_{j+1}(X, D))$ for

$$m_j(X, D) = \int_0^{t_j} \lambda(s, \alpha) e^{\beta' X(s) + \gamma D} ds$$

6.3.3 Estimation in practice

The statistic $S_n(\theta; W)$ is a multi-dimensional step-function. Therefore, the standard Newton-Raphson algorithm cannot be used to solve the minimizer of the quadratic form of the estimation equations in (6.9). One of the alternative methods for finding the roots of a non-differentiable function is the Powell-method. This method (see Press et al. 1986, §10.5 and Powell 1964) is a multidimensional version of the Brent algorithm.³

An additional difficulty in solving the estimation equations is that the (optimal) weight-functions may depend on the, unknown, distribution of U_0 . However, a consistent first stage estimator based on weight-functions that are independent of the distribution of U_0 is easy to find. For example, in a GAFT model with a piecewise constant λ and a time-invariant coefficient of the endogenous variable, the choice for the first-step weight functions could be: $W = (X, R, I_1(u), \dots, I_{J-1}(u))$, with X is the weight-function for the effect of the exogenous covariates, R is the weight-function for the (time-constant) endogenous variable and, $I_1(u), \dots, I_{J-1}(u)$ are the weight-functions for the parameters of the piecewise constant baseline hazard. Then based on the first stage estimator we can calculate the optimal weight functions.⁴

Related to the computation of optimal weight function is the estimation of the variance matrix for an arbitrary weight function.⁵ The difficulty in estimating the covariance matrix lies in the calculation of the matrix $Q(W)$ and not in the calculation of the variance matrix of the estimating equation. The latter can be consistently estimated by

$$\hat{\Omega} = \frac{1}{n} \sum_{i=1}^n \Delta_i^{\hat{U}} [W(u, \bar{X}_i^{\hat{U}}(u), R_i) - \bar{W}(u, \hat{\theta})][W(u, \bar{X}_i^{\hat{U}}(u), R_i) - \bar{W}(u, \hat{\theta})]' \quad (6.18)$$

where \hat{U} is the value of $U(\hat{\theta})$.

Thus, the optimal weight functions, the covariance matrix and the most efficient estimators are estimated in two steps. The first step consists of obtaining a consistent estimate of θ_0 using a weight function that does not depend on the distribution of U_0 . The second step concerns the estimation of the unknown distribution of U_0 , based on the transformed durations implied by the first step estimates. Many different methods are available to get a reasonable estimate of an unknown distribution. We shall *not* apply the commonly used kernel based method. Although kernel-smoothed hazard rate estimators have been developed and adjusted to deal with the boundary problems inherent to hazard rates these methods can be difficult to implement due to the choice of the bandwidth. It is also unclear how the boundary corrections can

³ See the site of Bo Honore <http://www.princeton.edu/honore/> for the Powell method in Gauss.

⁴ The estimation procedures written in Gauss are available upon request from the author.

⁵ Robins and Tsiatis (1991) suggested to use a numerical derivative of $n^{-1}S_n(\theta; W)$ that does not need an estimate of the optimal W -function to get $\hat{Q}(W)$. This numerical derivative is sensitive to the choice of the difference in θ . We found it hard to get stable results.

be incorporated in the kernel estimates of the derivative of the hazard. We therefore choose to use a series approximation of the distribution.

Suppose the distribution of U_0 can be approximated arbitrary well using orthonormal polynomials. We base our approximation on Hermite polynomials using the exponential distribution as a weighting function:

$$g_0(u) = \frac{ae^{-au}}{\sum_{l=0}^L b_l^2} \left[\sum_{l=0}^L b_l L_l(u) \right]^2 \quad (6.19)$$

where

$$L_l(u) = \sum_{k=0}^l \binom{l}{k} \frac{(-au)^k}{k!} \quad (6.20)$$

are the Laguerre polynomials. The unknown parameters of this approximation are a and b_0, \dots, b_L . If $b_l \equiv 0$ for all $l > 0$ the distribution of U_0 is exponential. Even for L as small as three (6.19) allows for many different shapes of $\kappa_0(u)$ and its derivative. Both can be derived analytically given the estimates of the parameters. The parameter estimators can be obtained from standard maximum likelihood procedures on the observed transformed durations implied by the first step estimates.

If a consistent but inefficient estimator $\hat{\theta}_n(W)$ of θ_0 is available, e.g. the first stage estimator, and we have estimated the parameters of the polynomial approximation of the distribution of U_0 we can obtain an efficient estimator $\hat{\theta}_{opt}$ in just one additional step. From the linearization of the estimating equations, given in (6.36), we obtain an efficient estimator from

$$\hat{\theta}_{opt} = \hat{\theta}_n(W) - \hat{Q}(W)^{-1} S_n(\hat{\theta}_n(W); W_{opt})/n \quad (6.21)$$

This procedure is related to obtaining an efficient GMM estimator in two steps from a consistent, but possible, inefficient GMM estimator. It also possible to obtain the efficient estimator directly from minimizing the quadratic form. However, this involves again the minimization of a multi-dimensional step function.

6.4 Application to the Illinois Re-employment Bonus Experiment

Between mid-1984 and mid-1985, the Illinois Department of Employment Security conducted a controlled social experiment.⁶ This experiment provides the opportunity to explore, within a controlled experimental setting, whether bonuses paid

⁶ A complete description of the experiment and a summary of its results can be found in Woodbury and Spiegelman (1987).

Table 6.1 Average unemployment durations: control group and (non-)compliers

	Control group	Claimant bonus			Employer bonus		
		All	Compl.	Non-compl.	All	Compl.	Non-compl.
Benefit weeks	18.33	16.96	16.74	18.18	17.65	17.62	17.72
	(0.20)	(0.20)	(0.22)	(0.50)	(0.21)	(0.26)	(0.35)
N	3952	4186	3527	659	3963	2586	1377

Note: Standard error of average in brackets.

to Unemployment Insurance (UI) beneficiaries or their employers reduce the time spend in unemployment relative to a randomly selected control group. In the experiment, newly unemployed claimants were randomly divided into three groups: a *Claimant Bonus Group*, a *Employer Bonus Group* and, a *control group*. The members of both bonus groups were instructed that they (Claimant group) or their employer (Employer group) would qualify for a cash bonus of \$500 if they found a job (of at least 30 hours) within 11 weeks and, if they held that job for at least four months. Each newly unemployed individual who was randomly assigned to one of the two bonus groups had the possibility to refuse participation in the experiment.

Woodbury and Spiegelman (1987) concluded from a direct comparison of the control group and the two bonus groups that the claimant bonus group had a significantly smaller average unemployment duration. The average unemployment duration was also smaller for the employer bonus group, but the difference was not significantly different from zero. These results are confirmed in Table 6.1. Note that the response variable is insured weeks of unemployment. Because UI benefits end after 26 weeks, all unemployment durations are censored at 26 weeks. In Table 6.1 no allowance is made for censoring. In the table we distinguish between compliers, those who agreed to be eligible for a bonus if assigned to a bonus group, and non-compliers. We see that the claimant bonus only affects the compliers and that the average unemployment duration of the non-compliers and the control group are almost equal.

About 15% of Claimant group and 35% of the employer group declined participation. The reason for this refusal is unknown. Bijwaard and Ridder (2005) showed that the participation rate is significantly related to some observed characteristics of the individuals that also influence that re-employment hazard. Hence, we cannot exclude the possibility of unmeasured variables that affect both the compliance decision and the re-employment hazard. Meyer (1996) analyzed the same data using a PH model with a piecewise constant baseline hazard. He used the randomization indicator instead of the actual bonus-group agreement indicator as an explanatory variable. Thus he used the ITT estimator. He found a significantly positive effect of the claimant bonus. However, as shown by Bijwaard and Ridder (2005), the ITT may have a downward bias.

We calculate the IVLR estimate of the effect of the claimant and employer bonus on the unemployment duration in a GAFT model and compare these estimates with the IVLR estimates of an AFT model, with ITT estimates in an MPH model and

the ML estimates of an MPH model that ignores the endogeneity of the decision to participate in the bonus group. We consider the two interventions separately: thus Claimant Bonus group versus Control group and Employer Bonus group versus Control.

We shall consider two alternative specifications for the effect of the bonus on unemployment duration: (i) constant effect and, (ii) a change in the effect after 10 weeks, in line with the end of the eligibility period of the bonuses. Thus, the implied transformed durations are

$$U(\theta) = \int_0^T \lambda(s; \alpha) e^{\beta'X + (\gamma_1 I_1(s) + \gamma_2 I_2(s))D} ds \quad (6.22)$$

with $I_1(t) = I(0 \leq t < 11)$ and $I_2(t)$ is its complement. Note that the covariates are all time-constant because the individual characteristics available in the data are all determined when the individuals register at the unemployment office. We include the following: the logarithm of the age (LNAGE), the logarithm of the pre-unemployment earnings (LNBPE), gender (MALE = 1), ethnicity (BLACK = 1), and the logarithm of the weekly amount of UI benefits plus dependence allowance (LNBEN). We employ two different specifications for $\lambda(t; \alpha_0)$: (i) AFT model, i.e. $\lambda(t; \alpha_0) \equiv 1$; and (ii) GAFT model with a piecewise constant λ on six intervals 0–2, 2–4, 4–6, 6–10, 10–25 and 25 and beyond.

For identification we need to set one of the parameters of the piecewise constant λ equal to one (or the log equal to zero). We let the base interval, the interval on which $\lambda = 1$, start on the last week before the end of the observation period, at 25 weeks. This allows us to capture the spike in the observed unemployment duration just before the UI eligibility period ends. The end of the UI eligibility period, at 26 weeks, is for all individuals the same and thus provides the potential censoring time.

For both the AFT and the GAFT specifications we estimate a first stage IVLR using the Powell-method and the one step optimal IVLR. The first stage IVLR uses the values of the covariates, X , the interval indicators on the transformed duration (only for the GAFT-model), $I_j(u)$ and, the bonus group assignment indicator times the interval indicators on the transformed duration, $R \cdot I_1(u)$ and $R \cdot I_2(u)$, as the weight functions. From these first stage IVLR's the implied transformed duration are obtained. Then, we estimate the parameters of the polynomial approximation of the distribution of U conditional on R and D as mentioned in Section 6.3.3. From these estimated parameters we calculate the hazard and its derivative of the transformed duration. These functions are then used as inputs to derive the optimal weight functions (see Theorem 3), which in turn are necessary to calculate the covariance matrix. We also calculate the 1-step efficient estimates with these optimal weight functions. In the case of a constant bonus effect, the optimal weight function are given in (6.15), (6.16) and (6.17). When we assume that the effect of the bonus changes after 11 weeks the optimal weight function in (6.17) is more complicated and therefore not spelled out here.

The estimation results for the bonus effects are reported in Table 6.2. The results for the piecewise constant λ and for the regression coefficients in the AFT and

Table 6.2 Instrumental Variable Linear Rank estimates for the effect of the Bonus

	Claimant group			
	AFT	GAFT ^a	MLE	ITT
<i>Constant effect</i>				
First stage	0.1446 (0.0493)	0.1024 (0.0523)	–	–
1-step optimal	0.1596 (0.0460)	0.0932 (0.0380)	0.1039 (0.0285)	0.1117 (0.0303)
<i>Time varying effect</i>				
First stage				
0–10	0.2955 (0.0523)	0.1433 (0.0907)	–	–
10+	–0.0720 (0.0608)	0.0063 (0.0886)	–	–
1-step optimal				
0–10	0.3865 (0.0486)	0.1439 (0.0578)	0.1601 (0.0361)	0.1516 (0.0378)
10+	–0.0437 (0.0572)	–0.0411 (0.0850)	–	–
Employer group				
First stage	0.1011 (0.0646)	0.0721 (0.0470)	–	–
1-step optimal	0.1332 (0.0612)	0.0696 (0.0425)	0.0387 (0.0318)	0.0516 (0.0307)
<i>Time varying effect</i>				
First stage				
0–10	0.2304 (0.0710)	0.1103 (0.0736)	–	–
10+	–0.0783 (0.0836)	–0.0048 (0.1253)	–	–
1-step optimal				
0–10	0.6334 (0.0674)	0.1279 (0.0521)	0.0881 (0.0402)	0.0800 (0.0348)
10+	0.0330 (0.0745)	–0.0747 (0.0882)	–	–

^a GAFT piecewise constant intervals: 0–2, 2–4, 4–6, 6–10, 10–25, 25 →;

Note: Standard error in brackets.

GAFT models can be found in Appendix 4. A comparison of the results shows that AFT overestimates the effect and that both ML and ITT estimators underestimate the effect of the employer bonus. For the claimant bonus the ML and ITT estimates are very close to the IVLR estimates. This indicates that endogeneity of the compliance decision is rather limited for the claimant group. The compliance rate in the claimant group is much higher and most probably the compliance decision of the individuals in the claimant bonus group is less related to their expected unemployment duration. The results clearly indicate that the bonuses only influence the chances to find a job in the first ten weeks. This is in line with the bonus eligibility period: those who find

a job after that period would not get the bonus. The effect of the Claimant Bonus increases from about 10% higher probability to find a job at every unemployment duration to about 15% higher probability to find a job in the first ten weeks (and no effect thereafter). The bonus for the Employer group raises the job finding probability with about 7% at every unemployment duration or with about 12% in the first ten weeks of unemployment.

In the GAFT (and AFT) model the effect of the bonus is defined in terms of the change in the quantiles, see (6.3). In an AFT model with a time-constant coefficient for the bonus this effect is constant and independent of the other covariates. In a GAFT model the λ function influences this effect directly and indirectly as the other covariates determine the quantiles. Using the distribution of U_0 , already calculated to estimate the optimal IVLR and the variance-covariance matrix, we can derive the effect of the bonus in the GAFT depending on the quantile of the distribution. In Table 6.3 we present the effect of the bonus on the unemployment duration at the 80%, 60% and 40% survival for the reference individual and for a black individual, together with the AFT effect (first stage). Figures 6.1, 6.2, 6.3, and 6.4 depict the change over the whole 90%–25% survival range of the effect of the bonus in the GAFT model.

Note that an effect smaller than one indicates that the bonus decreases the duration till re-employment and an effect bigger than one increases the duration. We

Table 6.3 Effect of the Bonus on the length of unemployment duration

		Claimant		Employer	
		<i>Constant</i>	<i>Time-varying</i>	<i>Constant</i>	<i>Time-varying</i>
AFT	0–10	0.865	0.744	0.904	0.794
	10+	0.865	1.075	0.904	1.081
GAFT		<i>Reference individual</i>			
80%	$t_q(0)$	3.9	3.7	2.8	4.3
	$t_q(1)$	3.5	2.9	2.5	3.7
	effect	0.911	0.866	0.933	0.823
60%	$t_q(0)$	12.8	12.6	8.9	12.7
	$t_q(1)$	10.4	9.4	7.8	10.0
	effect	0.911	0.571	0.933	1.078
40%	$t_q(0)$	25.7	25.7	20.7	24.3
	$t_q(1)$	22.8	23.1	18.3	22.5
	effect	1.772	1.973	0.933	1.078
GAFT		<i>Black individual</i>			
80%	$t_q(0)$	7.5	6.8	4.8	8.1
	$t_q(1)$	6.5	5.3	4.1	6.4
	effect	0.911	0.681	0.933	0.880
60%	$t_q(0)$	25.3	24.4	18.44	24.22
	$t_q(1)$	22.1	21.0	16.2	22.5
	effect	1.772	1.042	0.933	1.078
40%	$t_q(0)$	35.6	35.1	30.7	34.2
	$t_q(1)$	32.9	33.8	28.9	33.9
	effect	0.911	1.042	0.933	1.078

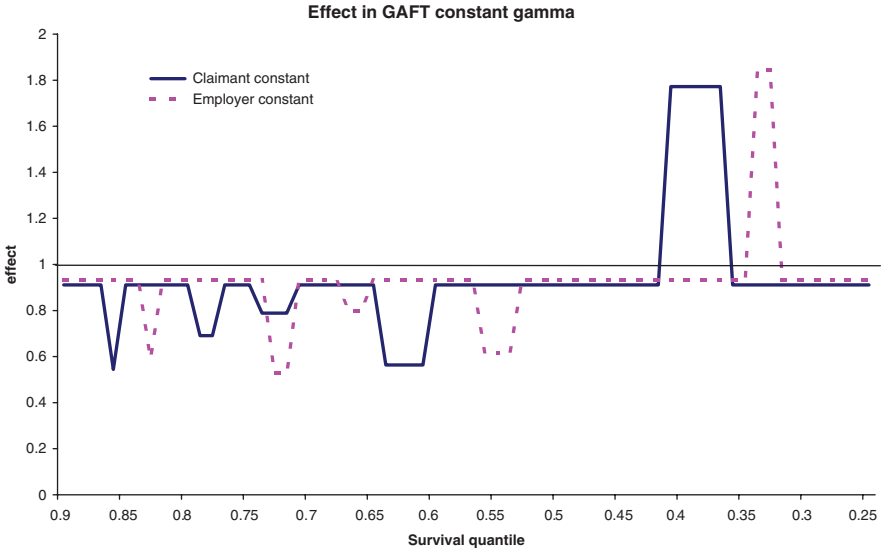


Fig. 6.1 Effect of Bonus on quantiles of unemployment duration $\frac{dt_q(1)}{dt_q(0)}$ (constant γ)

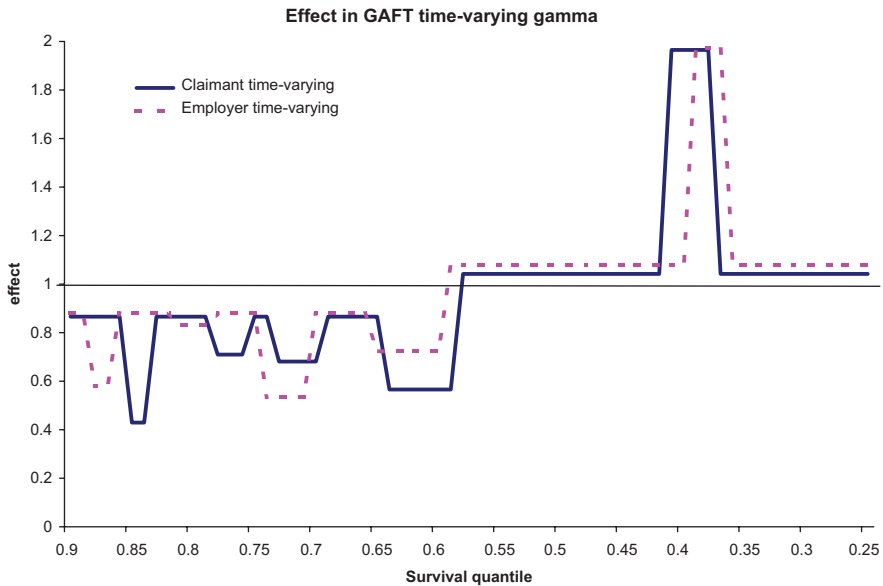


Fig. 6.2 Effect of Bonus on quantiles of unemployment duration, $\frac{dt_q(1)}{dt_q(0)}$ (time-varying γ)

see from the table (and more pronounced in Figs. 6.1 and 6.3) that even for a time-constant γ the effect of the bonus on the unemployment duration in the GAFT model changes with the duration. The huge spike in the effect at the survival quantile of 40% for the claimant group is because the re-employment rate exhibits a spike

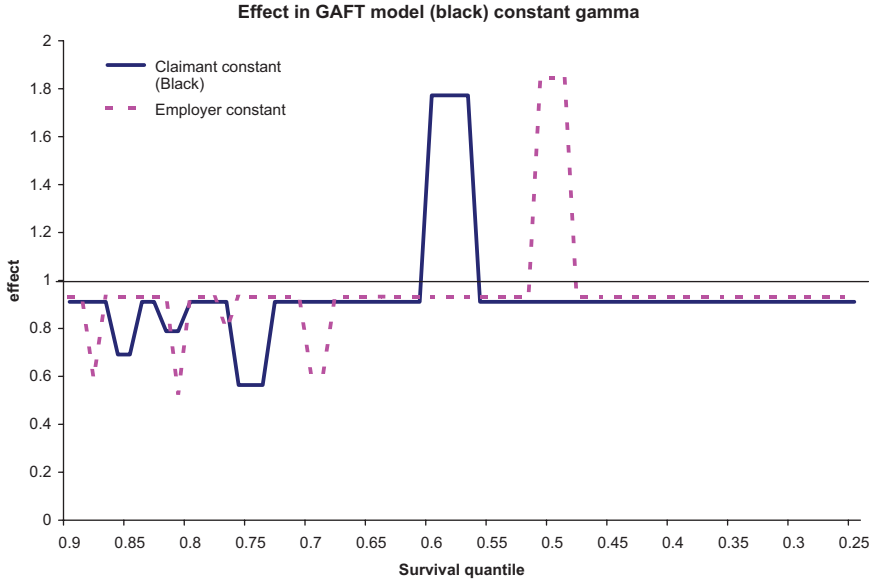


Fig. 6.3 Effect of Bonus on quantiles of unemployment duration of BLACKS, $\frac{dT_q(1)}{dT_q(0)}$ (constant γ)

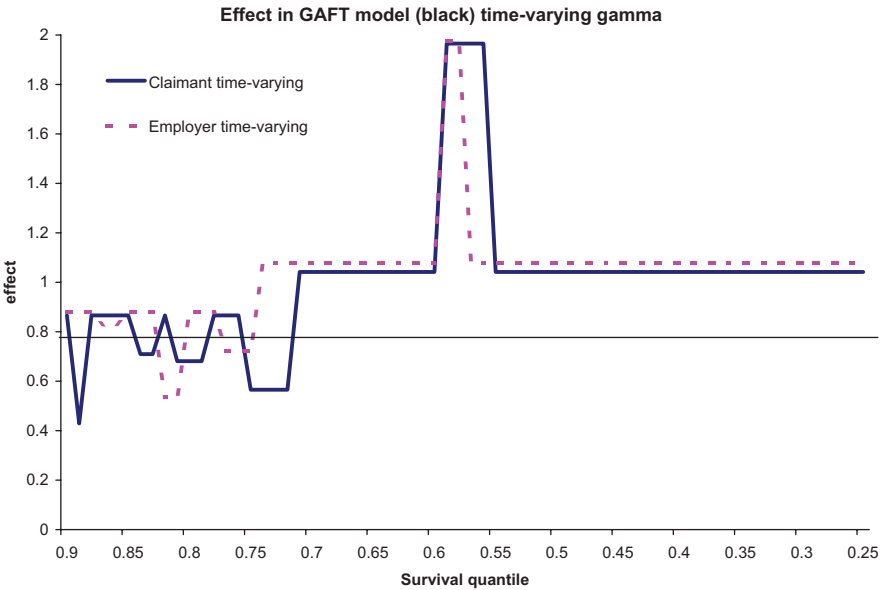


Fig. 6.4 Effect of Bonus on quantiles of unemployment duration of BLACKS, $\frac{dT_q(1)}{dT_q(0)}$ (time-varying γ)

just before the time that unemployment benefits are exhausted, which is at 26 weeks. For the individuals in the control group the 40% survival time is just before 26 weeks, while in the claimant bonus group it is at 23 weeks. Thus the control group individuals are in the re-employment spike while the claimant bonus group are not. The interval boundaries of the other intervals of λ also cause, although not as pronounced, spikes. These spikes are downward because the λ is jumping to a lower level at these boundaries. The spikes are also visible in the effect of a time-varying coefficient of the bonus, see Figs. 6.2 and 6.4. Here, the change in γ at a duration of 10 weeks, after which the coefficient is negative, is reflected is a upward shift of the effect curve.

An indication that the AFT is not the right model is the difference between the first stage and one-step optimal estimators for the AFT model. For a correctly specified model both estimators are consistent and, therefore, do not differ much. In the GAFT model the first stage and one-step estimator are of the same magnitude. The estimated standard errors of the latter are, as expected, substantially lower in most situations.

Although the focus in this article is on the estimation of the effect of a possibly endogenous variable on the duration we also give a short discussion on the estimation results of the other parameters. These estimators can be found in the Tables 6.4, 6.5, 6.6, and 6.7 in Appendix 4. The regression parameters are overestimated (in absolute terms) if we assume an AFT model. These regression parameters hardly change from a model with constant bonus effect (Table 6.5) to a model with time-varying bonus effect (Table 6.6). The regression parameters for the Claimant data and the Employer data (both including the control group) are almost identical. Gender, MALE, is the exception; Gender has no significant influence on the re-employment probability in the Employer data. The shape of the estimated λ 's indicate a U-shaped λ .

We end with a discussion on the selectivity in the bonus data. The compliance rate in the Claimant group, 85%, was much higher than the compliance rate in the employer group, 65%. Many individuals in the Employer group, apparently and contrary to our findings, did not perceive a bonus paid to their new employer beneficiary for their job search. Following Moffitt (1983) this partial compliance may be explained by a stigma effect. However, this is a tentative explanation because our analysis only adjust for (possible) selective compliance. It does not provide a model for the selection process. Thus, both an advantage and a drawback of our method is that we do not make any assumptions on the selection process and therefore cannot tell why individuals make such a selective decision.

6.5 Conclusion

In this article we proposed and implemented an instrumental variable estimation procedure for duration models. We show how the effect of an endogenous variable on the duration in a Generalized Accelerated Failure Time (GAFT) model can be estimated. The GAFT model is based on a transformation of the durations that

encompasses both the Accelerated Failure Time (AFT), very popular in biostatistics, and the Mixed Proportional Hazards (MPH) model, very popular in econometrics. The interpretation of regression coefficients in the GAFT is in terms of shifting the quantiles of the distribution.

The basis of the Instrumental Variable Linear Rank estimator is that for the true GAFT model the instrument does not influence the hazard of the transformed duration. This implies that a rank test on the significance of the effect of instrument on the hazard of the transformed duration is zero. The IVLR estimation procedure is based on the inverse of an extended, including all the parameters of the GAFT model, rank-test. The estimation procedure is related to the rank estimation procedures of Robins and Tsiatis (1991) and of Bijwaard and Ridder (2005). The Two Stage Linear Rank procedure of Bijwaard and Ridder (2005) is based on a semi-parametric MPH and requires preliminary estimates of the baseline hazard. The Rank Preserving Structural Failure Time Model of Robins and Tsiatis (1991) is based on the strong version of the Accelerated Failure Time model. Their model imposes a strong non-interaction assumption. This implies that if two individuals have the identical observed durations and observed treatment histories then they would have had identical durations had treatment always been withheld. The IVLR estimator does neither impose the non-interaction assumption nor requires preliminary estimates of the baseline hazard.

The estimation procedure is also related to quantile-regression, in particular Koenker and Biliias (2001) and Koenker and Geling (2001). It is, however, unclear how these methods can handle time-varying endogenous variables. Because the IVLR is based on a vector of mean restrictions it is related to the well-known GMM estimation procedure. Similar to the application of GMM estimation choosing the right weight functions can improve the efficiency. However, again similar to the GMM, these optimal weight functions are not directly observable. Fortunately, an adaptive (or even 2 step) procedure can provide the efficient IVLR.

We can give a causal interpretation to the effect of the endogenous variable the IVLR identifies for the GAFT model. However, the causal effect is defined in terms of shifting the quantiles of the outcome distribution and not in terms of the (Local) Average Treatment Effect, common in the treatment evaluation literature. But averages are less useful to base treatment effects on for duration data, due to censoring and time-varying treatment.

The empirical application shows that the ML and ITT estimates for the employer group, in which the new employer of the claimant receives the bonus, are downward biased due to endogeneity. In the claimant group, in which the claimant himself receives the bonus, the ML and ITT estimates are close to the IVLR estimates. This might indicate that the endogeneity of the decision to participate in this group is rather small. Incorrectly assuming an AFT model can give misleading conclusions about the effects of a bonus on the re-employment hazard. In the Illinois bonus re-employment experiment many unemployed found a job just before their UI-benefits expires. This induces a spike in the re-employment hazard. In the GAFT, even with a constant regression coefficient, such a spike leads to an effect that changes over

the quantiles. This has important implications for the evaluation of the effect of a possible endogenous variable on a duration.

Social experiments may provide instruments for an endogenous variable. With good instruments available the proposed method can be very useful in analyzing the effects of a possible endogenous variable on an inherently duration outcome. Examples in population studies include the effect of training programs on the unemployment duration, policies to increase the birth rate and migration policies.

There are several issues that need further research. First, the current approach to adjust for endogenous censoring implies loss of information and depends on the (unknown) parameters of the model. An important improvement would be to find a method to adjust for endogenous censoring that is parameter independent and minimizes the loss of information. Another related issue is that if the IVLR assumes that the censoring time is (potentially) known in advance. Further research on more general censoring patterns deserve attention. Second, in our empirical application we have, because of random assignment, a perfect assignment. Such an instrument is, however, not always available. Finding good instruments is therefore an important issue just as the influence of weak instruments on the properties of the estimator. A final issue for further research is the extension of the IVLR to recurrent duration data, like repeated unemployment spells.

References

- Abbring, J.H. and G.J. van den Berg (2005). Social experiments and instrumental variables with duration outcomes. Tinbergen Institute, discussion paper, TI 2005–47.
- Andersen, P.K., O. Borgan, R.D. Gill and N. Keiding (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.
- Angrist, J.D. and A.B. Krueger (1999). Empirical strategies in labor economics. In: *Handbook of Labor Economics*, Vol. 3A, eds. O. Ashenfelter and D. Card. Amsterdam: North-Holland.
- Ashenfelter, O., D. Ashmore and O. Deschène (2005). Do unemployment insurance recipients actively seek work? Evidence from randomized trials in four U.S. states. *Journal of Econometrics* 125: 53–75.
- Bijwaard, G.E. and G. Ridder (2005). Correcting for selective compliance in a re-employment bonus experiment. *Journal of Econometrics* 125: 77–111.
- Brännäs, K. (1992). *Econometrics of the Accelerated Duration Model*. Umeå: Solfjäders Offset AB.
- Cox, D.R. and D. Oakes (1984). *Analysis of Survival Data*. London: Chapman and Hall.
- Han, A.K. (1987). Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator. *Journal of Econometrics* 35: 303–316.
- Heckman, J.J., R.J. LaLonde and J.A. Smith (1999). The economics and econometrics of active labor market programs. In: *Handbook of Labor Economics*, Vol. 3A, eds. O. Ashenfelter and D. Card. Amsterdam: North-Holland.
- Kalbfleisch, J.D. and R.L. Prentice (2002). *The Statistical Analysis of Failure Time Data (second edition)*. New York: John Wiley & Sons.
- Klein, J.P. and M.L. Moeschberger (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer-Verlag.
- Koenker, R. and Y. Biliias (2001). Quantile regression for duration data: A reappraisal of the Pennsylvania reemployment bonus experiments. *Empirical Economics* 26: 199–220.

- Koenker, R. and O. Geling (2001). Reappraising medfly longevity: A quantile regression survival analysis. *Journal of the American Statistical Association* 96: 458–468.
- Lalive, R., J.C. van Ours and J. Zweimüller (2005). The effect of benefit sanctions on the duration of unemployment. *Journal of the European Economic Association* 3: 1386–1417.
- Meyer, B.D. (1995). Lessons from the U.S. unemployment insurance experiments. *Journal of Economic Literature* 33: 91–131.
- Meyer, B.D. (1996). What have we learned from the Illinois reemployment bonus experiment? *Journal of Labor Economics* 14: 26–51.
- Moffitt, R. (1983). An economic model of welfare stigma. *American Economic Review* 73: 1023–1035.
- Powell, M.J.D. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal* 7: 155–162.
- Prentice, R.L. (1978). Linear rank tests with right censored data. *Biometrika* 65: 167–179.
- Press, W.H., B.P. Flannert, S.A. Teukolsky and W.T. Vetterling (1986). *Numerical Recipes: The Art of Scientific Computing*. Cambridge: Cambridge University Press.
- Ridder, G. (1990). The non-parametric identification of generalized accelerated failure-time models. *Review of Economic Studies* 57: 167–182.
- Robins, J.M. and A.A. Tsiatis (1991). Correcting for non-compliance in randomized trials using rank-preserving structural failure time models. *Communications in Statistics Part A: Theory and Methods* 20: 2609–2631.
- Tsiatis, A.A. (1990). Estimating regression parameters using linear rank tests for censored data. *Annals of Statistics* 18: 354–372.
- Van den Berg, G.J. (2001). Duration models: Specification, identification, and multiple duration. In: *Handbook of Econometrics*, Vol. 5, eds. J. Heckman and E. Leamer. Amsterdam: North-Holland.
- Van den Berg, G.J., J.C. van Ours and B. van der Klaauw (2004). Punitive sanctions and the transition rate from welfare to work. *Journal of Labor Economics* 22: 211–241.
- Woodbury, S.A. and R.G. Spiegelman (1987). Bonuses to workers and employers to reduce unemployment: Randomized trials in Illinois. *American Economic Review* 77: 513–530.
- Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *Annals of Statistics* 21: 76–99.

Appendix 1: Identification of the GAFT Model

Assume that the regression function in the GAFT model is log-linear. Then, the model is characterized by the non-negative function $\lambda(t; \alpha)$ defined on $[0, \infty)$, the distribution of U_0 and the regression parameter β . Ridder (1990) has shown that if the covariates are time constant, all observationally equivalent GAFT models, i.e. models that give the same conditional distribution of T given X , have regression parameters $d\beta$, integrated transformation $c_1(\int_0^t \lambda_0(s; \alpha_0) ds)^{c_2}$ and U_0 distribution $G_0((\frac{u}{c_1})^{1/c_2})$ for some constants $c_1, c_2 > 0$. The equivalent class follows from the fact that a GAFT model with time constant covariates can be expressed as a transformation model

$$\ln \left(\int_0^T \lambda_0(s; \alpha_0) ds \right) \stackrel{d}{=} -\beta'_0 X + \ln U,$$

and the constants c_1, c_2 correspond to addition of e^{c_1} to and division by c_2 of the left- and right-hand sides.

With time-varying covariates, the set of observationally equivalent GAFT models is generally smaller. In particular, the power transformation that gives an observationally equivalent model if the covariates are time constant, in general does not result in a GAFT model. As an example consider the GAFT model with time-varying regressors that differ between two groups. In group I

$$X(t) = \begin{cases} 1 & \text{if } 0 \leq t \leq 1, \\ 0 & \text{if } t > 0. \end{cases}$$

and in group II , $X(t) = 0$; $t \geq 0$. Moreover $\lambda_0(t; \alpha) = \alpha t^{\alpha-1}$. With time constant regressors the parameter α is not identified. It can be shown that the observationally equivalent GAFT models have transformation $c_1 t^\alpha$ and U -distribution with survival $G_u(\frac{u}{c_1})$. Hence, with time-varying covariates α is identified (and so is β).

We conclude that identification depends on whether the covariates are time constant or time-varying. If the covariates are time constant we can identify the transformation $h(T, \bar{X}(T); \theta_0)$ up to a power and β up to scale (with the power and the scale being equal). Moreover, if we fix the power we can identify $h(T, \bar{X}(T); \theta_0)^{c_2}$ up to scale and the distribution of U_0 up to the same scale parameter.

If the covariates are time-varying we can, except in special cases, identify $h(T, \bar{X}(T); \theta_0)$ and the distribution of U_0 up to a common scale parameter. Because we leave the distribution of U_0 unspecified in our estimation method, we can not use restrictions on U_0 to find the scale parameter. For that reason we normalize $h(T, \bar{X}(T); \theta_0)$ by setting $h(T, 0; \theta_0) = 1$ for some $t_0 > 0$. With time constant regressors we need the same normalisation, but in addition we need to set one regression coefficient equal to one. Of course, we could choose a class of transformations that is not closed under the power transformation. This amounts to identification by functional form.

Finally, we need a condition on the sample paths of X in the population. If we rewrite (6.1) as

$$\int_0^T e^{\ln \lambda(s; \alpha_0) + \beta_0 X(s)} ds = U_0 \quad (6.23)$$

we require that

$$\Pr(\ln \lambda(s; \alpha_0) + \beta_0 X(\cdot) = 0) = 0 \quad (6.24)$$

where the probability is computed over the distribution of X as a random function of t and 0 is the zero function. In other words, $\ln \lambda$ is not collinear with X .

For the identification in the GAFT model with endogenous variables we need additional assumptions on the instrument. First, the instrument should only affect the duration through the endogenous variable and not directly. Second, the value of the instrument should influence the value of the endogenous variable in a non-trivial way. For example, if both the instrument and the endogenous variable are binary then $\Pr(D = 1 | R = 1) > 0$ and $\Pr(D = 0 | R = 0) > 0$.

Appendix 2: Counting Process Interpretation

The density and the survival function of a duration T can be expressed as functions of the hazard rate. These expressions can be used to obtain a likelihood function. In this appendix we use a different (but of course equivalent) representation of the relation between the hazard rate and the random duration. In particular, we use the framework of counting processes (see e.g. Andersen et al. 1993 and Klein and Moeschberger 1997). The main advantage of this framework is that it allows us to express the duration distribution as a regression model with an error term that is a martingale difference. This simplifies the analysis of the estimator. The conditions for non selective observation can be precisely stated in this framework. The same is true for conditions on time-varying covariates.

The starting point is that the hazard of T is the intensity of the counting process $\{N(t); t \geq 0\}$ that counts the number of times that the event occurs during $[0, t]$. The counting process has a jump $+1$ at the time of occurrence of the event.⁷ A jump occurs if and only if $dN(t) = N(t) - N(t-) = 1$. For duration data, the event can only occur once. In many unemployment studies the individuals are only observed until re-employment. So, at most one jump is observed for any unit. To account for this we introduce the observation indicator $Y(t) = I(T \geq t)$ that is zero after re-employment. By specifying the intensity as the product of this observation indicator and the hazard rate we effectively limit the number of occurrences of the event to one. We assume that the observation indicator only depends on events up to time t . The observation process is assumed to have left-continuous sample paths. We define the history of the process up to time t by $H(t) = \{\bar{Y}(t), D, \bar{X}(t)\}$, where $\bar{Y}(t) = \{Y(s), 0 \leq s \leq t\}$. The history $H(t)$ only contains observable events.

Let V be some unobserved variable that both influence the endogenous variable and the duration. An example is the, usually, unobserved search intensity of unemployed looking for a job. We assume that V and $\bar{X}(t)$ are stochastically independent. Denote $H^V(t) = \{H(t), V\}$, the history that also includes the unobservables. As with dynamic regressors in time-series models, the time-varying $X(t)$ may depend on the dependent variable up to time t but not after time t (conditionally on V). Thus D only depends on $H^V(t)$ and $X(t)$ only on $H(t)$. In the counting process literature such a time-varying covariate is called predictable. We will use the econometric term predetermined.

If the conditional distributions of $N(t)$ given $H^V(t)$ or $H(t)$ are well-defined (see Andersen et al. (1993) for assumptions that ensure this) we can express the probability of an event in $(t - dt, t]$ as⁸

$$\Pr(dN(t) = 1 | H^V(t)) = Y(t) \kappa(t | \bar{X}(t), D, V) dt \quad (6.25)$$

⁷ The sample paths are assumed to be right-continuous.

⁸ Because the sample paths of $\{Y(t), X(t), t \geq 0\}$ are assumed to be left-continuous (as is the baseline hazard), we can substitute t for $t - dt$ in (B.1).

with $\kappa(t | \cdot)$ is the hazard of T at t given $\bar{X}(t)$, D and V . By the Doob-Meier decomposition

$$dN(t) = Y(t) \kappa(t | \bar{X}(t), D, V) dt + dM(t) \tag{6.26}$$

with $\{M(t); t \geq 0\}$ a (local square integrable) martingale. The conditional mean and variance of this martingale are

$$E(dM(t) | H(t)) = 0 \tag{6.27}$$

$$\text{Var}(dM(t) | H(t)) = Y(t) \kappa(t | \bar{X}(t), D, V) dt \tag{6.28}$$

The (conditional on $H(t)$) mean and variance of the counting process are equal, so that the disturbances in Equation (6.26) are heteroscedastic. The probability in Equation (6.21) is zero, if the individual is not at risk.

A counting process can be considered as a sequence of Bernoulli experiments, because if dt is small Equations (6.21) and (6.28) give the mean and variance of a Bernoulli random variable. The relation between the counting process and the sequence of Bernoulli experiments is given in Equation (6.26), which can be considered as a regression model with an additive error that is a martingale difference. This equation resembles a time-series regression model. The Doob-Meier decomposition is the key to the derivation of the distribution of the estimator, because the asymptotic behavior of partial sums of martingales is well-known.

The GAFT model transforms the observed duration T to a transformed duration U_0 . The transformation involved a parameter vector $\theta_0 = (\beta'_0, \gamma'_0, \alpha'_0)'$. We denote the transformation for parameter vectors $\theta \neq \theta_0$ by $U(\theta)$ with $U_0 = U(\theta_0)$. The distribution of $U(\theta)$ can also be represented by a (transformed) counting process $\{N^U(u); u \geq 0\}$. The relation between the original and transformed counting process, the observation indicator, and the time-varying exogenous covariates is

$$\begin{aligned} N^U(u; \theta) &= N(h^{-1}(u; \theta)) & Y^U(u; \theta) &= Y(h^{-1}(u; \theta)) \\ X^U(u; \theta) &= X(h^{-1}(u; \theta)) & I_k^U(u; \theta) &= I_k(h^{-1}(u; \theta)) \end{aligned}$$

with $h(T; \theta) = h(T, \bar{X}(T), \bar{D}(T); \theta)$, defined in (6.4), and $I_k(t) = I(t_k < t \leq t_{k+1})$. For $\theta = \theta_0$ we denote $h_0(T) = h(T; \theta_0)$. The corresponding history is $H^U(u; \theta) = \{\bar{Y}^U(u; \theta), \bar{X}^U(u; \theta), \bar{I}_k^U(u; \theta), D\}$. In the sequel we suppress θ and write $Y^U(u)$, $N^U(u)$, $\bar{X}^U(u)$, $\bar{I}_k^U(u)$ and $H^U(u)$ for $\theta \neq \theta_0$ and $Y_0(u)$, $N_0(u)$, $\bar{X}_0(u)$, $\bar{I}_{k0}(u)$ and $H_0(u)$ for $\theta = \theta_0$. The intensity of the transformed counting process with respect to history $H^U(u)$ is obtained by the innovation theorem (see Andersen et al. 1993, pp. 80, 87)⁹

⁹ If $U = h(T)$ and κ_T is the hazard rate of the distribution of T , then the hazard rate of the distribution of U is $\kappa_U(u) = \kappa_T(h^{-1}(u)) \frac{1}{h'(h^{-1}(u))}$.

$$\Pr(dN^U(u) = 1 | H^U(u)) = Y^U(u) \mathbb{E} \left[\frac{\lambda(h^{-1}(u; \theta); \alpha_0)}{\lambda(h^{-1}(u; \theta); \alpha)} e^{(\beta_0 - \beta)' X^U(u)} \right. \\ \left. \times \exp \left(\sum_{k=1}^K (\gamma_{k0} - \gamma_k) I_k^U(u) D \right) \kappa_o(h_0(h^{-1}(u; \theta))) | H^U(u) \right] du \quad (6.29)$$

We implicitly integrate with respect to the distribution of the unobserved V conditional on $H^U(u)$. Note that these unobserved covariates are only introduced to ascertain the predictability of the endogenous covariate process. Although the distribution of those variables determines the distribution of U_0 , the consistency of the IVLR is independent of that distribution. Unfortunately, even for the population parameters θ_0 the hazard of U_0 , $\kappa_0(u)$, still depends on the intervention path (through the correlation with V). If we condition on the history of the instruments instead of the actual endogenous covariates we do get the desired independence.

We must add the instrument R to the conditioning variables in (6.29) if we consider instrumenting the endogenous variable. Let the UR -history, $H^{UR}(u) = \{Y^U(s), X^U(s), R; 0 \leq s \leq u\}$, be the history on the transformed durations in which the endogenous variable D is replaced by the instrument. Then, another application of the innovation theorem gives the intensity of the transformed process on the UR -history

$$\Pr(dN^U(u) = 1 | H^{UR}(u)) = Y^U(u) \mathbb{E} \left[\frac{\lambda(h^{-1}(u; \theta); \alpha_0)}{\lambda(h^{-1}(u; \theta); \alpha)} e^{(\beta_0 - \beta)' X^U(u)} \right. \\ \left. \times \exp \left(\sum_{k=1}^K (\gamma_{k0} - \gamma_k) I_k^U(u) D \right) \kappa_o(h_0(h^{-1}(u; \theta))) | H^{UR}(u) \right] du \quad (6.30)$$

which for the population parameters simplifies to $Y_0^U(u) \kappa_0(u) du$ with $H_0^{UR}(u) = H^{UR}(u; \theta_0)$. Note that (6.29) and (6.30) only differ in the history the intensities are conditioned on.

For further reference we denote the intensity in (6.30) by $\kappa_i^U(u; \theta)$ such that

$$\Pr(dN^U(u) = 1 | H^{UR}(u)) = Y^U(u) \kappa_i^U(u; \theta) du$$

which reduces to $\kappa_0(u)$ for the population parameters.

A common feature of duration data is that some of the observations are censored. Assume the censoring time, C , is (potentially) known. Then, the potential censoring time is known and the observed durations are $\tilde{T} = \min(T, C)$ and $\Delta = I(T \leq C)$, where Δ is one if T is observed.

Assume the piecewise constant structure for the effect of the endogenous variable in (6.5). This implies that for $t_k < t \leq t_{k+1}$, the coefficient of $D = 1$ is e^{γ_k} . We define the transformed censoring time $C^U(\theta)$ (possibly depending on the observed

history of other covariates) such that: (a) $T \geq C$ implies $h(T; \theta) \geq C^U(\theta)$ and (b) U_0 and R are independent on the interval bounded above by $C^U(\theta)$.

Note that we either observe $T \leq C$ and $\Delta = 1$, or $T > C$ and $\Delta = 0$. If some of the other covariates are also time-varying we have another identification problem, because these covariates are only observed up until \tilde{T} . The transformed censoring times (conditional on $T, C > t_k$) that take all these considerations into account are the sum of the transformed duration up to t_k , $h(t_k; \theta)$ and the censoring adjustment, i.e.

$$C^U(\theta) = \begin{cases} \int_0^C \lambda(s; \alpha) e^{\beta' X(s)} P(s; \gamma) ds & \text{if } T > C, \\ \int_0^T \lambda(s; \alpha) e^{\beta' X(s)} P(s; \gamma) ds + \int_T^C \lambda(s; \alpha) ds & \text{if } T \leq C. \end{cases} \quad (6.31)$$

where $P(s; \gamma) = I(s \geq t_k) \prod_{j=0}^k \min(e^{\gamma_j}, 1)$. From the last term on the right-hand side of (6.31) we see why we need to know C even for the uncensored observations. Otherwise we can not compute $C^U(\theta)$ for these observations. We can estimate the parameters of the model from the following observed data

$$\tilde{U}(\theta) = \min(U(\theta), C^U(\theta)), \quad \Delta^U(\theta) = I(U(\theta) < C^U(\theta)) \quad (6.32)$$

and $Y^U(u; \theta) = I(\tilde{U}(\theta) \geq u)$. Now $\tilde{U}(\theta_0)$ is independent of R for $\Delta^U(\theta_0) = 1$. Note that if, at least, one of the γ 's is different from zero, we introduce extra censoring on the transformed durations, because then some units with $\Delta = 1$ have $\Delta^U(\theta) = 0$.

The counting process interpretation allows for an alternative formulation of the estimating equations in (6.8). The relevant counting measure, $N_i^U(u)$, can be seen as a discrete 'probability distribution' that assigns weight unity to uncensored transformed durations and is zero elsewhere. Then the estimating equations can be expressed as an integral with respect to that counting process

$$S_n(\theta; W) = \sum_{i=1}^n \int_0^{C_i^U} \{W(u, R_i) - \bar{W}(u; \theta)\} dN_i^{\tilde{U}}(u) \quad (6.33)$$

where C_i^U is the transformed censoring time defined in (6.31).

Appendix 3 Asymptotic Properties of the IVLR

In this section we discuss the asymptotic behavior of the Instrumental Variable Linear Rank estimator. The counting process framework simplifies the derivation of these asymptotic properties. We assume a piecewise constant λ for the GAFT model.

We make the following assumptions:

1. The covariate process $X(t)$ is predetermined, i.e. its distribution is independent of $\{H(s), s > t\}$. The sample paths of the covariate process are bounded and at least one of time-varying covariates is a continuous variable.
2. The observation process $Y(t)$ is cadlag and $Y(t)$ is predetermined. Moreover,

$$\Pr(dN(t) = 1 | Y(t) = 1, H(t)) = \Pr(dN(t) = 0 | Y(t) = 0, H(t))$$

3. The population distribution of T given \bar{X} and \bar{D} satisfies

$$\int_0^T \lambda(s; \alpha_0) e^{\beta'_0 X(s) + \psi(s, D; \gamma_0)} ds = U_0$$

The absolutely continuous distribution of U_0 does not depend on \bar{X} or \bar{R} . The p.d.f. of U_0 is bounded.

4. The transformed observation process $Y^U(u) = I(\tilde{U}(\theta) \geq u)$ is cadlag and pre-determined, with $\tilde{U}(\theta) = \min(U(\theta), C^U)$ and C^U defined in (6.31).
5. The instrumental function W is bounded and left-continuous.
6. The intensity of $U(\theta)$, $\kappa_i^U(u)$ given history $H^{UR}(u)$ in (6.30) can be linearized in a neighborhood of θ_0 as a function of θ , i.e. there exist $\mu(u)$ and $\varepsilon > 0$ such that for $\|\theta - \theta_0\| < \varepsilon$

$$|\kappa_i^U(u; \theta) - \kappa_0(u) - (\theta - \theta_0)' d_{i0}(u)| \leq \|\theta - \theta_0\|^2 \mu(u)$$

for $u \leq C_0 = C^U(\theta_0)$ with

$$d_{i0}(u) = \left. \frac{\partial \kappa_i^U(u; \theta)}{\partial \theta} \right|_{\theta = \theta_0}$$

7. There exists a continuous function $a(u; \theta)$ of θ in a neighborhood B of θ_0 such that

$$\sup_{u \leq C_0} \sup_{\theta \in B} \|\bar{W}(u; \theta) - a(u; \theta)\| = 0$$

where

$$\bar{W}(u; \theta) = \frac{\sum_{j=1}^n Y_j^U(u) W(u, \bar{X}_j^U(u), R_j)}{\sum_{j=1}^n Y_j^U(u)}$$

8. There exists a continuous matrix function $A(u; \theta)$ of θ in a neighborhood B of θ_0 such that

$$\sup_{u \leq C_0} \sup_{\theta \in B} \left\| \frac{1}{n} \sum_{i=1}^n [W(u, \bar{X}_i^U(u), R_i) - \bar{W}(u; \theta)] \right. \\ \left. \times [W(u, \bar{X}_i^U(u), R_i) - \bar{W}(u; \theta)]' Y_i^U(u) - A(u; \theta) \right\| \xrightarrow{p} 0$$

9. There exists a continuous matrix-function $V(u; \theta)$ of θ in a neighborhood B of θ_0 such that

$$\sup_{u \leq C_0} \sup_{\theta \in B} \left\| \frac{1}{n} \sum_{i=1}^n [W(u, \bar{X}_i^U(u), R_i) - \bar{W}(u; \theta)] \right. \\ \left. \times d_{i0}(u)' Y_i^U(u) - V(u; \theta) \right\| \xrightarrow{p} 0$$

The starting point is (6.32), which can, for θ in a small neighborhood of θ_0 , be rewritten as

$$S_n(\theta; W) = \sum_{i=1}^n \int_0^{C_{i0}} \{W(u, \bar{X}_i^U(u), R_i) - \bar{W}(u; \theta)\} dN_i^U(u) \\ + \sum_{i=1}^n \int_{C_i^U}^{C_{i0}} \{W(u, \bar{X}_i^U(u), R_i) - \bar{W}(u; \theta)\} dN_i^U(u) \quad (6.34)$$

Substitution of the Doob-Meier composition in the first term on the right for N_i^U gives

$$S_n(\theta; W) = \sum_{i=1}^n \int_0^{C_{i0}} \{W(u, \bar{X}_i^U(u), R_i) - \bar{W}(u; \theta)\} dM_i^U(u) \\ + \sum_{i=1}^n \int_0^{C_{i0}} \{W(u, \bar{X}_i^U(u), R_i) - \bar{W}(u; \theta)\} \kappa_i^U(u) Y_i^U(u) ddu \quad (6.35)$$

We consider both terms separately. The first term is, for θ close to θ_0 , close to $S_n(\theta_0; W)$ and for the second term we have

$$(\theta - \theta_0) \cdot \sum_{i=1}^n \int_0^{C_{i0}} \{W(u, \bar{X}_i^U(u), R_i) - \bar{W}(u; \theta)\} \\ \times \frac{\partial \kappa_i^U(u)'}{\partial \theta} Y_i^U(u) du + O_p(\|\theta - \theta_0\|^2)$$

Returning to (6.34) we note that the second term in this equation equals

$$\sum_{i=1}^n \{ [W(C_{i0}, \bar{X}_{i0}(C_{i0}), R_i) - \bar{W}(C_{i0}; \theta_0)] \times \theta_0(C_{i0}) Y_i(C_{i0}) \} + O_p(\|\theta - \theta_0\|^2)$$

The term between brackets is the covariance between $\theta_0(C_{i0})$ and $W(C_{i0}, \bar{X}_{i0}(C_{i0}), R_i)$ which is zero. Thus this whole term is zero for θ close to θ_0 and we have

$$S_n(\theta; W) \approx S_n(\theta_0; W) + n \int_0^{C_0} Z(u; \theta_0) du \cdot (\theta - \theta_0) \quad (6.36)$$

Hence, approximately for the IVLR estimator $\hat{\theta}_n(W)$

$$\sqrt{n}(\theta_n(W) - \theta_0) = \left[\int_0^{C_0} Z(u; \theta_0) du \right]^{-1} \frac{1}{\sqrt{n}} S_n(\theta_0; W) \quad (6.37)$$

The proof of the consistency and asymptotic normality are both based upon the asymptotic linearity of $S_n(\theta; W)$ in the neighborhood of the true value θ_0 . We follow the reasoning of Tsiatis (1990). Instead of a mean and variance condition, we have a mean and three covariance conditions. Let $\tilde{S}_n(\theta; W)$ be the right-hand side of (6.36). The following lemma shows that the linearization in (6.36) is uniformly close to the original estimating function

Lemma 1. *In neighbourhoods of $O(n^{-1/2})$ of θ_0*

$$n^{-1/2} \|\tilde{S}_n(\theta; W) - S_n(\theta; W)\|$$

converges uniformly to zero.

This lemma implies that $n^{-1/2}\tilde{S}_n(\theta; W)$ and $n^{-1/2}S_n(\theta; W)$ are asymptotically equivalent in a neighbourhood close to θ_0 .

Proof. This can be proved in lines of Tsiatis (1990) Lemma (3.1) and (3.2) and Theorem (3.2) and this is, because of the analogy, not repeated here.

Proof of Theorem 1 and Theorem 2 According to Lemma 1 are $n^{-1/2}S_n(\theta; W)$ in a neighbourhood close to θ_0 asymptotically equivalent to $n^{-1/2}\tilde{S}_n(\theta; W)$. Then the estimates θ^* and $\hat{\theta}$, with $\tilde{S}_n(\theta^*; W) = 0$, will also be asymptotically equivalent. Clearly, θ^* converges in probability to θ_0 . Hence, if we show that $\sqrt{n}(\hat{\theta} - \theta^*) \rightarrow 0$ then this would imply that $\hat{\theta}$ also converges in probability to θ_0 . Tsiatis (1990) argues that Lemma 1 suffices to proof this. This proves Theorem 1.

According to the Mann-Wald theorem convergence in probability implies convergence in distribution. We note that $\sqrt{n}(\theta^* - \theta_0) = n^{-1/2}Q^{-1}(W)S_n(\theta_0; W)$ clearly converges to a normal distribution with mean zero and variance matrix $Q^{-1}(W)\Omega(W)Q^{-1}(W)$. This completes the proof of Theorem 2.

Remark 1. To establish detailed conditions on when $\tilde{S}_n(\theta; W)$ has a unique root is rather tedious; however Ying (1993) gave an excellent general treatment on rank estimation, which can also be used for the estimating equations in this article.

Appendix 4 Additional Tables for the IVLR of Reemployment Bonus Experiment

Table 6.4 Descriptive statistics for control, claimant bonus and employer bonus group

	Control group	Claimant bonus	Employer bonus
White	0.632	0.651	0.647
Black	0.271	0.251	0.256
Other	0.097	0.099	0.097
Male	0.547	0.563	0.538
Age 20–29	0.425	0.436	0.424
Age 30–39	0.333	0.324	0.326
Age 40–49	0.179	0.185	0.187
Age 50–54	0.063	0.054	0.064
Weekly benefit			
–\$51	0.088	0.085	0.084
\$52–\$90	0.201	0.212	0.217
\$91–\$120	0.169	0.176	0.179
\$121–\$160	0.190	0.196	0.181
\$161–	0.353	0.331	0.339
Dependence allowance	0.323	0.345	0.332
Average pre-claim earnings	3188	3222	3215
Average age	33.0	32.9	33.1
Average weekly	119.9	118.8	118.5
N	3952	4186	3963

Table 6.5 Instrumental Variable Linear Rank estimates for the regression coefficients of the Illinois data (Constant Bonus Effect)

	<i>Claimant</i>		<i>Employer</i>	
	AFT	GAFT ^a	AFT	GAFT ^a
<i>First stage</i>				
LNAGE	-0.5718 (0.0734)	-0.3424 (0.0897)	-0.5219 (0.0717)	-0.3379 (0.0699)
LNBPE	0.3528 (0.0510)	0.2146 (0.0601)	0.3188 (0.0512)	0.2036 (0.0482)
BLACK	-0.6636 (0.0526)	-0.3770 (0.0842)	-0.6264 (0.0510)	-0.3792 (0.0641)
MALE	0.1135 (0.0377)	0.0663 (0.0330)	0.0464 (0.0376)	0.0295 (0.0305)
LNBEN	-0.5841 (0.0867)	-0.3558 (0.1011)	-0.6263 (0.0871)	-0.4010 (0.0865)
<i>One step optimal</i>				
LNAGE	-0.5204 (0.0693)	-0.3612 (0.0653)	-0.4733 (0.0683)	-0.3110 (0.0603)
LNBPE	0.3537 (0.0473)	0.2266 (0.0449)	0.3133 (0.0483)	0.1871 (0.0424)
BLACK	-0.6162 (0.0509)	-0.3982 (0.0510)	-0.5646 (0.0495)	-0.3574 (0.0443)
MALE	0.1293 (0.0355)	0.0691 (0.0303)	0.0698 (0.0355)	0.0227 (0.0303)
LNBEN	-0.5924 (0.0813)	-0.3692 (0.0762)	-0.6040 (0.0826)	-0.3610 (0.0727)

^a GAFT piecewise constant intervals: 0-2, 2-4, 4-6, 6-10, 10-25, 25 →;

Note: Standard error in brackets.

Table 6.6 Instrumental Variable Linear Rank estimates for the regression coefficients of the Illinois data (Time-varying bonus effect)

<i>First stage</i>				
	<i>Claimant</i>		<i>Employer</i>	
	AFT	GAFT ^a	AFT	GAFT ^a
LNAGE	-0.5361 (0.0693)	-0.3285 (0.0897)	-0.5233 (0.0706)	-0.3355 (0.0763)
LNBPE	0.3313 (0.0481)	0.2139 (0.0617)	0.3153 (0.0506)	0.2029 (0.0530)
BLACK	-0.6086 (0.0494)	-0.3665 (0.0861)	-0.6268 (0.0501)	-0.3771 (0.0740)
MALE	0.1036 (0.0352)	0.0668 (0.0337)	0.0461 (0.0371)	0.0294 (0.0304)
LN BEN	-0.5470 (0.0867)	-0.3564 (0.1043)	-0.6187 (0.0859)	-0.3989 (0.0959)
<i>One step optimal</i>				
LNAGE	-0.4861 (0.0653)	-0.3288 (0.0664)	-0.4529 (0.0675)	-0.3660 (0.0622)
BPE	0.3332 (0.0442)	0.2061 (0.0455)	0.3017 (0.0474)	0.2236 (0.0434)
BLACK	-0.5644 (0.0476)	-0.3615 (0.0533)	-0.5286 (0.0488)	-0.4189 (0.0480)
MALE	0.1176 (0.0332)	0.0626 (0.0304)	0.0622 (0.0349)	0.0283 (0.0302)
LN BEN	-0.5501 (0.0765)	-0.3343 (0.0770)	-0.5813 (0.0815)	-0.4284 (0.0752)

^a GAFT piecewise constant intervals: 0-2, 2-4, 4-6, 6-10, 10-25, 25 →;

Note: Standard error in brackets.

Table 6.7 Estimated λ in GAFT model for the bonus data

Interval	<i>Constant bonus effect</i>		<i>Time-varying bonus effect</i>	
	First	Opt.	First	Opt.
<i>Claimant</i>				
0–2	0.8098 (0.4638)	0.7500 (0.2052)	0.8625 (0.5262)	0.9328 (0.2409)
2–4	0.3146 (0.3691)	0.2348 (0.1462)	0.3542 (0.4048)	0.2309 (0.1799)
4–6	–0.0782 (0.2646)	–0.0415 (0.1220)	–0.0390 (0.3015)	–0.0318 (0.1552)
6–10	–0.2743 (0.2392)	–0.1859 (0.1133)	–0.2341 (0.2807)	–0.2085 (0.1369)
10–25	–0.6868 (0.1626)	–0.6655 (0.1006)	–0.6077 (0.1758)	–0.6345 (0.1261)
<i>Employer</i>				
0–2	0.7095 (0.3063)	0.8929 (0.1450)	0.7088 (0.4375)	0.5647 (1716)
2–4	0.2540 (0.2134)	0.4451 (0.0939)	0.2542 (0.3344)	0.1464 (0.1227)
4–6	–0.1217 (0.2008)	–0.1178 (0.0925)	–0.1195 (0.2330)	0.0875 (0.1050)
6–10	–0.4552 (0.1516)	–0.2707 (0.0751)	–0.4526 (0.2255)	–0.4098 (0.0975)
10–25	–0.7492 (0.0971)	–0.6826 (0.0372)	–0.7180 (0.1015)	–0.6057 (0.0491)

Note: Standard error in brackets.

Chapter 7

Female Labour Participation with Concurrent Demographic Processes: An Estimation for Italy

Gustavo De Santis and Antonino Di Pino

7.1 Introduction

This paper sets out to measure the “true” influence of partnering and fertility decisions on women’s participation in the labour market in Italy in 2002. Our model is rather complex for the following reasons. Firstly, because we consider several demographic processes, all of which are potentially affected by endogeneity (i.e. are in turn influenced by labour market decisions). Secondly because we use a cross sectional data source with retrospective questions, which calls into question two additional issues: selectivity and treatment effects. Selectivity arises because only a few, non-random individuals (women in our case) are observed in a given state (e.g. at work, or with children). Treatment effects arise because certain experiences of the past (e.g. having found a husband), may later put a woman on a different life course, which affects her approach towards family formation and labour participation.

After a quick look at the main issues at stake and the solutions offered by the relevant literature (Section 7.2), we present our model (Section 7.3) and the data (Section 7.4). The results that we obtain (Section 7.5) are discussed (in Section 7.6) in the light of the institutional setting that characterises Italy.

7.2 Background

Female labour participation decisions cannot be fully understood if one ignores the demographic setting that surrounds such decisions (marital status, living arrangement, fertility, etc.). Unfortunately, however, the analysis of this type of influence, from “demography” to labour participation, is hampered by a problem of endogeneity, since these demographic variables are, in turn, influenced by the past and current work status of the woman herself, and, if relevant, of other household members, including her partner (Browning 1992). So, how can one say anything sensible about

Gustavo De Santis (✉)

Department of Statistics, University of Florence, Viale Morgagni, 59, Firenze, Italy
e-mail: desantis@ds.unifi.it

the effect of demography on labour market decisions, at the micro level? And what type of data does one need for such a study?

The specialised literature offers a few alternative solutions to this problem. In one, demographic decisions are assumed to be taken together with those on labour participation, jointly by husband and wife at the beginning of their life as a couple. Partners are thus assumed to know, right from the start, their optimum equilibrium in terms of standard of living, progeny (number and quality of children), and time allocation between work, indoors and outdoors, and leisure (Mincer 1963; Becker 1981; Cigno 1991). In such a setting, marriage, reproductive behaviour, and work schedules of both partners cannot be analysed in terms of cause and effect, because they are all determined jointly, and they all depend on some other prior variables, for instance the characteristics of the parental home. But these assumptions are probably too strong: they imply that the context is static and that the need to adapt to unforeseen circumstances never arises. Besides, this approach underplays the stochastic nature of human reproduction, as well as uncertainty and variability in parental income streams and market wage rate (Heckman and Macurdy 1980; Hotz and Miller 1988).¹

An alternative approach admits that there is a trade-off between labour supply and housework, including the time devoted to childcare (e.g. Joshi 1990; Dankmeyer 1996; Sousa-Poza, Schmid and Widmer 2001). This trade-off applies particularly to women, for whom children constitute an indirect or “opportunity” cost that, together with other variables (for instance, labour market prospects), affects women’s and couples’ decisions on labour supply.

The evaluation of the causal relationship between demographic and labour market choices is further complicated by the fact that there may be variables that affect both marital instability and female labour market participation: an increase in female wages, for instance, increases both (Becker, Landes and Michael 1977). Besides, what actually happens within a household is difficult to evaluate. For example, in case of marital instability, partners may not divorce immediately: instead, they may find various forms of non-cooperative equilibrium within their marriage, which influences, among other things, the allocation of their working time (Lundberg and Pollak 1993, 2003). Therefore, a correct estimate of the impact of marital instability on the economic status of women should be based on a fully exogenous assessment of the probability of divorcing, which, unfortunately, is not available (Bedard and Deschênes 2003).

As is now standard in this type of study, endogeneity can be reduced (or, at best, altogether eliminated) with instrumental variables: instead of regressing on actual variables (e.g. fertility and marital status), one can put in their place *theoretical* values, i.e. the fertility and marital status that would have been observed, given certain assumptions, had labour market behaviour not interfered with these processes. In order to estimate instruments correctly, one needs exogenous variables that satisfy

¹ Heckman and Macurdy (1976), for example, find evidence of a married female labour supply response to transitory shocks in household income.

two criteria: they must not be correlated with the disturbances of the participation equation, and they must be correlated with the allegedly endogenous variables (here: fertility and marital status). Unfortunately, variables with these characteristics are rare. Take education, for instance: it is not exogenous to work and fertility choices, and it should ideally be transformed into an instrument itself, with the use of extra control variables, taken from the background of women (e.g. their parental home; Bratti 2003). But this procedure cannot always be followed, because the relevant information may be missing, the model becomes complicate, and identification is not always assured. What researchers often do (and this is also our case) is to accept a few approximations, as long as the “exogenous” variable does not violate too patently the two previous basic conditions.

We note, finally, that the same caveats hold also for geographical mobility. This, too, is in part an endogenous variable, that is a conscious decision that women take with certain professional careers in mind, because living in, or moving to, the north of Italy significantly increases the chances of finding an occupation, especially for women. However, since geographical mobility is relatively low in Italy, where there is a strong tendency to live close to one’s parents, relatives and friends, we do not deem it inappropriate to focus mainly on the opposite influence: that of the geographical context on work chances. In short, we will treat the area of residence as another weakly exogenous variable.

7.3 Model Specification: Theoretical and Methodological Issues

Several methodological difficulties arise from the specification and estimation of models with endogenous regressors. The main problem is how to substitute a latent, partially endogenous regressor with an instrument that is observable only as a dichotomous or count variable. In doing this, we distinguish between two types of instruments: one is a “treatment effect estimate”, and concentrates on the consequences that may derive from *being* in a given situation (Angrist, Imbens and Rubin 1996; Angrist 2000), while the other is a “propensity score estimate”, and keeps into account the theoretical *propensity* each individual has of actually finding him/herself precisely in that situation.

Let us develop these concepts with specific reference to our case. In the estimation of female labour market participation L , endogeneity is an obstacle, because L is influenced by, but also influences, what happens in the demographic sphere D . Let us generically model this demographic process as $D = g(X_D; L; \varepsilon_D)$ and let the participation function be $L = f(X_L; D; \varepsilon_L)$, where X_D and X_L are exogenous, possibly partly overlapping sets of variables that influence D and L respectively, and where ε_D and ε_L are unobserved error terms. If the demographic factors D are endogenous, the errors ε_D and ε_L will be correlated.

The demographic process D may be characterized by a binary dummy variable, for example marital status: having ($D = 1$) or not having ($D = 0$) a partner affects a woman’s attitudes towards the labour market, and therefore affects the equation

$L = f(X_L; D; \varepsilon_L)$. We may consider this structural change as an “endogenous switching” due to the treatment effect of the endogenous dummy variable D (Goldfeld and Quandt 1973; Maddala and Nelson 1975; Maddala 1983; Terza 1998).

As mentioned, we attempt to correct the estimates for endogeneity by first estimating the reduced form $\hat{D} = g(X_D)$ on the basis of a set of exogenous regressors X_D that we assume to be independent of the individual choice of participating in the labour market. Then, \hat{D} can be included as an instrumental, non-endogenous regressor in the second stage estimation: $\hat{L} = f(X_L; \hat{D})$. The reduced form estimation $\hat{D} = g(X_D)$ originates a series of residuals, $\check{D} = D - \hat{D}$, representing the part of the demographic behaviour that our model cannot explain. They may depend on some unobservable components or on pure hazard, say, having or not having found the “right” partner in the marriage market.

The standard assumption, in these cases, is that \check{D} is a stochastic component, uncorrelated with the disturbances ε_L . However, we will also introduce another assumption: that the difference \check{D} has an interest in itself, because it can affect labour market participation. Consider, for instance, two otherwise identical women, whose reduced-form estimates of the chances to marry are, say, 70%, and imagine that one did actually get married some time in the past, while the other did not. Do we expect them to behave similarly in the labour market? And, if not, should the difference \check{D} (+0.3 for the ever-married one, -0.7 for the other) not be considered among the explanatory variables? It may be worthwhile, then, to rewrite the previous identity as $\hat{D} = D - \check{D}$, and to use the following model $L = f(X_L; D; \check{D}; \varepsilon_L)$,² as we did in a few cases in our application (e.g. in Equation 7, below).

If the disturbance terms ε_D and ε_L are correlated (a consequence of the presence of “endogenous switching”), the condition that permits us to identify the treatment effect is that at least one exogenous variable included in the set X_D does not appear among the explanatory variables of L (Imbens and Angrist 1994; Angrist 2000).

In this paper, we take into account that endogenous structural modifications in the model can take different forms. For instance, let us assume the following relationship between women’s fertility, C (= 0, 1, 2, ... children ever born), and a set of explanatory variables X_C

$$C | P = X_C \beta_C + \varepsilon_C \quad (7.1)$$

where β_C is a vector of coefficients, intercept included, and ε_C is a vector of random disturbances $N(0; \sigma_{\varepsilon_C}^2)$. The count, dependent variable $C | P$ depends, among other things, on the outcome of the binary switching variable $P = (0; 1)$, i.e. on whether the woman actually has a partner. The dummy variable P is characterized by a dichotomous rule of the following form:

$$P = \begin{cases} 1 & \text{if } X_p \beta_p + \varepsilon_p > 0 \\ 0 & \text{otherwise} \end{cases} \quad (7.2)$$

² Cfr. *inter alia*, Terza (1998).

where X_P is a matrix of observable exogenous variables “explaining” the presence of a partner, and β_P is the corresponding vector of parameters. The random vectors ε_P and ε_C are assumed to be jointly normal with mean vector zero. If there is endogenous switching, the covariance $\sigma_{\varepsilon_C; \varepsilon_P}$ differs from zero.

Notice that endogenous switching may take various forms:

- (i) “endogenous treatments effects”. The structural change in $X_C\beta_C$ is only influenced by the treatment effects (in this case, only by the presence or absence of a sexual partner);
- (ii) presence of “partially observable” endogenous explanatory variables. The dependent variable C is influenced by one or more extra explanatory variables related to P , which, unfortunately, cannot always be observed. For instance, a potentially relevant explanatory variable of a woman’s fertility C may be the age difference between herself and her partner, but this variable can be observed only if a partner is there;
- (iii) presence of an “endogenous latent variable”, which means that the dependent variable C can be explained both by the theoretical propensity of the subject to choose a specific outcome of P , and by her actual choice. As a matter of fact, several characteristics that enhance a woman’s fertility C also impact positively on her chances of having a partner P .

In the estimation, however, our specification is general enough, so as to encompass all these cases (as in Maddala and Nelson (1975), and in Terza (1998)).

Notice, finally, that in the two-stage procedure that we adopt here, we obtain a generally good “diagnostic” for instrumental variable (*IV*) estimates at the first stage (see, *infra*, Tables 7.2 and 7.3), which protects us from the risks that derive from poor specification and the use of “weak instruments” (Staiger and Stock 1997).

7.3.1 *The Model*

Let us sum things up: we want to model the labour force participation of Italian women in the year 2002: this will be our Equation (11). This participation is influenced by several “explanatory” variables, among which some are truly exogenous, while others need to be explained themselves, and this leads to a multiple equation model. On top of this, some of these “explanatory” variables are endogenous: i.e. they depend, at least partly, on the same variables that we want them to explain. We circumvent this difficulty in the standard way, using instrumental variables (*IV*).

The model, in its structural form, consists of the eleven equations of table 7.1, where the most interesting to us is Equation (11), a consistent estimate of the latent probability an Italian woman has to participate in the labour market, which empirically translates into a dichotomy (yes/no). The first ten equations are introduced either to estimate instrumental variables (Equations 7.1, 3, 5, 6, 7, and 9), or to

Table 7.1 The structural form of the model

(1)	$L^* = \alpha_1 + \alpha_2 Age_F + \alpha_3 Age_F^2 + \alpha_4 Age_M + \alpha_5 Age_M^2 + \alpha_6 Edu_M + \alpha_7 Edu_F + \alpha_8 South + u_1$
(2)	$H_L = L - [L^*] (L = 1 \text{ if } L^* > 0; L = 0 \text{ otherwise})$
(3)	$P^* = \alpha_9 + \alpha_{10} Age_F + \alpha_{11} Age_F^2 + \alpha_{12} Sib + \alpha_{13} Edu_{Par} + \alpha_{14} Edu_F + u_3$
(4)	$H_P = P - [P^*] (P = 1 \text{ if } P^* > 0; P = 0 \text{ otherwise})$
(5)	$Sen = \alpha_{15} + \alpha_{16} Sex + \alpha_{17} Edu_F + \alpha_{18} South + \alpha_{19} Jobs + u_5$
(6)	$W = \alpha_{20} + \alpha_{21} Sex + \alpha_{22} Edu + \alpha_{23} GNP_{reg} + \alpha_{24} White + \alpha_{25} [H_L] + u_6$
(7)	$C^* = \alpha_{26} + \alpha_{27} Age_F + \alpha_{28} Age_F^2 + \alpha_{29} Age_{diff} + \alpha_{30} Par + \alpha_{31} Edu_F + \alpha_{32} Sib + \alpha_{33} [P^*] + \alpha_{34} [H_P] + u_7$
(8)	$H_C = C - [C^*]$
(9)	$Div^* = \alpha_{35} + \alpha_{36} Age_F + \alpha_{37} South + \alpha_{38} Edu_{Par} + \alpha_{39} [H_P] + u_9$
(10)	$H_{Div} = Div - [Div^*] (Div = 1 \text{ if } Div^* > 0; Div = 0 \text{ otherwise})$

Finally, and most importantly for this study,

(11)	$L_F^* = \alpha_{41} + \alpha_{42} Age_F + \alpha_{43} Age_F^2 + \alpha_{44} Edu_F + \alpha_{45} GNP_{reg} + [\alpha_{46} W_{MN} + \alpha_{47} W_{FN} + \alpha_{48} W_{FS}] + \alpha_{49} [Sen_{FN}] + \alpha_{50} [H_{PN}] + \alpha_{51} [H_{PS}] + \alpha_{52} [H_{Div}] + \begin{cases} \alpha_{53} [C^*] \\ \alpha_{53} H_C \end{cases} + u_{11}$
------	---

($L_F = 1$ if $L^* > 0$; $L_F = 0$ otherwise; we use both alternatives, first $[C^*]$, in Equation (11a), and then H_C in Equation (11b) – see further in the text)

where there are 10 endogenous variables (in brackets):

C^* = expected number of children ever born {or H_C = unexpected component of fertility}

Div^* = propensity to divorce

H_L, H_P, H_{Div} = Heckman (or treatment) correction factors

L^* = Labour participation (of men and women)

L_F^* = Labour participation of women

P^* = propensity to have a Partner

$Sen = \ln(\text{Seniority}/\text{Age})$ (Seniority = Number of years of work)

$W = \ln(\text{Wage})$ [with exogenous fixed effects for Males/Females and for geographical area, North-centre, or South]

Besides, we consider 15 exogenous variables:

$Age; Age_F; Age_M = age$ (and fixed effects Male/female)

$Age_F^2; Age_M^2 = \text{squares of age}$ (and fixed effects Male/female)

$Age_{diff} = Age_F - Age_M$ (partner)

$Edu = \text{Education}$ (number of years of school)³

$Edu_{Par} = \text{Education of parents}$ (number of years of school of the parent with the highest education level)

$GNP_{reg} = \text{log of the per capita GNP of the administrative region}$

$Jobs = \text{Number of previously held jobs}$

$Par = \text{Dummy variable}$ (if at least one parent of the woman is living)

$Sex = \text{Dummy variable}$ (man = 0; woman = 1)

$Sib = \text{number of siblings still alive at the time of the interview}$

$South = \text{Geographical dummy variable}$ (South = 1; North-Centre = 0)

$White = \text{Dummy for type of occupation}$ (White collar = 1; Blue collar = 0) plus constant terms

Note that Equation (11) can be identified, because the number of the exogenous variables considered in the system, but excluded from the final equation (15–5=10) is larger than the number of endogenous variables included minus one (5).

An asterisk denotes a latent (continuous) variable, the empirical counterpart of which is ordinarily a dummy (e.g. L or P), or a count variable (e.g. C).

We tried several specifications for several variables: gender (M/F), region (Reg for single Region, out of the 20 that make up Italy; S for South and N for North-Centre) and cohort (Par refers to the parent generations – fathers and mothers of the adult women we are considering here). Only those that proved significant and consistent in the various attempts were retained in the final model.

Table 7.2 Estimation results of the reduced-form equations

Sample Estimator dependent variable	Equation (1)	Equation (3)		Equation (5)		
	10408 Probit L (dummy)	4526 (women) Probit P (dummy)		10408 OLS Sen		
	0 = unemployed <i>S.E.</i>	0 = no partner <i>S.E.</i>		log (years of work /age)	<i>S.E.</i>	
Intercept	-3.3650	0.0124	-2.2671	0.5480	-1.3705	0.0474
Age _M	0.2808	0.0124				
Age _M ²	-0.0038	0.0001				
Age _F	0.1968	0.0138	0.1757	0.0266		
Age _F ²	-0.0026	0.0002	-0.0018	0.0003		
Edu _M	0.0660	0.0060				
Edu _F	0.1065	0.0054				
South (North- Centre = 0)	-0.3870	0.0318			-1.4310	0.0507
Sib			0.0703	0.0211		
Edu _{par}			-0.0348	0.0082		
Edu _F			-0.0494	0.0093	-2.2491	0.0484
Sex (M=0; F=1)					0.1085	0.0060
Edu						
Jobs					0.3422	0.0137
H _L [^]						
GNP _{reg} (×10 ⁶)						
White						
Age _{Diff}						
Par						
P						
H _p [^]						
% correctly classified	0.7932		0.9450			
R ² adj.					0.3313	

estimate residuals (Equations 7.2, 4, and 8), which correct for selectivity both in the final Equation (11), and in a few intermediate Equations (6, 7 and 9).

³ We first calculated the “equivalent (or theoretical) number of years of school”, i.e. those in principle necessary for a non-repeating pupil to attain a given grade. However, the meaning of schooling changes over time, because the youngest cohorts tend to study considerably more than their predecessors. This trend is only interrupted by censoring, because the most recent cohorts haven’t completed their education, yet. In short, to make schooling comparable over the generations, we consider how better (or worse) off in this respect each individual is in comparison with the average of his or her own cohort. For instance, it normally takes 8 years to complete junior high school (*scuola media*), compulsory since 1963. For women aged 65 in 2000–02, whose average number of years of schooling was about 5, this case translates into +3, i.e. considerably above average. For women born around 1968, whose average is nearly 12 years spent at school, this translates into about -4 (i.e. strongly below average).

Table 7.3 Estimation results of the reduced-form equations

Sample Estimator Dependent variable	Equation (6) Employed: 5542 OLS W		Equation (7) Women: 4526 Log-Poisson C		Equation (9) Ever married women: 4221 Probit Div (dummy)	
	log (yearly wage)		Children ever born	<i>S.E.</i>	0 = no divorce	<i>S.E.</i>
Intercept	8.5443	<i>0.1169</i>	3.3697	<i>0.2884</i>	-1.0397	<i>0.6362</i>
Age _M						
Age _M ²						
Age _F			0.1423	<i>0.0131</i>	-0.0779	<i>0.0097</i>
Age _F ²			-0.0014	<i>0.0001</i>		
Edu _M						
Edu _F			-0.0370	<i>0.0033</i>		
South (North- Centre = 0)					-0.4106	<i>0.1580</i>
Sib			0.0494	<i>0.0052</i>		
Edu _{par}					0.0872	<i>0.0221</i>
Edu _F			-0.0370	<i>0.0033</i>		
Sex (M=0; F=1)	-0.3312	<i>0.0242</i>				
Edu	0.0350	<i>0.0030</i>				
Jobs						
H _L [^]	-0.1929	<i>0.0542</i>				
GNP _{reg} (× 10 ⁶)	19.2137	<i>2.2116</i>				
White	0.2307	<i>0.0229</i>				
Age _{Diff}			-0.0074	<i>0.0029</i>		
Par			0.0205	<i>0.0307</i>		
P			0.1201	<i>0.1207</i>		
H _P [^]			2.1784	<i>0.1577</i>	-4.9324	<i>0.5340</i>
% correctly classified					0.9754	
R ² adj.	0.1615					
Pseudo R ²			0.2155			

All the endogenous variables marked with an asterisk are latent, and not directly observable. For instance, we have the latent probabilities of working (L^*), having a partner (P^*) and having experienced a divorce (Div^* – for married women only). Endogenous fertility C^* (Equations 7 and 8), is the expected, unobserved number of children ever born to women of given characteristics (age, age difference with partner, number of siblings, ...), which, together with the difference between the observed and the expected number of children H_C , translates into actual fertility.

H_L (H for Heckman) in Equation (2) represents the random factors that may make actual employment status L differ from its propensity score L^* – for both men and women. Similarly, H_P in Equation (4) links actual to expected partner status. Notice that H_L appears as a regressor in Equation (6), \log of yearly

wage,⁴ and can be interpreted as a Heckman correction for the selectivity of actual employment status on wage, for men and women. The correction H_L proves necessary because only a non random sub-sample of individuals work, and the wage that can be observed on them may not be automatically extended to the whole of the population. Analogously, H_P represents the correction for the endogeneity of treatment effects in three cases: Equation (7) (expected fertility), Equation (9) (probability of marital dissolution),⁵ and Equation (11) (labour participation of women).⁶

Equations (5) and (6) describe seniority (relative to age, in *log*) and *log*-wage, and both variables help explain women's participation in Equation (11).

We are basically adopting a two-stage procedure. In the first stage, i.e. the first 10 equations, we estimate the instrumental variables; in the second stage (Equation 11), we launch the *Probit* estimation that we are most interested in, that of female labour participation, on the basis of exogenous and instrumental variables, i.e. avoiding endogeneity. This allows us to tentatively interpret our parameters in terms of cause and effect: although we cannot exclude that other types of correlation are at play, too, we can at least rule out inverse causation.

$[\hat{P}^*]$ and \hat{H}_P are, respectively, the theoretical values and the residuals of the *Probit* estimation of Equation (3) – having a partner. $[\hat{P}^*]$ measures how much the “theoretical” probability of entering a marital union affects fertility (Equation 7), while \hat{H}_P is an estimate of unobserved factors that lead to having or not having a partner. The inclusion of \hat{H}_P in Equation (7) tries to measure how much these unobserved factors influence fertility. \hat{H}_P is also adopted to correct the selection bias for the estimation of the probability to divorce, in Equation (9).

Finally, we introduce the instrumental variable $[\hat{H}_L]$ (residuals of the *probit* estimate of Equation 7.1), in order to correct for the bias that we may introduce in Equation (6), where we estimate wages on the selected sub-sample of the employed (Heckman 1974, 1979).

The path we are following here is rather complex, but, in our opinion, necessary, in order to avoid the several types of potential biases we described before. However, there are still a few problems that we cannot tackle with our data. Let us mention just two of them:

⁴ We use yearly and not hourly wage, so as to avoid the need to estimate separately the number of hours worked, not asked in the survey, and endogenous itself (see. e.g. Fortin-Lacroix 1997; for Italy, see Di Pino 2004).

⁵ The probability of marital dissolution can be defined only for women ever in marital union. Note that H_{Div} proves negatively correlated with the latent probability of being with a partner (not shown here). Therefore, H_P may be interpreted as a proxy for “search costs”, which impact on the decision to divorce (Becker, Landes and Michael 1977): the higher the probability of having a partner, the lower the search costs incurred to find a new one, should this search prove necessary.

⁶ Equations (7) and (11) are examples of the “endogenous switching” we discussed in the preceding paragraph.

- 1) we do not have sufficient information on the number of hours worked, either in the household or in the labour market. This is why we are modelling the simple dichotomous variable work/no work;
- 2) the number of children ever born depends, among other things, on having or not having experienced couple dissolution, but we ignore this influence in our model, because, once again, this would introduce endogeneity (what causes what?), the correct treatment of which would have further complicated our study. Note, however, that we are not trying to fully explain fertility,⁷ here: we are merely introducing it as an instrument in the explanation of labour force participation.

Let us briefly discuss the stochastic properties of the model. We assume that the error terms of Equations (7.1), (3), (9), and (11) are distributed as a normal standard. The distribution of the error terms of Equations (5), (6), and (7) is assumed to be normal, with zero mean. The existence of endogenous switching in fertility (Equation 7), influenced by the presence of a partner (Equation 3), can be modelled by the inclusion in the regression both of the dummy variable P (to have a partner or not) and the correction term for endogeneity, H_P . If the presence of a partner is endogenous, the error terms of Equations (3) and (7) are correlated, as discussed in Section 7.2. Analogously, the existence of an endogenous correlation between marital instability and labour market participation, or between marital status and participation implies a non-null correlation in the error terms of Equations (9) and (11) and in the error terms of Equations (3) and (11).

The term H_C , resulting from the fertility equation, is introduced as an explanatory variable of labour participation in Equation (11), as an alternative to the theoretical expected fertility, C^* . We estimate two distinct alternative specification of Equation (11). The first (11a), under the assumption that expected fertility affects female participation, and the second (11b) under the assumption that the unexplained component of fertility affects woman's behaviour in the labour market.

7.4 The Data

Our sample derives from the 2002 SHIW dataset, the Survey on Household Income and Wealth, carried out under the initiative of the Bank of Italy (2004). All the details of this survey, and the data itself, are freely available on the internet (<http://www.bancaditalia.it/statistiche/ibf>). Let us just mention that it is a biennial survey, on about 8 thousand private households (about 20 thousand people), mainly aimed at estimating income, saving and wealth, but with a large set of ancillary questions on work, household composition, and families of origin. We exploit especially this part of the survey, here. The survey has also a panel part in it, which however, we

⁷ Besides, we miss several important pieces of information in this respect, and notably contraception.

cannot use for our exercise, since it is too small, too short and with non trivial (and probably not random) attrition.

We selected 10408 of the surveyed persons: 4526 women of working age (18 to 60 years), and 5782 men. These are either their partners (not necessarily husbands) of any age, or other men aged 18 to 75. There are 3997 couples in our sample, either husband and wife, or simply cohabiting sexual partners.

Retrospective questions are not very many in the SHIW and, most importantly, not asked of all respondents: only of the “reference couple”, i.e. the first person in the household roster and his/her partner, if present. Fertility (here: number of children ever born, up to the date of the interview), is not asked directly. We reconstructed it from two separate questions: on the number of children present in the household and on the number of children living elsewhere. The latter question, once again, is asked only of the reference couple, which leaves out a certain proportion of women. We assumed that these women, relatively young and not in the reference couple, had no children living elsewhere: comparison of fertility levels (ours and the official one, estimated by Istat), indicates that this assumption is tenable, especially if one considers that fertility enters our estimates only instrumentally, because what we are really modelling is female labour participation. Notice that our calculations on the number of children ever born do not take into account those who died some time before the interview. This should not be too much of a problem given the very low level of infant and child mortality in Italy.

7.5 Results

Tables 7.2 and 7.3, below, show our first stage IV estimates. Notice that our parameters are always meaningful (with only a couple of minor exceptions, in Equation 7), and that the overall goodness of fit is relatively good. This protects us from the potential bias that may derive from the use of “weak instruments” (Bound, Jaeger and Baker 1995; Staiger and Stock 1997).

Although these are instrumental regressions, it may be worthwhile to stop a minute to consider what we get. Labour force participation (Equation 7.1) is bell-shaped with respect to age, with a maximum between 35 and 40 years, for both men and women. Work is more frequent for the well educated (especially for well educated women) and for those who live in the Centre-North of Italy.

Being currently with a sexual partner is more frequent among women around the age of 50, when own and parents’ education is low, and the number of siblings high (Equation 3). Seniority (the ratio between the number of years worked and own age, in log) is lower in the South, for women, and for people with low own or familial education (Equation 5). Wage (Equation 6) is estimated for the employed, but with Heckman’s factor (H_L) so as to correct for the distortion that derives from leaving aside those who do not work, and who are therefore selected. Wage is higher for men, for the well educated, for those who live in rich regions and for those who have “white-collar type” jobs.

Table 7.4 Female participation estimation

Sample (women) Estimator Dependent variable	Equation (11a)			Equation (11b)		
	4526 (2181 with $W_F = 1$)			Probit		
	L _F (Female work, dummy)					
	(0 = unemployed; 1 = employed)					
	Coeff.	S.E.	From equation	Coeff.	S.E.	From equation
Intercept	-6.1078	0.5150		-5.7587	0.5040	
Age _F	0.2419	0.0221		0.2264	0.0199	
Age _F ²	-0.0031	0.0002		-0.0029	0.0002	
Edu _F	0.1029	0.0123		0.1137	0.0118	
GNP _{reg} ($\times 10^6$)	34.9366	8.7894		33.3571	8.8462	
$W_{MN}^{\wedge} (\times 10^3)$	0.0111	0.0082	(6)	0.0093	0.0083	(6)
$W_{FN}^{\wedge} (\times 10^3)$	-0.1039	0.0205	(6)	-0.0102	0.0206	(6)
$W_{FS}^{\wedge} (\times 10^3)$	0.0689	0.0239	(6)	0.0634	0.0241	(6)
Sen_{FN}^{\wedge}	4.6803	0.6779	(5)	4.4688	0.6671	(5)
H_{PN}^{\wedge}	-0.5013	0.0855	(4)	-0.5949	0.0833	(4)
H_{PS}^{\wedge}	-0.6416	0.1208	(4)	-0.7337	0.1119	(4)
H_{Div}^{\wedge}	0.6022	0.1640	(10)	0.5861	0.1661	(10)
C^{\wedge}	-0.1257	0.0669	(7)			
H_C^{\wedge}				-0.1060	0.0347	(8)
% correctly classified	72.02%			72.09%		
Loglikelihood	-2555.9			-2545.6		

The number of children ever born (our indicator of past fertility – Equation 7), is estimated on women only. The number of children ever born, which peaks at the age of about 50, is higher for women with a partner (but we include Heckman's correction H_P here, too, because having a partner is endogenous to having had children), for those whose partner is younger, and for those whose parents are still alive, or at least one of them. We interpret this as an indication that family support with child care is (or has been) potentially available, because in Italy the adults very frequently live close to their (now old) parents. Finally, Equation 9 focuses on couple dissolution (divorce or separation), which, by definition, can characterise only ever-married women, and which therefore needs Heckman's correction factor H_P . Divorce decreases with age, because a cohort effect prevails: the young divorce much more than their predecessors, and this counts more than the fact that the young have had less time to go through this experience. Divorce is higher in the North-Centre of Italy, and among well-educated women.

Table 7.4 below shows our *Probit* estimates of Equation (11),⁸ which is the most important result of our research. There are two versions of this final equation, be-

⁸ The standard errors of the estimated coefficients are calculated by applying the correction of the variance of residuals generally utilized in two-stage estimates. To compute unbiased residuals, we use the observed endogenous variables, not the instrumental variables estimated at the first stage.

cause in one case we try estimated fertility \hat{C} among the regressors (Equation 11a), and in the other we try its “unexpected” (residual) component H_C (Equation 11a).⁹ In both cases, the parameters of the other variables change only marginally (same sign and same order of magnitude) and we may therefore comment on them considering indifferently Equations (11a) or (11b). Our exogenous variables affect the chances of female employment in the expected direction: it has an inverted U-shaped evolution with age (maximum at about 40 years), and both education, GNP_{reg} (a rough measure of the economic performance of the area where our respondents live), and (not very significantly, and only in them North-Centre of Italy) high partner’s wage make female employment more likely.¹⁰ As for the instrumental variables, having a partner (or, better: the residual, and therefore the “unexplained” part of this variable, H_P) reduces the likelihood of female employment, especially in the southern regions (H_{PS}). Fertility, too, depresses female employment: both in its “expected” (instrumental) part \hat{C} (Equation 11a) and in its “unexpected” (residual) component H_C (Equation 11b) and the order of magnitude is comparable. Divorce and separations induce women to work more.

7.6 Discussion

What do we learn from this study? That female employment depends in part on what happens in the demographic sphere (having children, living with a partner, etc.) and in part on the forces that shape these demographic decisions. With a few retrospective questions, and a sufficient number of variables, it seems possible to separate these two types of influence, at least partly, and therefore to obtain parameter estimates that get us closer to an interpretation in causal terms. A direct regression of female labour force participation on several of these variables (although generally yielding the same sign – not shown here), would not be interpretable in this sense, because of endogeneity and, in a few cases, selection.

Regression parameters are not always easy to interpret: the sign is clear, but what about the order of magnitude? One possible answer is to tabulate the expected employment profiles of women with given characteristics (age, education, marital

⁹ We cannot use both of them simultaneously because, when used jointly, they produce endogeneity.

¹⁰ The introduction of the regional GNP as a “fixed effect” among the explanatory variables in Equations (6), (11a) and (11b) may introduce some heteroskedasticity in the error terms “between” different regions, and a possible autocorrelation in the error terms “within” the observations in each region. All in all, however, we think that the opposite effect prevails: our equations are thus better identified, and this results in more (not less) efficiency in our estimates. In fact, Equations (11a), (11b) and (6) can be thought of as, respectively, a “supply function” of labour, and a “demand function” of wage: they are influenced by the level of economic activity and labour demand, both proxied by the regional GNP.

Table 7.5 Labour market participation probability of Italian women. Classification for age, fertility, and geographical area (in percent)

Age No. of Children	Married									
	25		30		35		40		50	
	CN	<i>South</i>	CN	<i>South</i>	CN	<i>South</i>	CN	<i>South</i>	CN	<i>South</i>
0	43.87	32.68	57.73	47.66	64.99	56.04	66.46	57.71	53.54	41.96
1	39.74	28.95	53.55	43.46	60.99	51.83	62.52	53.53	49.32	37.87
2	35.71	25.44	49.33	39.33	56.87	47.61	58.44	49.30	45.10	33.91
3	31.84	22.16	45.11	35.32	52.68	43.41	54.27	45.09	40.94	30.13
4	28.16	19.14	40.95	31.47	48.45	39.28	50.05	40.93	36.88	26.54
Unmarried										
0	62.49	56.46	75.09	71.36	80.96	78.36	82.21	79.79	72.37	67.12
1	58.41	52.26	71.61	67.65	77.95	75.13	79.32	76.67	68.72	63.21
2	54.23	48.03	67.92	63.76	74.68	71.65	76.16	73.31	64.88	59.15
3	50.02	43.83	64.03	59.72	71.18	67.96	72.76	69.71	60.88	54.99
4	45.80	39.69	60.00	55.57	67.46	64.08	69.13	65.92	56.75	50.78
Divorced										
0	66.71	55.46	78.27	70.11	83.43	76.98	84.40	78.25	75.01	64.92
1	62.78	51.25	75.02	66.33	80.65	73.63	81.73	75.00	71.53	60.92
2	58.71	47.02	71.54	62.38	77.61	70.06	78.79	71.52	67.83	56.80
3	54.54	42.83	67.84	58.30	74.32	66.28	75.60	67.82	63.95	52.61
4	50.32	38.72	63.96	54.12	70.79	62.33	72.15	63.94	59.91	48.38

state, region of residence, etc.) (Table 7.5).¹¹ Labour force participation for a typical Italian woman increases up to the age of about 40, and decreases subsequently. It is higher for the never married and the divorced, especially in the central and northern part of the country. And, not surprisingly, decreases with the number of children, by about 4 percentage points for each child.

Another possible use is to study elasticities. Let us just consider two examples. In the first (Fig. 7.1), we can see that the propensity to go to work, for Italian women, increases up to about age 40, which confirms our previous finding. The extra information, here, is that this increase, stronger for the younger generations, is especially marked in the south, and most particularly for the married. Their starting point (not shown here) is lower, but their trend is more steeply on the increase: in other words, they might be (slowly) catching up.

Similarly, Fig. 7.2 considers the elasticity of labour participation with respect to education. Extra years of school always increase female labour supply, but the increase is strongest in the South, where, as mentioned, starting levels are lower, but potential for growth apparently greater.

¹¹ Some of these, as discussed before, are not (fully) exogenous to labour market participation. Therefore, the simulated values of Table 7.5 only give us a rough idea of the true probabilities of the hypothetical women considered.

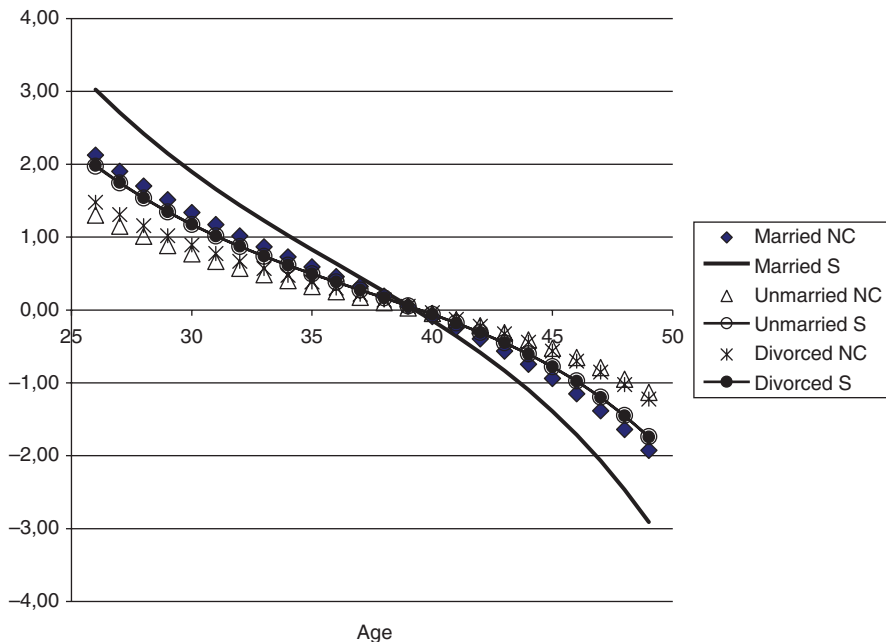


Fig. 7.1 Age-elasticity of Italian women labour market participation

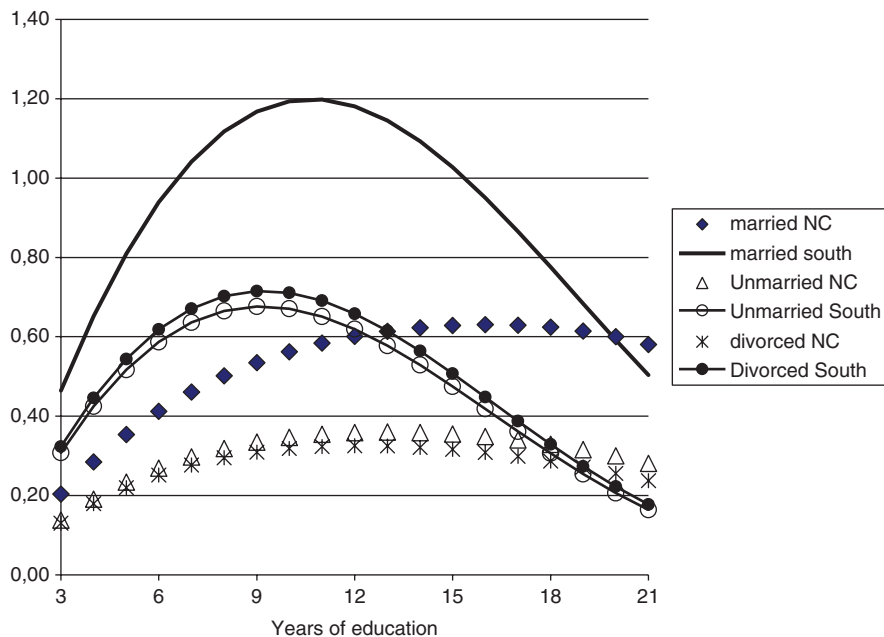


Fig. 7.2 Education-elasticity of Italian women's labour market participation

Acknowledgements We acknowledge financial support from the Italian Ministry of University and Research (PRIN 2008–2010), and useful comments from an anonymous referee.

References

- Angrist, J.D. (2000). Estimation of Limited Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice. *NBER Working Paper* 248.
- Angrist J.D., G. Imbens and D.B. Rubin (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association* 91 (434): 444–455.
- Bank of Italy (2004). *Indagine sui bilanci delle famiglie italiane per l'anno 2002*.
- Becker, G.S. (1981). *Treatise on Family*. Cambridge: Harvard University Press.
- Becker G.S., E.M. Landes and R.T. Michael (1977). An Economic Analysis of Marital Instability. *The Journal of Political Economy* 85 (6): 1141–1188.
- Bedard, K. and O. Deschênes (2003). *Sex Preferences, Marital Dissolution and the Economic Status of Women*. Working Paper, June 2003. Department of Economics, University of California Santa Barbara.
- Bound, J., D.A. Jaeger and R.M. Baker (1995). Problems with Instrumental Variables Estimation when the Correlation between the Instrument and the Endogenous Variable is Weak. *Journal of American Statistical Association* 90: 443–450.
- Bratti, M. (2003). Labour Force Participation and Marital Fertility of Italian Women: The Role of Education. *Journal of Population Economics* 16: 525–554.
- Browning, M. (1992). Children and Household Economic Behaviour. *Journal of Economic Literature* 30 (3): 1434–1475.
- Cigno, A. (1991). *Economics of the Family*. Oxford: Clarendon Press.
- Dankmeyer, B. (1996). Long Run Opportunity-Cost of Children According to Education of the Mother in the Netherlands. *Journal of Population Economics* 9: 349–361.
- Di Pino, A. (2004). On the Economic Estimation of the Time Devoted to Household Chores and Childcare in Italy. *Genus* 60(1): 139–160.
- Fortin, B. and G. Lacroix (1997). A Test of the Unitary and Collective Models of Household Labour Supply. *The Economic Journal* 107: 933–955.
- Goldfeld, S.M. and R.E. Quandt (1973). The Estimation of Structural Shifts by Switching Regression. *Annals of Economic and Social Measurement* 2: 475–485.
- Heckman, J.J. (1974). Shadow Prices, Market Wages, and Labour Supply. *Econometrica* 42: 679–694.
- Heckman, J.J. (1979). Sample Selection Bias as a Specification Error. *Econometrica* 47: 153–161.
- Heckman, J.J. and T.E. MaCurdy (1976). Labor Econometrics. In: *Handbook of Econometrics*, Vol.3, eds. Z. Griliches and M.D. Intriligator. Amsterdam: North Holland.
- Hotz, V.J. and R.A. Miller (1988). An Empirical Analysis of Life Cycle Fertility and Female Labor Supply. *Econometrica* 56(1): 91–118.
- Imbens, G. and J.D. Angrist (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica* 62(2): 467–475.
- Joshi, H. (1990). The Cash Opportunity Costs of Childbearing: An Approach to Estimation Using British Data. *Population Studies* 44 (1): 41–60.
- Lundberg, S. and R.A. Pollak (1993). Separate Sphere Bargaining and the Marriage Market. *The Journal of Political Economy* 101(6): 988–1010.
- Lundberg, S. and R.A. Pollak (2003). Efficiency in Marriage. *Review of Economics of the Household* 1: 153–167.
- Maddala, G.S. (1983). *Limited-dependent and Qualitative Variables in Econometrics*. Cambridge: Harvard University Press.

- Maddala, G.S. and F.D. Nelson (1975). Switching Regression Models with Exogenous and Endogenous Switching. *Proceedings of the American Statistical Association, Business and Economic Statistics Section*: 423–426.
- Mincer, J. (1963). Labor Force Participation of Married Women: A Study of Labor Supply. In: *Aspects of Labor Economics*, ed. H.G. Lewis. New Jersey: Princeton University Press.
- Sousa-Poza, A., R. Schmid R. and H. Widmer (2001). The Allocation and Value of Time Assigned to Housework and Child-care: An Analysis for Switzerland. *Journal of Population Economics* 14(4): 599–618.
- Staiger, D. and J.H. Stock (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica* 65: 557–586.
- Terza, J.V. (1998). Estimating Count Data Models with Endogenous Switching, Sample Selection and Endogenous Treatment Effects. *Journal of Econometrics* 84(1): 129–154.

Chapter 8

New Estimates on the Effect of Parental Separation on Child Health

Shirley H. Liu and Frank Heiland

8.1 Introduction

While marriage remains the most common foundation of family life in the U.S., the prominence of the traditional process of family formation, namely marriage before having children, is diminishing. Today, more than one-third of all births in the U.S. occur outside of marriage (Martin et al. 2006). Although most unmarried parents are romantically involved when their child is born (Carlson et al. 2004), many separate before their child reaches age three (Osborne and McLanahan 2006). While the consequences of *marital* dissolution on children have been studied extensively,¹ the effect of separation of never-married parents on child wellbeing has rarely been examined. This is mainly due to the lack of large representative surveys that collect detailed information on men who father children born out of wedlock.² If the characteristics of the parents and their relationship that determine the risk of union dissolution also affect child wellbeing, then estimates of the effect of separation on child outcomes that fail to account for these factors may suffer from confounding or “selection bias”.

Even when detailed information on the determinants of child wellbeing is available and can therefore be accounted for, however, conventional regression approaches such as Ordinary Least Squares (OLS) may produce invalid estimates of the effect of

S.H. Liu (✉)

Department of Economics, University of Miami, P.O.Box 248126, Coral Gables, FL, USA
e-mail: s.liu2@miami.edu

¹ See Cherlin (1999) and Liu (2006) for recent surveys of this literature. See Morrison and Ritualo (2000) for evidence on the economic consequences of cohabitation and remarriage for children who experienced parental divorce.

² Finding a representative sample of nonresident fathers has proved extraordinarily difficult. In U.S. nationally representative surveys such as the CPS, NSFH, and SIPP, researchers estimated that more than one fifth and perhaps as many as one-half of nonresident fathers are “missing”, i.e. not identified as fathers (e.g., Cherlin et al. 1983; Garfinkel et al. 1998; Sorenson 1997). The problem is especially pronounced for men who fathered children outside of marriage: More than half appear to be missing. Although longitudinal studies of divorced fathers offer a more complete picture, even these suffer from non-inclusion and non-response bias (Garfinkel et al. 1998).

separation on child wellbeing. Regressions rely on strong functional form assumptions (linearity between the covariates and the outcome of interest). In the present context we expect that children who experienced separation (“treated”) may have very different characteristics or environments than children whose parents remained involved (“untreated”). Not only may the treated children differ in terms of the means of their characteristics and environmental variables from the untreated, but also the distribution of these variables could overlap relatively little across groups (“lack of common support”). In this case the regression will project the outcome of the untreated children outside the observed range to form a comparison (“counterfactual outcome”) for the treated children at common values of the covariates. The concern is that such projections, which are highly sensitive to functional form assumptions, will be invalid.

To measure the effect of relationship dissolution on child wellbeing, ideally researchers would use data from randomized experiments or controlled social experiments where parental separation (the treatment) was randomly assigned. In the absence of such data, one strategy is to only compare outcomes between children who experienced parental separation and otherwise similar children whose parents remained together, thereby minimizing potential bias from confounding factors. The challenge of this matching strategy in practice is to identify those children in the untreated group who can serve as good comparisons to the children in the treatment group, i.e. to balance out the children being compared in terms of their characteristics and environmental factors. This approach makes extensive use of the observed characteristics, provides a direct test of whether the observables have common support, and is non-parametric as it does not require assumptions regarding the functional form of the relationship between characteristics and child outcomes.

This study employs a matching strategy to identify whether union dissolution between unmarried parents (defined as the dissolution of a romantic relationship) has a causal effect on child health. We focus on the effect of parental relationship dissolution within three years since childbirth on the child’s likelihood of developing asthma by age three.³ The analysis utilizes data from the Fragile Families and Child Wellbeing Study (FFCWS), which provides detailed information on both biological parents of a large sample of children born out of wedlock. The FFCWS allows us to estimate the separation effect accounting for an unusually large set of characteristics of the child’s parents and their relationship. We present estimates from standard parametric regressions as well as a semi-nonparametric approach based on propensity score matching (Rubin 1979; Rosenbaum and Rubin 1983; Heckman and Hotz 1989; Heckman et al. 1997, 1998). The latter method matches each child whose parents separated with children whose parents remained romantically involved but share similar (observable) characteristics, then compare the outcomes

³ Much of the existing evidence on the effects of family structure and child outcome stems from studies using data on the wellbeing of school-age children and adolescents. We focus on early child outcomes since unmarried families tend to be less stable and hence more short-lived (Bumpass and Lu 2000; Manning et al. 2004), findings from these previous studies may be characteristic of stable unmarried families only.

of these matches. By only using those children that are very similar to children of separated parents to estimate the counterfactual child outcome, the matching method helps us identify the causal relationship between separation and child health. We find that parental separation increases a child's odds of developing asthma by age three by 6% ~ 7%, relative to the situation where their parents *had remained romantically involved*.

8.2 Background

This section provides the conceptual and empirical background for analyzing the effects of separation on child wellbeing, with special emphasis on how separation of the biological parents may harm children born out of wedlock. We draw on the literatures on family formation, dissolution, and resource allocation (e.g. Becker 1973, 1974; Becker et al. 1977; Weiss and Willis 1997; Willis 1999; Ribar 2006), which stress the importance of family resources (time and money) and endowments (care-givers' ability) in the production of family public goods such as child health ("child quality").

8.2.1 Consequences of Separation

Parental separation is expected to lead to a reduction in parental involvement with and resources for the children as benefits associated with growing up in a (parental) union are at best temporarily interrupted and potentially discontinued for a prolonged amount of time.⁴ McLanahan (1985) shows that income explains up to half of the differences in child wellbeing between one- and two-parent families. Unions yield gains from specialization and exchange in the presence of comparative advantages of the partners. Couples may also pool individuals' resources, and realize economies of scale in household production and gains from exploiting risk-sharing opportunities.⁵ Individuals may also be more productive as part of a family due to social learning or other positive externalities.⁶ Lastly, the effective use of monetary transfers from one partner to the other on behalf of the child is more easily monitored within a union (Willis and Haaga 1996; 1999).

⁴ For a detailed discussion of the benefits of a parental union, see Becker (1991), Michael (1973), Shaw (1987), Drewianka (2004).

⁵ Following Becker (1991), the pooling of all resources arises if the dominant decision-maker is altruistic or if the partners have the same objectives. However, if these assumptions are relaxed (McElroy 1990; Manser and Brown 1980; McElroy and Horney 1981), one person's resources cannot be treated as common household income.

⁶ Waite and Gallagher (2000) find some evidence that living together may induce a stabilizing effect on the partners, which can increase resources as a result of greater productivity at home and in the labor market.

8.2.2 Existing Evidence

Parents' economic resources have been shown to be important determinants of child wellbeing (Blau 1999). While caregivers' time and income are substitutable to a certain extent as money can buy childcare services and working in the labor market increases available financial resources, both time and material resources are needed for healthy child development (Coleman 1988). Especially, parenting resources – the services provided by the parents using their time and childrearing ability are believed to be important complements to economics resources (McLanahan and Sandefur 1994).⁷ Studies that compare children across living arrangements have shown that children in single-parent families experience fewer economic and parenting resources (Brown 2002; Hofferth 2001). Single parents may be unable to perform the multiple roles and tasks required for childrearing, which can result in heightened stress levels and insufficient monitoring, demands, and warmth in their parenting practices (Cherlin 1992; Thomson et al. 1994; Wu 1996). Conflicts over visitation may also encumber parenting effectiveness (Brown 2004).

While a large body of research consistently shows a negative correlation between marital dissolution and child outcomes,⁸ until very recently, the relationship between non-marital separation and child wellbeing has received little attention. Heiland and Liu (2006) report that children born to cohabiting or visiting (i.e. romantically involved but living apart) biological parents who end their relationship within a year after birth are up to 9% more likely to have asthma compared to children whose parents stayed together. They also report an increase in child behavioral problems associated with a break-up among children born to romantically involved but not co-residing parents but no effect on mother-reported child health status measures. However, their estimates are obtained from conventional (parametric) models and whether these correlations reflect causal relationships is unclear.

8.2.3 Separation and Selection

A change in the parental relationship towards no (romantic) involvement is expected to decrease the availability of resources and paternal investments in children. However, the environment provided by and the characteristics of parents who separate may differ substantially from parents who remain together. In examining the effect of separation on child outcomes, potential differences in the characteristics of the parents who break up and those who stay together, need to be addressed.

⁷ For example, parental interaction with the child has been found to foster the development of the child by providing support, stimulation, and control (e.g., Maccoby and Martin 1983).

⁸ See Ribar (2006) and Liu and Heiland (2009) for recent surveys of the literature on the effect of marriage on child wellbeing.

Economic theories of relationship dissolution posit that couples break up when the value of the ‘outside opportunity’ of one partner exceeds the benefits from continuing the relationship (Becker et al. 1977; Weiss and Willis 1997). This implies that dissolution does not occur randomly across couples which complicates the identification of the effect of separation on child wellbeing. Simple comparisons of child outcomes by parental relationship status can be misleading if, for example, couples with characteristics that benefit child health are also more likely to break up after childbearing (ceasing a source of positive influence), compared to those who remain together, then the (negative) consequences of separation may be understated (e.g. Steele et al. 2007; Liu 2006). Conversely, if arrangements that induce adverse effects on the child – such as having an abusive father – are more likely to end in a break-up, the association between separation and child wellbeing may even become positive (e.g. Jekielek 1998).

The benefits of father involvement in childrearing are increasingly recognized (see e.g. Lamb 2004). The father’s involvement in the child’s life may depend on the quality of his relationship with the mother. Couples in good relationships tend to communicate more effectively and mothers are more likely to encourage the father’s active involvement in both her and the child’s lives (Carlson et al. 2004). In contrast, when mothers are not able to cooperate with the father and do not perceive that he has the child’s best interests at heart (or are unable to provide for her and their children), they may discourage his involvement and end the romantic relationship. Sigle-Rushton (2005) found that men who fathered children outside of marriage are more likely to come from socioeconomically disadvantaged backgrounds and receive public assistance. Separating from a “deadbeat” dad may reduce the mother’s stress level and allow her to increase available resources for the child through forming new partnerships (e.g. Waller and Swisher 2006).⁹

8.3 Statistical Framework and Estimation Strategy

8.3.1 *Conceptual Model*

Consider a (romantically involved) couple i who has a child out of wedlock. Borrowing from the standard formulation of a selection problem in econometrics, the interrelation of child outcomes, parental investments in children, and relationship status may be formalized as follows:

$$C_i = \beta S_i + \gamma X_i + \varepsilon_i \quad (8.1)$$

$$S_i = \delta X_i + v_i \quad (8.2)$$

⁹ McLanahan and Sandefur (1994) found that children living in stepparent families generally have better outcomes than children in single-parent families.

where C_i denotes the observed child outcome of couple i . S_i is equal to 1 if the couple separates (i.e., dissolve their romantic relationship) and 0 otherwise. The vector X_i includes characteristics of the couple i that affect its willingness and ability to make child investments as well as the risk of relationship dissolution. Unobservables affecting child wellbeing and parental separation are captured by ε_i and ν_i , respectively.

Regression approaches seek to identify the effect of union dissolution on the wellbeing of children, β . Estimates of β based on standard regression methods such as Ordinary Least Squares (OLS) may be biased if S_i and ε_i are statistically dependent. This dependence can arise from two sources: First, couples characteristics (child investments) may be correlated with unmeasured health endowments, i.e. X_i and ε_i are correlated. There may also be bias due to unobservable factors that affect both the child outcomes and the couple's relationship status. In either case, at least part of the observed relationship between child outcomes and the indicator for parental separation is spurious (confounded). The existence of either source of bias would likely show that children of separated parents to have different outcomes from their peers whose parents remained together, independent of any true causal effect of parental separation on child outcomes (selection bias problem).

Selection bias arise in conventional regression analysis as these estimators employ data from all observations to be combined into one estimate of the separation effect. If parents who remain together tend to be very different regarding their child investments compared to couples who separate, then the validity of results from standard regression models is suspect since the combining functions operate over very different families. Specifically, the separation effect is identified by comparing the average outcome of children who experienced a dissolution to those who did not. In the presence of any characteristics that affect the couples' decision to separate as well as child wellbeing, the resulting estimates will reflect both the "true" effect of parental separation on children who experience union dissolution *and* the effects of factors that influence the parents' risk of separation in the first place.

In addition to estimates from conventional regression approaches, this study builds on a non-parametric strategy known as the potential outcome approach to investigate the effect of parental separation on child health. In this approach, the relationship between union dissolution and child outcome is formulated in a framework similar to a social experiment in which the treatment is randomly assigned. Pioneered in the program evaluation literature in economics (see e.g. Lechner 2002; Imbens 2004), the matching approach has been fruitfully employed to study the effect of an event ("treatment") on participant outcomes when participation ("selection into treatment") is expected to be non-random. For instance, when analyzing the effect of a welfare program on individuals, researchers want to know what the outcomes of the participants would have been had they not enroll in the program. Since data on the counterfactual are typically unavailable in observational data, one needs to rely on the behavior of the non-participants in the sample to construct the counterfactual outcome. However, since welfare participation is voluntary, the participation choice is non-random and participants tend to exhibit different characteristics from non-participants. As a result, standard regression estimates of the

effect of the treatment, obtained from comparing participants with non-participants who are systematically different, will be confounded with the effects of selection into participation. The matching method is particularly useful in this situation as it re-establishes the conditions of an experiment, by matching the sample of participants and non-participants with respect to characteristics that rule the selection into program participation (treatment).

In the present context, the “treatment” of interest – parental separation – is defined in terms of the potential outcomes for children whose parents separated. Children whose parents separated are in the treated group, and children whose parents remained romantically involved are defined as the control group (or “untreated”). We want to identify the effect of parental separation on children whose parents separated. To construct the counterfactual, i.e. the outcomes of children whose parents separated had their parents remained romantically involved, we draw on matching methods developed in the statistics literature (Rosenbaum and Rubin 1983; Heckman and Robb 1985) that exploit the full information of the observable characteristics. Unlike regression approaches, these methods balance out the groups being compared in terms of their covariates and do not require assumptions regarding the functional form of the relationship between family characteristics and child outcomes. Specifically, they provide systematic ways to construct a sample counterpart for the missing information on the counterfactual outcomes of the treated children by pairing treated and control children who share similar observable characteristics. Our application of propensity score matching to the study of parental separation on child health is novel and adds to the growing number of areas within population studies that have benefited from this technique (see Sigle-Rushton 2005; Liu and Heiland 2009, and the related chapters in this book for additional applications).

We note that the methodology adopted here addresses selection on observable factors and does not readily extend to selection on unobservables. If unobservable factors are proxied for by X_i then matching based on observables also reduces selection bias generated by unobserved factors. The extent to which the treatment bias is reduced will thus crucially depend on the richness and quality of the control variables, X_i , that are used to match treated and control observations. Typically, the information about the parents of out-of-wedlock children and their relationship is limited in large representative survey datasets. Fortunately, the FFCWS contains detailed information on the child as well as both biological parents and their romantic involvement, allowing us to capture factors believed to be important determinants of the separation risk including the degree to which the parents are assortatively matched.¹⁰

¹⁰ Approaches that seek to address selection bias due to unobservables directly include treatment effects estimators and instrumental variables estimators. The former essentially model the selection process directly and require strong distributional assumptions. In the context of divorce and child outcomes, variation in state and local divorce policy and costs have been used as instruments for divorce. However, to what extent these types of events can serve as valid instruments has been debated (see Steele et al. 2007; Liu 2006) and finding a suitable instrument for union dissolution among unmarried couples promises to be even more challenging.

8.3.2 Potential Outcome Approach

Consider the “treatment” to be the separation (i.e. romantic relationship dissolution) between the biological parents of child i : $S_i = 1$ denotes the “treatment group” (i.e. children whose parents separate), and $S_i = 0$ denotes the “control group” (i.e. children whose parents remain romantically involved). Let $C_i(1)$ denote the potential outcome of child i under the treatment state “parents separated” ($S_i = 1$), and $C_i(0)$ the potential outcome if the same child receives no treatment, “parents remained romantically involved” ($S_i = 0$). Thus, $C_i = S_i C_i(1) + (1 - S_i) C_i(0)$ is the observed outcome of child i . The individual treatment effect is $\beta_i = C_i(1) - C_i(0)$, which is unobserved since either $C_i(1)$ or $C_i(0)$ is missing.¹¹

Ordinary least squares estimates the average treatment effect (ATE) by taking the average outcome difference between the treated and control groups: $\beta_{OLS} = E[C_i(1) | S_i = 1] - E[C_i(0) | S_i = 0]$. The ATE is the average of the treatment effect on the treated *and* the treatment effect on the controls. Given that many children whose parents remained involved may never be at risk of parental separation, the ATE may not be particularly illuminating when our interest lies in how parental separation has affected children whose parents did separate. Hence, alternatively, one might focus on the average effect of treatment on the treated only (“effect of parents’ separation on children whose parents separate”), i.e. the ATET henceforth:

$$\beta_{S_i=1} = E[\beta_i | S_i = 1] = E[C_i(1) | S_i = 1] - E[C_i(0) | S_i = 1] \quad (8.3)$$

which is the difference between the expected outcome of a child whose parents separate, and the expected outcome of the same child if his/her parents had remained romantically involved. While we do observe the outcomes of children whose parents separate, and are thus able to construct the first expectation $E[C_i(1) | S_i = 1]$, we cannot identify the counterfactual expectation $E[C_i(0) | S_i = 1]$ without invoking further assumptions. To overcome this problem, one has to rely on children whose parents remained romantically involved to obtain information on the counterfactual outcome. Since treatment status is likely non-random, replacing $E[C_i(0) | S_i = 1]$ with $E[C_i(0) | S_i = 0]$ is inappropriate since the treated and untreated might differ in their characteristics determining the outcome.

An ideal randomized experiment would solve this problem because random assignment of couples to treatment ensures that potential outcomes are independent of treatment status;¹² and if such data exist, conventional regression methods would

¹¹ The individual treatment effect is equivalent to taking the difference between the outcome of child i if his/her parents separated, and the outcome of the same child if his/her parents remained together. Since for any given child, his/her parents can only be observed as either “separated” or “remained involved”, we can never observe the outcomes of a given child in *both* of these situations.

¹² Randomization implies that $S_i \perp (C_i(0), C_i(1))$ and therefore: $E[C_i(0) | S_i = 1] = E[C_i(0) | S_i = 0] = E[C_i(0)]$.

produce an unbiased estimate of β . However, this would require that couples who share similar characteristics are randomly assigned to separate or remain involved, which would be infeasible for obvious practical and ethical reasons. In this non-experimental setting, the couple's relationship status is likely non-random and depends on characteristics that may also influence the couple's child investment behavior. For instance, the couples' economic conditions can influence both their relationship stability and ability to care for their children. In what follows, the approach used to construct a suitable comparison group when random assignment is unavailable, namely the matching method, and the identifying assumptions on which it is based, are described.

8.3.3 Matching

Statistical matching is a way to identify a suitable control group that is comparable to the treated. This method is particularly useful in settings where data often do not come from randomized trials, but from (non-randomized) observational studies. Matching estimators try to re-establish the condition of an experiment by stratifying the sample of treated and untreated children with respect to covariates X that rule the selection into treatment. Selection bias is eliminated provided all variables in X are measured and comparable (or "balanced") between the two groups. In this case, outcome differences between the treated and controls provide an unbiased estimate of the treatment effect.

8.3.3.1 Conditional Independence Assumption (CIA)

The matching method pairs treated and control units with similar observable characteristics and assume that their relevant differences, in terms of potential outcomes, are captured in their observable attributes. This underlying assumption, called the conditional independence assumption (CIA henceforth), requires that conditional on observables X_i , the distribution of the counterfactual outcome $C_i(0)$ in the treated group is the same as the (observed) distribution of $C_i(0)$ in the non-treated group. In other words, the outcomes of the untreated are independent of participation into treatment S_i , conditional on observable characteristics $X_i : C_i(0) \perp S_i | X_i$. This rules out the possibility that variables not included in X_i , on which we cannot condition, affect both $C_i(0)$ and S_i (i.e., there is no selection on unobservables). It follows that, for a child whose parents separated with a given x , the outcomes of matched children whose parents remained romantically involved can be used to measure what his/her outcome *would have been*, on average, had his/her parents remained romantically involved. This assumes that there are untreated individuals for each $x : Pr(S_i = 0 | X_i = x) > 0$ for all x , implying that individuals are matched only over the common support region of X_i where the treated and untreated group overlap. Note that under the CIA, it is not necessary to make assumptions regarding the

functional forms of the outcome equations, decision processes, or distribution of the unobservables.¹³

8.3.3.2 Average Treatment Effect for the Treated (ATET)

Following the CIA, the average treatment effect on the treated can be computed as follows:

$$\begin{aligned}
 \beta_{|S_i=1} &= E [C_i (1) | S_i = 1] - E [C_i (0) | S_i = 1] & (8.4) \\
 &= E_X [E [C_i (1) | X_i, S_i = 1] - E [C_i (0) | X_i, S_i = 1] | S_i = 1] \\
 &= E_X [E [C_i (1) | X_i, S_i = 1] - E [C_i (0) | X_i, S_i = 0] | S_i = 1] \\
 &= E_X [E [C_i | X_i, S_i = 1] - E [C_i | X_i, S_i = 0] | S_i = 1]
 \end{aligned}$$

To estimate the ATET, one is to first take the outcome difference between the two treatment groups conditional on X_i , then average over the distribution of the observables in the treated population.¹⁴

Conditioning on X within a finite sample, however, can be problematic if the vector of observables is of high dimension. The number of matching cells increases exponentially as the number of covariates in X_i increases. Thus, it is possible that there will be some cells that contain only treated or untreated units, but not both, making the comparison impossible. Rubin (1979) and Rosenbaum and Rubin (1983) suggest the use of the *propensity score*, the conditional probability of selection into treatment: $p(X_i) = \Pr(S_i = 1 | X_i = x) = E(S_i | X_i)$, to stratify the sample. In the present context, the propensity score is simply the conditional probability that the parents of a given child would separate. They showed that by definition the treated and the non-treated with the same propensity score have the same distribution of X : $X_i \perp S_i(X_i)$. This is called the *balancing property* of the propensity score. Furthermore, if $C_i(0) \perp S_i | X_i$, then $C_i(0) \perp S_i(X_i) | p(X_i)$. This implies that matching can be performed on $p(X_i)$ alone, which is more parsimonious than the full set of interactions needed to match treated and untreated on the basis of observables, thus reducing the dimensionality problem into a single variable.

Matching treated and untreated with the same propensity scores and placing them into one cell (i.e., observations with propensity scores falling within a specific range) is as if the selection into treatment is random within each cell and the probability of participation within this cell equals the propensity score. Consequently, the difference between the treated and the untreated average outcomes at any value of $p(X_i)$

¹³ The CIA assumption is strong because it is based on the assumption that the conditioning variables in X_i be sufficiently rich to justify the application of matching. In particular, CIA requires that the set of X_i should contain all the variables that jointly influence the outcome without treatment $C_i(0)$ as well as selection into treatment S_i (Heckman et al. 1998). To justify this assumption, econometricians implicitly make conjectures about what variables enter in the decision set of couples, and unobserved relevant variables are related to observables.

¹⁴ The regression equivalent of this procedure requires the inclusion of all the possible interactions between the observables X_i .

is an unbiased estimate of the ATET at that value of $p(X_i)$. Therefore, an unbiased estimate of the ATET can be obtained by conditioning on $p(X_i)$:

$$\beta_{|S_i=1} = E_{p(X)} [(E(C_i | S_i = 1, p(X_i)) - E(C_i | S_i = 0, p(X_i))) | S_i = 1] \quad (8.5)$$

The implementation of this framework has several challenges. First, the propensity score itself needs to be estimated.¹⁵ Second, since it is a continuous variable, the probability of finding an exact match for each treated child is theoretically zero. Therefore, a certain distance between the treated and untreated has to be accepted.

8.3.4 Matching Estimators

Various methods exist to implement matching estimates, all are based on the same strategy of pairing individuals but with different weighting schemes given to counterfactual individuals. Let T and C be the set of treated and untreated individuals, respectively. The observed outcome of a treated individual be denoted Y_i^T , and Y_j^C denotes the observed outcome of an individual in the control group. Let $C(i)$ be the set of control individuals matched to the treated individual i with an estimated propensity score p_i .

In general, *Kernel matching* matches all treated observations with a weighted average of all control observations with weights that are inversely proportional to the distance between the propensity scores of treated and controls. The *kernel matching estimator* is given by:

$$\tau^k = (1 / N^T) \sum_{i \in T} \left[Y_i^T - \left[\left(\sum_{j \in C} Y_j^C K((p_j - p_i) / h_n) \right) / \left(\sum_{k \in C} Y_k^C K((p_k - p_i) / h_n) \right) \right] \right]$$

where $K(\cdot)$ is a kernel function and h_n is a bandwidth parameter. In this study, we consider three matching estimators, namely *Uniform* (also known as the “radius” matching estimator), *Epanechnikov*, and *Gaussian* kernels, each uses a specific kernel function:

Epanechnikov: $K(u) = (3 / 4)(1 - u)^2$ for $|u| < 1$, and 0 otherwise

Gaussian: $K(u) = (1 / \sqrt{2\pi}) \exp[-u^2 / 2]$ for all u

Uniform (Radius): $K(u) = 1 / 2$ for $|u| < 1$ and 0 otherwise

¹⁵ The propensity score, i.e., the conditional probability that the parents of a given child would separate, can be estimated using any standard probability model. For example, $\Pr(S_i | X_i) = F(h(X_i))$, where $F(\cdot)$ is the normal or the logistic cumulative distribution and $h(X_i)$ is a function of covariates with linear and higher ordered terms. See Dehejia and Wahba (1999) for a description of the algorithm used to estimate the propensity score.

Under the standard conditions on the bandwidth and kernel,

$$\sum_{j \in C} Y_j^C K((p_j - p_i) / h_n) / \sum_{k \in C} Y_k^C K((p_k - p_i) / h_n)$$

is a consistent estimator of the counterfactual outcome Y_{0i} .

The main difference between these matching estimators is in how weights are assigned to the matches. In *radius* matching, each treated unit is matched only with control units whose propensity score falls within a predefined neighborhood (i.e., radius) from its propensity score. All matches within this radius are assigned the same weight. If the dimension of the neighborhood (i.e., radius) is defined to be very small, it is possible that some treated units are not matched because the neighborhood does not contain any control units. Conversely, the smaller the size of the neighborhood the better the quality of the matches. With *Gaussian* and *Epanechnikov* kernel matching, all treated are matched with a weighted average of all controls, with the Gaussian kernel assigning weights that follow a normal distribution, and the Epanechnikov kernel assigning weights that follow a triangular distribution.¹⁶

Estimation using propensity score matching is now available via a set of Stata programs using the *pscore* package. Details of the algorithms used can be found in Becker and Ichino (2002). There are tradeoffs between the quantity and quality of the matches among these estimators but none is a priori superior. Relative to radius matching, the Gaussian and Epanechnikov matching tend to produce higher quantity of matches; however, the quality of the matches may be poorer since treated units are potentially matched with distant controls. Nevertheless, their joint consideration offers a way to assess the robustness of our results.

8.4 Data, Sample, and Descriptive Evidence

Our data are drawn from the Fragile Families and Child Wellbeing Study (FFCWS), which follows a cohort of 4,898 children and both of their biological parents in 20 U.S. cities from birth (1998 ~ 2000), at age one, and again when the child is about three years old.¹⁷ The FFCWS is unique as it includes a large set of children born to unmarried parents. Areas such as parent-parent and parent-child relationships, socioeconomic activities, and child development are covered.

¹⁶ Depending on the choice of the bandwidth, the Gaussian kernel assigns positive weights to potentially poor matches (matches in which distance between the treated and controls are very far), while the Epanechnikov kernel assigns no weight to some potentially bad matches.

¹⁷ See Reichman et al. (2001) for a detailed description of the study design and sampling methods.

8.4.1 Sample Selection

Our study sample consists of 1,419 children all born to parents who were unmarried but romantically involved at childbirth. The sample is selected in the following manner. First, given that the relationship arrangement between the biological parents is crucial for our study question, we exclude children whose parents' relationship status at either the one- or three-year follow-ups cannot be identified ($n = 1,733$ are dropped). Second, we focus on children born to unmarried biological parents who were romantically involved at childbirth (i.e. either in cohabiting or visiting unions), therefore children whose parents were either married (944 cases) or not romantically involved (302 cases) at childbirth are excluded. Third, we exclude children for whom we do not observe the outcome measure, i.e. whether they have developed asthma by age three (406 cases). Fourth, the parents of 32 of the remaining children had been married within the first year after childbirth, but divorced before their child reached age three. To avoid confounding the effect of separation between never-married parents and parental divorce, these observations are dropped.¹⁸ Fifth, we cross check the marriage date (available since the one-year follow-up) with parents' reported marital status at childbirth. Observations in which the reported marriage date contradicts the reported marital status of the parents at childbirth are dropped (9 observations). An additional 32 observations are dropped due to missing information on important socioeconomic and demographic characteristics.¹⁹ In the resulting sample, consisting of 1,434 children all born to unmarried parents, 37% of the parents have ended their (romantic) relationship by the time their child reaches age three.

Finally, we estimate the propensity score of selection into treatment (i.e. the probability of parental separation within three years since childbirth) within this sample of 1,434 children. To ensure sufficient overlap of the propensity scores between the treatment and control groups, observations with propensity scores falling outside of the common support region are excluded from the analysis (7 treated and 8 controls), resulting in the final sample size of 1,419 children.²⁰ Table 8.1 presents summary statistics of the measures employed in this study. Sample means are presented for the full sample (Columns 2 and 3) and by treatment status (Columns 4 and 5).

¹⁸ We note that our results are robust to the inclusion of these observations (results available upon request).

¹⁹ To ensure that exclusion of these observations does not result in a selected sample (i.e. if the tendency of under-reporting is correlated with the treatment), we constructed missing indicators for each of these covariates and conducted t-tests of means for each of the missing indicators between the treated and control groups. None of the t-tests showed significant differences in the prevalence of under-reporting across the two groups (results available upon request).

²⁰ Imposing the "common support" restriction implies that the test of the balancing property is performed only on the observations whose propensity score belongs to the intersection of the supports of the propensity score of treated and controls. Imposing the common support condition in the estimation of the propensity score may improve the quality of the matches used to estimate ATET.

Table 8.1 Sample means by relationship status three years after an out-of-wedlock birth

	Entire sample		Parents' relationship status (3 years after childbirth)	
	Mean	[S.D.]	Involved	Separated
Child developed asthma by age 3	0.249	[0.433]	0.221	0.298*
Parents separated by age 3	0.371	[0.483]		
Parents' relationship at childbirth				
Cohabiting	0.654	[0.476]	0.765	0.466*
Visiting	0.346	[0.476]	0.235	0.534*
Child characteristics				
Child is of low birth weight (< 88 oz)	0.107	[0.309]	0.108	0.105
Child is female	0.464	[0.499]	0.479	0.437
Child's birth order (mother):				
– 1 st	0.376	[0.485]	0.353	0.416*
– 2 nd	0.330	[0.470]	0.336	0.319
– 3 rd or higher	0.294	[0.456]	0.311	0.264 ⁺
Parent's demographic characteristics				
Mother's age < 20 at childbirth	0.228	[0.419]	0.197	0.279*
Father's age < 20 at childbirth	0.111	[0.315]	0.089	0.149*
Father is younger than mother	0.195	[0.380]	0.207	0.175
Mother's race/ethnicity:				
– white	0.165	[0.371]	0.185	0.129*
– black	0.523	[0.500]	0.456	0.635*
– Hispanic	0.285	[0.452]	0.331	0.207*
– other	0.028	[0.164]	0.027	0.029
Father's race/ethnicity:				
– white	0.126	[0.332]	0.144	0.095*
– black	0.557	[0.497]	0.495	0.662*
– Hispanic	0.285	[0.452]	0.328	0.213*
– other	0.032	[0.175]	0.032	0.030
Mother and father of different race/ethnicity	0.145	[0.353]	0.143	0.150
Mother is foreign-born	0.111	[0.315]	0.147	0.051*
Father is foreign-born	0.218	[0.413]	0.209	0.234
Child's household income				
Income less than \$10,000	0.202	[0.402]	0.163	0.269*
Income between \$10,000 and \$24,999	0.340	[0.474]	0.352	0.319
Income at least \$25,000	0.458	[0.498]	0.484	0.412*
Parents' education				
Mother' education:				
– less than H.S. diploma	0.367	[0.482]	0.364	0.373
– high school diploma/GED	0.356	[0.479]	0.345	0.375
– some college	0.247	[0.432]	0.285	0.230
– bachelor & beyond	0.030	[0.170]	0.034	0.023
Father's education:				
– less than H.S. diploma	0.357	[0.484]	0.386	0.357
– high school diploma/GED	0.383	[0.486]	0.354	0.434*
– some college	0.213	[0.410]	0.229	0.186 ⁺
– bachelor & beyond	0.028	[0.166]	0.032	0.023

Table 8.1 (Continued)

	Entire sample		Parents' relationship status (3 years after childbirth)	
	Mean	[S.D.]	Involved	Separated
Father is less educated than mother	0.271	[0.445]	0.279	0.257
Parents' labor market activities				
Mother works	0.190	[0.393]	0.199	0.175
Mother's weekly hours of work	35.75	[9.199]	36.08	35.10
Mother's annual labor income:				
– less than \$10,000	0.423	[0.495]	0.417	0.432
– between \$10,000 and \$24,999	0.432	[0.496]	0.424	0.444
– at least \$25,000	0.145	[0.353]	0.158	0.123
Father works	0.839	[0.368]	0.862	0.798*
Father's weekly hours of work	43.71	[11.52]	44.11	42.88
Father's annual labor income:				
– less than \$10,000	0.280	[0.449]	0.264	0.315+
– between \$10,000 and \$24,999	0.473	[0.500]	0.466	0.486
– at least \$25,000	0.247	[0.431]	0.270	0.199*
Mother's labor income > father's	0.121	[0.328]	0.145	0.071
Other characteristics				
Mother is catholic	0.281	[0.450]	0.326	0.204*
Mother reports no religious affiliation	0.128	[0.334]	0.123	0.137
Mother attends religious activities frequently	0.166	[0.372]	0.165	0.169
Parents' have known each other for < 1 year before pregnancy	0.245	[0.430]	0.236	0.260
Father suggested abortion during pregnancy	0.152	[0.359]	0.137	0.177*
Prenatal smoking and/or drinking (mother)	0.268	[0.443]	0.263	0.278
Maternal grandmother's education (> HS)	0.216	[0.412]	0.218	0.213
Mother's PPVT score (Year 3)	88.11	[11.15]	88.58	87.39+
<i>N</i>	1,419		893	526

Notes: Sample means between "children whose parents remained romantically involved" and "children whose parents separated" by age 3 is statistically significantly different at the * = 5% level, + = 10% level.

8.4.2 Measure of Child Health

Child health is measured by a child's likelihood of developing asthma by age three. Asthma is the most common chronic illness affecting children,²¹ with symptoms formulated since infancy (Klinnert et al. 2001). Genetic predispositions combined with exposure to environmental toxins are common risk factors for asthma onset (Weisch et al. 1999; Sporik et al. 1991; Cogswell et al. 1987; Weitzman et al. 1990). In the U.S., children from lower socioeconomic and minority backgrounds develop higher rates of asthma, a pattern attributable to toxic environmental exposures and poor health investments (Neidell 2004; Gergen et al. 2006; Oliveti et al. 1996).

²¹ "Asthma in Children Fact Sheet", American Lung Association 2004.

Psychological stress is also known to aggravate asthma, and the relationship between stressful life events and the onset of asthma has been well established among the adult population (Teiramaa 1979; Levitan 1985; Kilpeläinen et al. 2002). Recent research also points to stress experienced by a caretaker as an independent factor contributing to child asthma (Wright et al. 2002).²² Stressful life events, such as parental relationship conflicts, have been found to be associated with asthma onset in infants, mainly through the mother's coping abilities that translate into her parenting behavior (Klennert et al. 1994).

In the FFCWS, mothers are asked to report whether her child has asthma or asthma attacks (or were informed by a health care professional that the child has asthma)²³ by age one, and again by age three. Within our sample, 25% report having asthma or an asthma attack by age three.²⁴ The incidence of asthma differs markedly by treatment status: a significantly higher proportion of children whose parents separated by age three reports having asthma (30%), relative to children whose parents remained romantically involved (22%).

8.4.3 Who Gets Separated?

While a number of recent studies examine the determinants of marriage among unmarried parents (e.g. Carlson et al. 2004; Goldstein and Harknett 1988), the factors contributing to the dissolution of these unions have received little attention (see Liu and Heiland 2009). Relationships that dissolve within three years after childbirth were potentially less stable at the onset. Parents in visiting relationships at the time of childbirth are more likely than cohabiting parents to separate within three years after a premarital birth: 26% of cohabiting parents as opposed to 57% of visiting parents end their romantic ties within three years after childbirth (not shown). Children whose parents separate are more likely the result of unplanned pregnancies,

²² Wright et al. studied the role of caregiver stress on infant asthma. Using a birth cohort with family histories of asthma to account for genetic predisposition, they find that greater stress levels experienced by caregivers when the child is 2 to 3 months old (before any symptoms of asthma can be detected) is associated with increased risk of recurrent episodes of wheezing (clinical definition of asthma) in children during the first 14 months of life. The findings are robust to established controls and potential mediators (including socioeconomic status, birth weight, race/ethnicity, maternal smoking, breast-feeding, indoor allergen exposure, and lower respiratory infections). In addition, the direction of causality runs from caregiver stress to levels of infant wheezing, rather than the reverse.

²³ This is consistent with the standard definition of childhood asthma, which is measured based on the response of a parent or adult household member ("America's Children: Key National Indicators of Well-Being, 2001," Federal Interagency Forum on Child and Family Statistics, Washington D.C.: U.S. Printing Office).

²⁴ According to the 2002 National Health Interview Survey, about 12% of U.S. children under the age of 18 are diagnosed with asthma, but the incidence is much higher among minority children (CDC 2004, <http://www.cdc.gov/asthma/children.htm>). Diagnosing asthma in young children is more difficult than in older children, but an estimated 50% of kids with asthma develop symptoms by age two.

as indicated by the greater percentage of fathers who suggested abortion during the pregnancy. Having an unplanned pregnancy can strain a romantic relationship, as it has been found to be associated with less positive interactions between spouses (Cox et al. 1989).

Studies of married couples have found that husbands' socioeconomic characteristics to be positively correlated with marital stability, but not the wife's (e.g., Whyte 1990). One of the most important barriers to a stable relationship is financial instability, as a father that cannot contribute to the economic wellbeing of the family is seen as a liability (Edin 2000). Consistent with this argument, we find that fathers who separate from the child's mother tend to be younger, foreign-born, less educated, and less attached to the labor force, relative to fathers who remain romantically involved with their child's mother. Low levels of education and poverty are linked to risky and abusive behavior (e.g. Clark et al. 2004). Unmarried non-resident fathers have been found to exhibit these risk factors at higher rates than married or cohabiting fathers (Wilson and Brooks-Gunn 2001; Jaffee et al. 2001). These risk factors may lead to lower father involvement with children both directly, or indirectly by weakening his relationship with the mother. Mothers may further mediate father involvement with the child even after their romantic relationship with the father has ended (Fagan and Barnett 2003).

8.5 Estimation Results

Our descriptive evidence points to a negative association between parental separation and child's likelihood of developing asthma. However, one cannot readily conclude that this association is causal, as there may be factors that influence both the child outcomes and parental separation. Ideally, to determine whether this association is causal, we would have information on the potential outcomes of these children *if their parents had remained romantically involved*. Since the counterfactual outcome is never directly observed, and standard regression estimates based on the average outcomes of all control observations (many of whom may differ systematically from the treated) are potentially biased, an alternative statistical method to identify the counterfactual is needed. Matching methods is a semi-parametric method that can be used to reduce selection bias, by constructing a suitable control group whose outcomes are more likely to resemble the counterfactual outcomes of children whose parents separated if they had remained together.

In this setting, children who experience parental separation are compared only to children whose parents remain romantically involved but share very similar (environmental) characteristics, and not to children subjected to very different conditions in addition to their treatment status. Hence, the estimated effect of parental separation is the average of the typical effect of treatment on the treated only, and the differences in their outcomes are taken as driven only by their treatment status (i.e. the "causal" effect of parental separation on children whose parents separated).

8.5.1 The Propensity Score of Parental Relationship Dissolution

The first step in implementing the matching method is to estimate the propensity score for the treatment (“parental separation”) under study: $\Pr[S_i = 1 | X_i]$. Parents’ propensity to separate is defined as a function of each parent’s socioeconomic and demographic characteristics, child-specific characteristics observed at childbirth, and measures of union match quality. Parameter estimates for the probit model used to match the treated and control groups of children are presented in Table 8.2. Consistent with our descriptive evidence (holding everything else constant), parents who did not co-reside at the time of childbirth (“visiting relationships”) are significantly more likely to dissolve their romantic relationship within three years after childbirth. Unmarried fathers who are young (less than 20 years of age), foreign-born, poorly educated, and work few hours per week are significantly more likely to see their romantic relationship with the child’s mother end within three years since childbirth.

Once the propensity score is estimated, we need to make sure that the treated and controls are (statistically) identical in terms of their observable characteristics X and their estimated propensity scores, but differ only in terms of their treatment status (“test of the balancing property”). The sample is stratified into 5 equally spaced intervals (or blocks) based on the predicted propensity score. We test (1) whether the average propensity scores and means of each covariate in X are (statistically) identical between the treated and control units within each interval, and (2) there is sufficient overlap of the propensity scores between the treated and controls within each interval, to ensure that adequate number of matches can be found for the treated units.²⁵ Table 8.3 reports results of the test of the balancing property between the treated and controls. The test shows that the treated and controls are comparable in their observable characteristics within each interval. In addition, Fig. 8.1 reveals that there is sufficient overlap of the propensity scores between the treated and controls in each block.

8.5.2 Main Findings

Table 8.4 presents the estimated effect of parental separation on child’s propensity to develop asthma by age 3. We first report the OLS estimates: column 2 shows the unadjusted mean differences in the prevalence of child asthma between the treated and controls (i.e. OLS regression without any controls), and column 3 reports the mean outcome difference after adjusting for a full set of controls. The propensity score matching estimates based on the Gaussian, Epanechnikov, and uniform kernel (radius) estimators, respectively, are reported in columns 4–8. To assess the sensitivity of the matching estimates to the choice of bandwidth (or radius), we also report

²⁵ For details of this test, see Dehejia and Wahba (1999).

Table 8.2 Probit estimates of the propensity score

	Coefficient	Robust Standard Error	P > z
Child is of low birth weight (< 88 oz)	-0.034	0.120	[0.780]
Child is female	-0.080	0.073	[0.278]
Child's birth order (mother):			
- (Ref: 1st)			
- 2nd	-0.114	0.092	[0.214]
- 3rd or higher	-0.170	0.104	[0.101]
Mother's age < 20	0.048	0.107	[0.652]
Father's age < 20	0.227	0.134	[0.091]
Father is younger than mother	-0.059	0.103	[0.565]
Parents' race/ethnicity:			
- (Ref: both black)			
- both white	-0.274	0.144	[0.057]
- both Hispanic	-0.122	0.150	[0.413]
- both other	0.312	0.397	[0.432]
- mother is white, father is non-white	-0.002	0.198	[0.992]
- mother is black, father is non-black	0.213	0.224	[0.343]
- mother is Hispanic, father is non-Hispanic	0.074	0.203	[0.717]
- mother is other, father is non-other	-0.218	0.465	[0.639]
Parents' region of birth:			
- (Ref: both U.S.)			
- mother is foreign-born, father is not	-0.403	0.278	[0.147]
- father is foreign-born, mother is not	0.308	0.122	[0.011]
- both parents are foreign-born	-0.318	0.183	[0.081]
Mother's education:			
- (Ref: less than HS)			
- H.S. diploma/GED)	-0.059	0.156	[0.703]
- some college	-0.146	0.255	[0.567]
- bachelor & beyond	-0.440	0.424	[0.299]
Father's education:			
- (Ref: less than HS)			
- H.S. diploma/GED	0.250	0.150	[0.095]
- some college	0.174	0.251	[0.488]
- bachelor & beyond	0.344	0.422	[0.415]
Father's education relative to mother's:			
- (Ref: same)			
- less	0.061	0.174	[0.725]
- more	-0.131	0.169	[0.439]
Child's household income:			
- (Ref: less than \$10,000)			
- between \$10,000 and \$24,999	-0.153	0.112	[0.172]
- at least \$25,000	-0.092	0.117	[0.428]
Parents' labor force participation:			
- (Ref: neither parents work)			
- both parents work	-0.092	0.361	[0.800]
- only mother works	0.358	0.423	[0.397]
- only father works	0.219	0.137	[0.109]

Table 8.2 (Continued)

	Coefficient	Robust Standard Error	P > z
Mother's weekly hours of work	0.007	0.009	[0.450]
Father's weekly hours of work	-0.005	0.002	[0.034]
Mother's labor income exceeds father's	-0.538	0.336	[0.110]
Length of parents' relationship before pregnancy:			
- (Ref: more than 2 years)			
- less than 6 months	0.030	0.123	[0.807]
- 6 months to 1 year	0.173	0.112	[0.120]
- 1 to 2 years	0.029	0.095	[0.762]
Mother is catholic	-0.078	0.113	[0.490]
Mother has no religious affiliation	0.031	0.114	[0.786]
Mother attends religious activities frequently	-0.003	0.102	[0.978]
Father suggested abortion during pregnancy	-0.007	0.101	[0.946]
Maternal grandmother attained more than a high school education	-0.055	0.099	[0.576]
Prenatal smoking/drinking (mother)	0.105	0.089	[0.242]
Parents in visiting relationship at childbirth	0.604	0.085	[0.000]
Mother's PPVT score (Year 3)	-0.000	0.004	[0.965]
Constant	-0.570	0.441	[0.196]

Log Likelihood = -821.31

Pseudo R² = 0.132

Notes: 1. Additional controls for "mother's state of residence at childbirth" (14 state dummies) omitted here. 2. Region of Common Support $\in [0.05292221, 0.83660801]$.

results using different bandwidths (or radiuses). Details on the choice of bandwidth are discussed in the next section.

On average, children whose parents separate are 7.8% more likely to develop asthma by age 3 compared to children whose parents remain romantically involved. Differences in observable parental and child characteristics partially explain the outcome difference between the treated and controls: the separation effect is reduced to 5.2% (OLS) or 6.1% ~ 7.1% (matching) but remains statistically significant. This finding suggests that selection into relationship separation helps explain the child outcome differences between children whose parents separate and those who do not. A notable share of unmarried fathers have disadvantaged characteristics that may not be conducive to increase engagement (or sustain romantic involvement), hence their relationship with the child's mother may have been less stable (or sustainable) from the onset. Hence, these factors may help explain the poorer health among out-of-wedlock children whose parents separate.

Recall that the OLS estimates the average treatment effect (ATE) and matching estimates the average treatment effect on the treated only (ATET). While our matching estimates confirm the direction of the separation effect suggested by the

Table 8.3 Test of balancing properties between the control and treatment group (Two-sample T-Test of means): T-statistics reported

	Block 1	Block 2	Block 3	Block 4	Block 5
Range of the propensity score	[0.053, 0.200]	[0.200, 0.400]	[0.400, 0.600]	[0.600, 0.800]	[0.800, 0.837]
N Treated	37	166	175	133	15
N Controls	264	392	169	62	6
Two-Sample Test of Means: Significance Level = 0.01					
 T Statistic					
Propensity score	1.314	2.432	2.136	1.116	0.005
Child is of low birth weight (< 88 oz)	0.592	1.236	0.778	0.323	0.679
Child is female	0.105	1.006	0.150	0.897	0.400
Child birth order (mother):					
– (Ref: 1st)					
– 2nd	0.640	0.660	1.185	2.102	1.405
– 3rd or higher	1.173	0.751	0.308	0.226	0.679
Mother’s age (< 20)	1.372	0.619	0.262	0.149	0.535
Father’s age (< 20)	0.842	1.020	0.443	0.618	0.291
Father is younger than mother	0.316	0.906	1.587	0.120	0.623
Parents’ race/ethnicity:					
– (Ref: Both parents are black)					
– Both parents are white	0.274	0.643	0.449	1.011	0.000
– Both parents are Hispanic	0.225	1.206	0.779	0.538	0.000
– Both parents are of “other”	0.018	1.386	0.427	0.787	0.679
– Mother = white, Father ≠ non-white	0.755	0.144	0.157	0.293	0.000
– Mother = black, Father ≠ non-black	0.374	1.150	0.664	1.772	1.165
– Mother = Hispanic, Father ≠ non-Hispanic	0.515	1.308	0.891	0.420	0.000
– Mother = other, Father ≠ other	0.752	1.150	0.043	0.057	0.679
Parents’ region of birth:					
– (Ref: Both parents are born in U.S.)					
– Mother is foreign-born (not Father)	0.032	0.069	0.025	0.000	0.000

Table 8.3 (Continued)

	Block 1	Block 2	Block 3	Block 4	Block 5
– Father is foreign-born (not Mother)	1.114	1.490	0.717	1.140	0.400
– Both parents are foreign-born	0.966	1.210	2-104	0.682	0.000
Child household income: (Ref: < \$10,000)					
– Between \$10,000 and \$24,999	0.452	0.267	0.057	0.251	1.405
– More than \$25,000	0.338	0.185	0.341	0.515	0.623
Parents' educational backgrounds:					
– (Ref: Less than HS)					
– Mother's education: H.S. diploma/GED	1.898	1.198	0.801	1.247	0.400
– Mother's education: some college	0.859	1.383	1.410	1.047	0.914
– Mother's education: bachelor and beyond	1.026	0.018	1.227	0.553	0.000
– Father's education: H.S. diploma/GED	1.530	1.055	1.041	2.422	0.734
– Father's education: some college	0.070	0.091	0.408	1.403	0.914
– Father's education: bachelor and beyond	0.515	0.333	1.312	0.057	0.000
Mother's education relative to father's:					
– (Ref: Same)					
– Father is less educated than mother	1.355	1.897	1.229	0.230	0.167
– Father is more educated than mother	0.164	0.245	0.561	0.666	1.371
Parents' labor force participation:					
– (Ref: Neither parents work)					
– Both parents work	1.018	0.453	0.585	0.334	1.648
– Only Mother works	0.000	0.650	0.247	0.571	0.167
– Only Father works	1.024	0.727	0.306	0.167	0.291
Mother's weekly hours of work	0.627	0.404	0.451	0.450	0.035

Table 8.3 (Continued)

	Block 1	Block 2	Block 3	Block 4	Block 5
Father's weekly hours of work	0.396	0.713	1.918	0.506	0.077
Mother's labor income > Father's labor income	1.065	1.462	0.025	0.000	0.000
Length of parents' relationship prior to pregnancy					
– (Ref: > 2 years)					
– ≤ 6 months	1.527	0.293	0.781	0.509	1.165
– 6 months ~ 1 year	0.400	0.414	0.855	0.900	0.623
– 1 year ~ 2 years	1.050	0.587	1.673	0.230	1.031
Mother is catholic	0.451	0.084	0.291	0.862	0.623
Mother has no religious affiliation	1.547	1.691	0.837	0.148	0.914
Mother attends religious activities (at least few times a week)	1.608	1.482	1.005	0.874	0.465
Father suggested abortion during pregnancy	0.122	0.814	0.568	0.496	1.405
Maternal grandmother's education (some college and beyond)	0.450	0.439	0.742	0.077	0.679
Prenatal smoking or drinking (mother)	1.678	0.329	1.046	0.423	0.167
Parents in visiting relationship (baseline)	1.114	0.092	1.259	0.186	0.000
Mother's PPVT score (measured at year 3)	1.786	1.327	0.782	0.653	0.401

Notes: 1. $|T|$ statistics of the two-sample test of means for “mother’s state of residence at baseline” (14 indicators) not reported here (available upon request).

parametric estimate, they are consistently larger in magnitude. This indicates that non-marital relationship dissolution may not be as detrimental for child health as one might suspect (at least for some children whose parents separate). To see this, consider a child whose parents separate (treatment group). The finding that, on average, the outcome difference between a treated child and a child in the control group that does not (necessarily) share similar disadvantages is smaller (i.e., OLS) than the outcome difference between the same treated child and a control child that does share these disadvantages (i.e., matching) implies that at least for some children in the treated group, having their parents separate may not be as detrimental as if

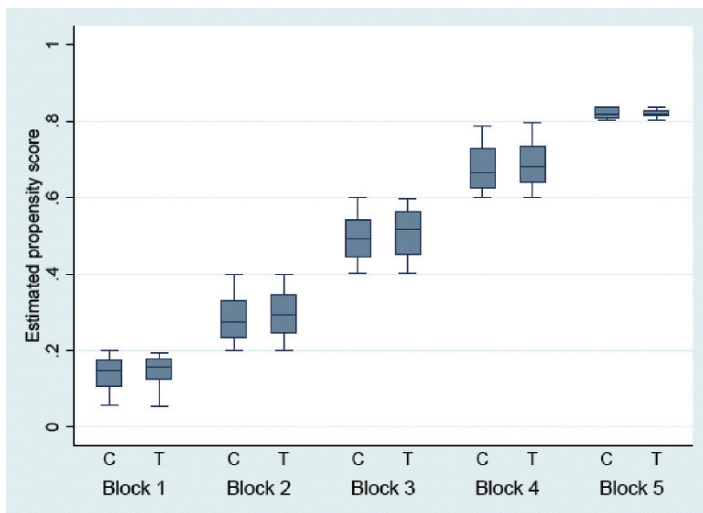


Fig. 8.1 Box plot of the propensity score overlap

their parents had remained romantically involved. Given that caretaker stress level has been identified as an independent determinant of child asthma onset (Wright et al. 2002), this result is consistent with the hypothesis that separating from a “deadbeat” dad may indirectly benefit some children by reducing the mother’s stress level and enhance her parenting (Waller and Swisher 2006), in addition to potential increases in available resources for the child by allowing the mother to form new relationships (e.g. McLanahan and Sandefur 1994).

8.5.3 Sensitivity Analysis

8.5.3.1 Choosing the Bandwidth

The matching estimates may be sensitive to the choice of bandwidth. The Silverman’s rule-of-thumb (1986) may be used to select the optimal bandwidth:

$$\hat{h} = 1.06 \times \text{Min} \{ \hat{\sigma}, R/1.34 \} \times n^{-\frac{1}{5}}$$

where $\hat{\sigma}$ = sample standard deviation, R = interquartile range (75th-quantile – 25th-quantile), and n = sample size. The method is based on the assumption that the underlying distribution of $p(X)$ (the propensity score) is normally distributed. The rule-of-thumb will give reasonable results for all distributions that are unimodal, fairly symmetric and do not have fat tails. However, the rule-of-thumb may not be applicable in our case as the distribution of the estimated propensity score is far from normal (see Appendix Fig. 8.2). As a result, the bandwidth suggested by the rule-of-

Table 8.4 Summary of the effect of Parents’ separation on the Child’s likelihood of developing Asthma by age 3

	OLS		Matching				
	Un-adjusted	Adjusted	Gaussian	Epanechnikov	Uniform		
				($h = 0.01$)	($h = 0.005$)	($r = 0.01$)	($r = 0.005$)
Estimate	0.078*	0.052*	0.061*	0.071*	0.071*	0.067*	0.069*
Standard error	[0.024]	[0.026]	[0.028]	[0.033]	[0.035]	[0.027]	[0.028]
<i>N</i> Treated	526	526	526	526	526	526	517
<i>N</i> Controls	893	893	893	893	893	880	862
% Matched treated			100	100	100	100	98

Notes. 1. The OLS estimates of the separation effect without controls (“unadjusted”) and with controls (“adjusted”) are reported. 2. h = bandwidth, and r = radius. 3. Robust standard error reported for the OLS estimate, standard errors for the matching estimates are obtained by bootstrapping with 500 replications. 4. Propensity score is re-estimated at each replication of the bootstrap procedure to account for the uncertainty associated with the estimation of the propensity score. 5. Estimated propensity score in region of common support [0.05292221, 0.83660801], which is defined by the minimum estimated propensity score within the treatment group, and the maximum estimated propensity score within the control group. 6. The propensity score is estimated using a probit model with the following specification: $Pr[S_i = 1] = F[\text{Parents’ relationship status at childbirth, child is of low birth weight, child gender, birth order of the child (mother), mother is less than 20 years old, father is less than 20 years old, father is younger than mother, both parents are white, both parents are Hispanic, both parents are of other race, mother is white (not father), mother is Hispanic (not father), mother is of other race (not father), mother is foreign-born (not father), father is foreign-born (not mother), both parents are foreign-born, mother’s education, father’s education, father is less educated than mother, father is more educated than mother, length of time parents knew each other before pregnancy, father suggested abortion during pregnancy, mother’s PPVT score, mother is catholic, mother has no religious affiliation, mother attends religious activities frequently, prenatal smoking and/or drinking (mother), household income at childbirth, mother works (not father), father works (not mother), both parents work, mother’s hours of work per week at childbirth, father’s hours of work per week at childbirth, mother’s labor income exceeds father’s, maternal grandmother has some college education (or more), mother’s state of residence at childbirth}]$

thumb may be far from optimal. If the choice of bandwidth is too large, the treated and their matches tend to differ more on observable characteristics. As a result, the matching estimates tend to converge to that produced by the OLS. Our matching estimates using the bandwidth suggested by the rule-of-thumb ($\hat{h} \approx 0.048$) is statistically equivalent to the OLS estimates. Hence, for our analysis smaller bandwidth(s) (0.010 and 0.005) are chosen to ensure closer matches between the treated and controls are used in the estimation.

8.5.3.2 Relaxing the Common Support Condition

Our estimates are based on observations with propensity scores falling within the common support, to ensure that there are sufficient overlap between the treated and control units to enhance comparability, which may improve the quality of our

estimates. A potential drawback of imposing the common support condition is that as the sample may be considerably reduced, since observations with propensity scores falling outside of the common support boundaries are dropped, the estimated treatment effect may be sensitive to this sample restriction. Hence imposing the common support restrictions is not necessarily better (Lechner 2001). Imposing the common support condition results in 8 control and 7 treated units being dropped from our main analysis. To ensure that our estimates are not sensitive to the exclusion of these observations, we relax the common support condition and re-estimate the ATET using all 1,434 observations. Appendix Fig. 8.3 presents the box plot of the propensity score overlap for this sample. Overall, the ATET estimates obtained by relaxing the common support condition are very similar to our main results (results available upon request).

8.5.3.3 Assessing the Conditional Independence Assumption

An identifying assumption of the matching method, namely CIA, requires that conditional on the observables, the distribution of the potential outcomes of the treated group in the absence of treatment is identical to the outcome distribution of the controls. Yet since the data are uninformative about the distribution of potential outcomes for the treated group in the absence of treatment, they cannot directly reject the CIA. Imbens (2004) proposes an indirect way of assessing its plausibility, relying on estimating a causal effect that is known to be zero. Specifically, the test involves estimating the causal effect of the treatment on a lagged outcome, with its value determined prior to the treatment itself. If it is not zero, this implies that the underlying conditional distribution of the potential outcomes of the treated under no treatment is not comparable to control outcomes. The power of this test is enhanced if the variable used in this proxy test is closely related to the outcome of interest.

A number of studies have found strong associations between low birthweight and subsequent poor lung function among children, including childhood asthma (e.g., Nepomnyaschy and Reichmann 2006). We estimate the “causal” effect of parents’ separation within three years after childbirth on whether the child was of low birthweight (< 88 oz). A child’s birthweight is realized before the treatment can take place, and potentially correlated with the child’s subsequent propensity of developing asthma. All of our matching estimates show that parental separation has no effect on whether the child was of low birthweight (results available upon request).

8.6 Conclusion

This study documents a causal relationship between parental non-marital separation and child health among out-of-wedlock children. Using a recent and representative sample of children all born to unmarried parents in large U.S. cities and adopting a potential outcome framework to account for self-selection into relationship dissolution, we find that parental separation has a detrimental effect on child health. By matching children who share similar backgrounds but differ only in terms of whether

their parents dissolve their romantic relationship, we find that out-of-wedlock children whose parents separate within the first three years after childbirth are 6% ~ 7% more likely to develop asthma by age 3, relative to if their parents *had remained romantically involved*.

Our findings are consistent with explanations that poor health investments and caretaker stress are important determinants of asthma among young children. In particular, we find that socioeconomic disadvantages of fathers are crucial in explaining relationship dissolution between unmarried parents. Similarly, the status and quality of unmarried parents' relationships seem to be important predictors of early paternal involvement (Carlson and McLanahan 2004; Johnson 2001). In addition to the lack of available resources as a result of having a "deadbeat" dad, having a partner who is unable (and potentially unwilling) to provide for the family may contribute to relationship instability and heightened stress level for the mother. If the mother were to maintain a romantic relationship with the father, as opposed to being single or forming new partnerships, she may experience greater socioeconomic hardships and tension with adverse effects on her parenting behavior. Our results are consistent with findings by Sigle-Rushton (2005), that men who fathered children out of wedlock are more likely to experience relationship instability, which are likely to militate against protective benefits of social bonds that a union may confer. Hence, promoting greater (or maintained) involvement between these parents may induce some parents to remain in unhealthy relationships (Allard et al. 1991; Raphael and Tolman 1997), with potentially undesirable consequences for the children involved.

The rise in unmarried parenthood and research suggesting that children from single parent families face disadvantages as adults, prompted recent policies geared toward responsible fatherhood initiatives and promoting greater involvement of fathers with their biological children (Harden 2002). While there is evidence suggesting that the majority of unmarried fathers are highly involved in their child's lives, especially during the first few years after childbirth (McLanahan et al. 1998), studies of divorced fathers indicate that men often disengage from their children when their romantic relationship with the mother ends (e.g., Furstenberg and Cherlin 1991). Even more controversial, government funding for programs promoting fathers' co-residence with their children through marriage are in place. While our findings generally support stronger paternal involvement and child support enforcements to protect out-of-wedlock children from socioeconomic hardship, policies that promote marriage between unmarried parents should be mindful that a notable share of the fathers that are targeted might have characteristics not conducive for healthy relationships.

Two caveats of this study should be noted. The matching approach addresses selection effects driven by differences in observable characteristics between children of separated and intact parents. It implicitly assumes that even if there are unobservable factors affecting both relationship dissolution and child outcomes, they are correlated and hence proxied by included controls. While we have access to an unusually detailed sets of observable characteristics including information on both parents' and the quality of their relationship, our estimates may still suffer from some selection bias due to unobservables affecting both parental relationship status

and child outcomes such as the home environment and other family-level influences. Within-cluster matching (or “Differences-in-differences” matching) makes further attempts to account for selection on unobservables by requiring that observations in the control groups be identical to the treated ones in a dimension believed to be particularly important to capture common (unobserved) background influences (for an application to the context of out-of-wedlock childbearing and schooling see Levine and Painter 2003). A possible application of this approach in our context is to require the children in the control group to come from the same family as the treated child. However, this is beyond the scope of the present study since it would require multiple children to be observed for each couple and such data are not available in the FFCWS.

Finally, while this study reports the effect of non-marital separation between the parents on child health, one may also be interested in how it compares to the effect of marital separation, holding union duration and other aspects constant. Although the FFCWS interviewed a sample of married parents with a newborn at baseline, the sample size (net of sample attrition by wave 3) of initially married parents is small and fewer than 5% (roughly 30 observations) divorced before their child reaches age 3. In addition, due to sample design, information on parents with a newborn in the FFCWS are limited to the observational period only: time after the birth of the focal child (who is more likely of higher parity than a child born to unmarried parents at baseline). As such, we have very little information on parents who are married at baseline prior to marriage (or even prior to childbirth) needed to account for important differences between married and divorced families. Hence, comparisons between the effects of marital vs. non-marital dissolution on child outcomes are beyond the scope of this study.

References

- Allard, M.A., R. Albeda, M.E. Colten and C. Cosenza (1991). *In Harm's Way? Domestic Violence, AFDC Receipt, and Welfare Reform in Massachusetts*. Boston, MA: University of Massachusetts, McCormack Institute and the Center for Survey Research.
- Becker, G.S. (1973). A Theory of Marriage: Part I. *Journal of Political Economy* 81(4): 813–46.
- Becker, G.S. (1974). A Theory of Marriage: Part II. *Journal of Political Economy* 82(2): 11–26.
- Becker, G.S. (1991). *A Treatise on the Family*. (Enl. Ed.) Cambridge, MA: Harvard University Press.
- Becker, G.S., E. Landes and R. Michael (1977). An Economic Analysis of Marital Instability. *Journal of Political Economy* 85: 1141–88.
- Becker, S.O. and A. Ichino (2002). Estimation of Average Treatment Effects Based on Propensity Scores. *The Stata Journal* 2(4): 358–77.
- Blau, D.M. (1999). The Effect of Child Care Characteristics on Child Development. *Journal of Human Resources* 34(4): 786–822.
- Brown, S. (2002). Child Well-Being in Cohabiting Families. In: *Just Living Together: Implications of Cohabitation on Families, Children, and Social Policy*, eds. A. Booth and A. Crouter. Manwah, NJ: Lawrence Erlbaum Associates.
- Brown, S. (2004). Family Structure and Child Well-Being: The Significance of Parental Cohabitation. *Journal of Marriage and The Family* 66(2): 351–67.

- Bumpass, L. and H.H. Lu (2000). Trends in Cohabitation and Implications for Children's Family Contexts in the United States. *Population Studies* 54(1): 29–41.
- Carlson, M. and S. McLanahan (2004). Early Father Involvement in Fragile Families. In: *Conceptualizing and Measuring Father Involvement*, eds. R. Day and M. Lamb. Mahwah, NJ: Lawrence Erlbaum Associates.
- Carlson, M., S. McLanahan and P. England (2004). Union Formation in Fragile Families. *Demography* 41(2): 237–61.
- Cherlin, A.J. (1992). *Marriage, Divorce, Remarriage*, (Rev. and Enl. Ed.). Cambridge, MA: Harvard University Press.
- Cherlin, A.J. (1999). Going to Extremes: Family Structure, Children's Well-Being, and Social Sciences. *Demography* 36: 421–28.
- Cherlin, A.J., J. Griffith and J. McCarthy (1983). A Note on Maritally-Disrupted Men's Reports of Child Support in the June 1980 Current Population Survey. *Demography* 20(3): 385–89.
- Clark, D., J. Cornelius, D. Wood and M. Vanyukov (2004). Psychopathology Risk Transmission in Children of Parents with Substance Use Disorders. *American Journal of Psychiatry* 161: 685–91.
- Cogswell, J.J., E.B. Mitchell and J. Alexander (1987). Parental Smoking, Breast Feeding, and Respiratory Infection in Development of Allergic Diseases. *Archives of Disease in Childhood* 62: 338–44.
- Coleman, J. (1988). Social Capital in the Creation of Human Capital. *American Journal of Sociology* 94: 95–120.
- Cox, M., M. Owen, M. Lewis and V. Henderson (1989). Marriage, adult adjustment, and early parenting. *Child Development* 60(5): 1015–24.
- Dehejia, R. and S. Wahba (1999). Causal Effect in Nonexperimental Studies: Reevaluation of the Evaluation of Training Programs. *Journal of the American Statistical Association* 94(488): 1053–62.
- Drewianka, S. (2004). How Will Reforms of Marital Institutions Influence Marital Commitment? A Theoretical Analysis. *Review of Economics of the Household* 2(3): 303–23.
- Edin, K. (2000). Few Good Men: Why Poor Women Don't Remarry. *American Prospect* 11(4): 1–8.
- Fagan, J. and M. Barnett (2003). The Relationship between Maternal Gatekeeping, Paternal Competence, Mothers' Attitudes about the Father's Role, and Father Involvement. *Journal of Family Issues* 24: 1020–43.
- Furstenberg, F.F., Jr. and A. Cherlin (1991). *Divided Families: What Happens to Children when Parents Part*. Cambridge, MA: Harvard University Press.
- Garfinkel, I., S.S. McLanahan and T.L. Hanson (1998). A Patchwork Portrait of Nonresident Fathers. In: *Fathers Under Fire*, ed. I. Garfinkel, S. McLanahan, D. Meyer and J. Seltzer. New York: Russell Sage Foundation.
- Goldstein, J.R. and K. Harknett (1988). National survey of prevalence of asthma among children in the United States, 1976 to 1980. *Pediatrics* 88(1): 1–7.
- Gergen, P.J., D.I. Mullally and R. Evans (2006). Parenting Across Racial and Class Lines: Assortative Mating Patterns of New Parents Who Are Married, Cohabiting, Dating or No Longer Romantically Involved. *Social Forces* 85(1): 121–43.
- Harden, B. (2002). Finding Common Ground on Poor Deadbeat Dads. *The New York Times*, February 3: 3.
- Heckman, J.J. and R. Robb Jr. (1985). Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics* 30(1): 239–67.
- Heckman, J.J. and V.J. Hotz (1989). Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training. *Journal of The American Statistical Association* 84: 862–80.
- Heckman, J.J., H. Ichimura and P. Todd (1997). Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *Review of Economic Studies* 64(4): 605–54.

- Heckman, J.J., H. Ichimura and P. Todd (1998). Matching as an Econometric Evaluation Estimator. *Review of Economic Studies* 65(2): 261–94.
- Heiland, F. and S.H. Liu (2006). Family Structure and Wellbeing of Out-of-Wedlock Children: The Significance of the Biological Parents' Relationship. *Demographic Research* 15(4): 61–104.
- Hofferth, S.L. (2001). Women's Employment and Care of Children in the United States. In: *Women's Employment in a Comparative Perspective*, eds. T. Van der Lippe and L. Van Dijk. New York: Aldine de Gruyter.
- Imbens, G. (2004). Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review. *Review of Economics and Statistics* 86(1): 4–29.
- Jaffee, S.R., A. Caspit, T.E. Moffitt, A. Taylor and N. Dickson (2001). Predicting Early Fatherhood and Whether Young Fathers Live with Their Children: Perspective Findings and Policy Considerations. *Journal of Child Psychology and Psychiatry* 42: 803–15.
- Jekielek, S.M. (1998). Parental Conflict, Marital Disruption, and Children's Emotional Well-Being. *Social Forces* 76: 905–35.
- Johnson, W.E. Jr. (2001). Paternal Involvement among Unwed Fathers. *Children and Youth Services Review* 23: 513–36.
- Kilpeläinen, M., M. Koskenvuo, H. Helenius and E.O. Terho (2002). Stressful Life Events Promote the Manifestation of Asthma and Atopic Diseases. *Clinical & Experimental Allergy* 32(2a): 256–63.
- Klinnert, M.D., P.J. Mrazek and D.A. Mrazek (1994). Early Asthma Onset: The Interaction between Family Stressors and Adaptive Parenting. *Psychiatry* 57(1): 51–61.
- Klinnert, M.D., H.S. Nelson, M.R. Price, A.D. Adinoff, D.Y.M. Leung and D.A. Mrazek (2001). Onset and Persistence of Childhood Asthma: Predictors from Infancy. *Pediatrics* 108(4): e69.
- Lamb, M.E. (2004). *The Role of the Father in Child Development*. New York: John Wiley & Sons.
- Lechner, M. (2001). A Note on the Common Support Problem in Applied Evaluation Studies. Discussion Paper 2001–01, Department of Economics, University of St. Gallen.
- Lechner, M. (2002). Program Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labor Market Policies. *Review of Economics and Statistics* 84(2): 205–20.
- Levine, D.I. and G. Painter (2003). The Schooling Costs of Teenage Out-of-Wedlock Childbearing: Analysis with a Within-School Propensity-Score-Matching Estimator. *Review of Economics and Statistics* 85(4): 884–900.
- Levitan, H. (1985). Onset of Asthma during Intense Mourning. *Psychosomatic* 26: 939–41.
- Liu, S.H. (2006). Is my Parents' Divorce to Blame for my Failure in Life? A joint Model of Child Educational Attainments and Parental Divorce. Department of Economics Working Paper WP2006-10, University of Miami.
- Liu, S.H. and F. Heiland (2009). Should we get Married? The Effect of Parents' Marriage on Out-of-Wedlock Children. *Economic Inquiry* (forthcoming).
- Maccoby, E.E. and J.A. Martin (1983). Socialization in the Context of the Family: Parent-Child Interaction. In: *Handbook of Child Psychology: Socialization, Personality, and Social Development*, Vol. 4., eds. P.H. Mussen and E.M. Hetherington. New York: John Wiley & Sons.
- Manser, M. and M. Brown (1980). Marriage and Household Decision-Making: A Bargaining Analysis. *International Economic Review* 21(1): 31–44.
- Manning, W., P. Smock and D. Majumbar (2004). The Relative Stability of Cohabiting and Marital Unions for Children. *Population Research and Policy Review* 23(2): 135–59.
- Martin, J.A., B.E. Hamilton, P.D. Sutton, S.J. Ventura, F. Menacker and S. Kirmeyer (2006). *Births: Final Data for 2004*, Vol. 55. Department of Health and Human Services, Center for Disease Control and Prevention, National Center for Health Statistics.
- McElroy, M.B. and M.J. Horney (1981). Nash-Bargained Household Decisions: Toward a Generalization of the Theory of Demand. *International Economic Review* 22(2): 333–49.
- McElroy, M.B. (1990). The Empirical Content of Nash-Bargained Household Behavior. *Journal of Human Resources* 25(4): 559–83.

- McLanahan, S. (1985). Family Structure and the Reproduction of Poverty. *American Journal of Sociology* 90(4): 873–901.
- McLanahan, S., I. Garfinkel, J. Brooks-Gunn, H. Zhao, W. Johnson, L. Rich and M. Turner (1998). Unwed Fathers and Fragile Families. *Center for Research on Child Wellbeing*, Working Paper 98–12.
- McLanahan, S. and G. Sandefur (1994). *Growing Up with a Single Parent: What Hurts, What Helps*. Cambridge, MA: Harvard University Press.
- Michael, R.T. (1973). Education in Nonmarket Production. *Journal of Political Economy* 81(2): 306–27.
- Morrison, D. and A. Ritualo (2000). Routes to Children’s Economic Recovery after Divorce: Are Cohabitation and Remarriage Equivalent? *American Sociological Review* 65(4): 560–80.
- Neidell, M.J. (2004). Air Pollution, Health, and Socio-economic Status: The Effect of Outdoor Air Quality on Childhood Asthma. *Journal of Health Economics* 23(6): 1209–36.
- Nepomnyaschy, L. and N.E. Reichmann (2006). Low Birthweight and Asthma Among Young Urban Children. *American Journal of Public Health* 96(9): 1604–10.
- Olivetti, J., C. Kercksmar and S. Redline (1996). Pre- and Perinatal Risk Factors for Asthma in Inner City African-American Children. *American Journal of Epidemiology*, 143(6): 570–77.
- Osborne, C. and S. McLanahan (2006). The Effects of Partnership Instability on Parenting and Young Children’s Health and Behavior. Center for Research on Child Well-Being, Working Paper No. 04-16-FF.
- Raphael, J. and R.M. Tolman (1997). Trapped by Poverty, Trapped by Abuse: New Evidence Documenting the Relationship between Domestic Violence and Welfare. The Taylor Institute and University of Michigan Research Development Center on Poverty, Risk, and Mental Health (<http://humanservices.ucdavis.edu/resource/uploadfiles/x%20Trapped%20by%20Poverty,%20Trapped%20by%20Abuse.pdf>).
- Reichman, N., I. Garfinkel, S. McLanahan and J. Teitler (2001). The Fragile Families: Sample and Design. *Children and Youth Services Review* 23(4/5): 303–26.
- Ribar, D.C. (2006). What Do Social Scientists know About the Benefits of Marriage? A Review of Quantitative Methodologies. IZA Discussion Paper No. 998.
- Rosenbaum, P.R. and D.B. Rubin (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70(1): 41–55.
- Rubin, D.B. (1979). Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observation Studies. *Journal of the American Statistical Association* 74(366): 318–28.
- Shaw, K. (1987). The Quit Propensity of Married Men. *Journal of Labor Economics* 5(4): 533–60.
- Sigle-Rushton, W. (2005). Young Fatherhood and Subsequent Disadvantage in the United Kingdom. *Journal of Marriage and the Family* 67: 735–53.
- Silverman, B.W. (1986). *Density Estimation*. London: Chapman and Hall.
- Sorenson, E. (1997). A National Profile of Nonresident Fathers and Their Ability to Pay Child Support. *Journal of Marriage and the Family* 59(4): 785–97.
- Sporik, R., S.T. Holgate and J.J. Cogswell (1991). Natural History of Asthma in Childhood: A Birth Cohort Study. *Archives of Disease in Childhood* 66: 1050–53.
- Steele, F., W. Sigle-Rushton and O. Kravdal (2007). Consequences of Family Disruption on Children’s Educational Outcomes in Norway. Working Paper (http://www.eui.edu/Personal/Dronkers/Divorce/Divorceconference2007/Steele_Sigle-Kraval.pdf).
- Teiramaa, E. (1979). Psychosocial and Psychic Factors and Age at Onset of Asthma. *Journal of Psychosomatic Research* 23: 27–37.
- Thomson, E., T.L. Hanson and S. McLanahan (1994). Family Structure and Child Wellbeing: Economic Resources vs. Parental Behaviors. *Social Forces* 73(1): 221–42.
- Waller, M.R. and R.R. Swisher (2006). Fathers’ Risk Behaviors in Fragile Families: Implications for ‘Healthy’ Relationships and Father Involvement. *Social Problems* 53(3): 392–420.
- Waite, L.J. and M. Gallagher (2000). *The Case for Marriage: Why People are Happier, Healthier, and Better Off Financially*. New York: Broadway Books.

- Weisch, D., D. Meyers and E. Bleeker (1999). Genetics of Asthma. *Journal of Allergy and Clinical Immunology* 104: 895–901.
- Weiss, Y. and R.J. Willis (1997). Match Quality, New Information, and Marital Dissolution. *Journal of Labor Economics* 15(1): 293–329.
- Weitzmann, M., S. Gortmaker, D. Klein Walker and A. School (1990). Maternal Smoking and Childhood Asthma. *Pediatrics* 85: 505–11.
- Willis, R.J. and J.G. Haaga (1996). Economic Approaches to Understanding Nonmarital Fertility. *Population and Development Review* 22: 67–86.
- Willis, R.J. (1999). A Theory of Out-of-Wedlock Childbearing. *Journal of Political Economy* 107(6): 33–64.
- Wilson, M. and J. Brooks-Gunn (2001). Health Status and Behaviors of Unwed Fathers. *Children and Youth Services Review* 23: 377–401.
- Wu, L.L. (1996). Effects of Family Instability, Income, and Income Instability on the Risk of a Premarital Birth. *American Sociological Review* 61(3): 386–406.
- Whyte, M.K. (1990). *Dating, Mating, and Marriage*. New York: Aldine de Gruyter.
- Wright, R.J., S. Cohen, V. Carey, S.T. Weiss and D.R. Gold (2002). Parental Stress as a Predictor of Wheezing in Infancy. *American Journal of Respiratory and Critical Care Medicine* 165(3): 358–65.

Appendix

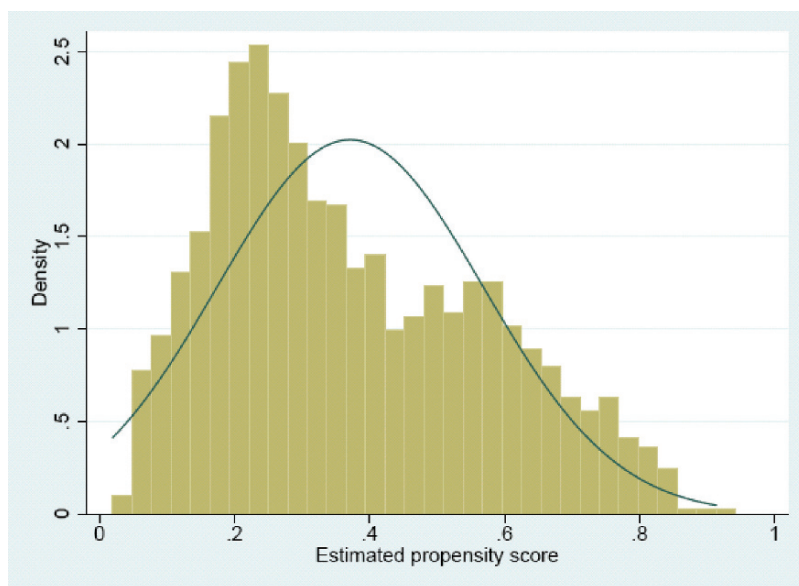


Fig. 8.2 Distribution of the estimated propensity score (Relaxing the common support condition)

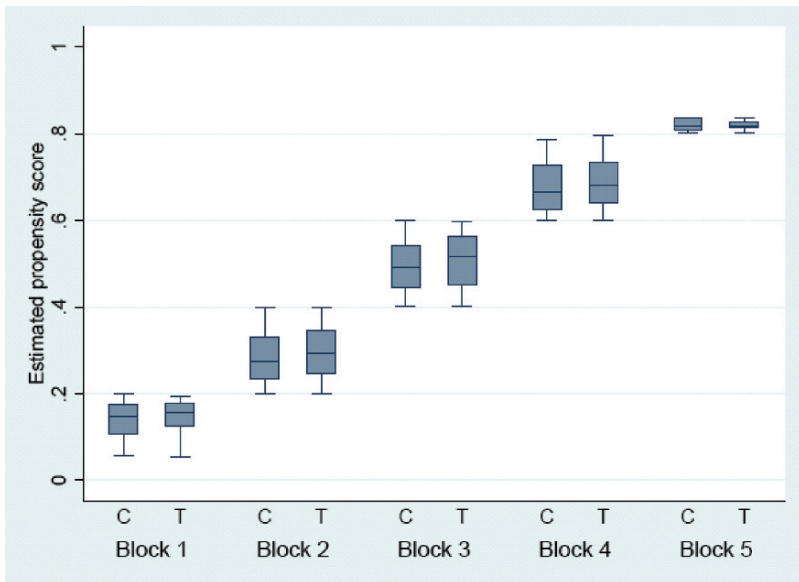


Fig. 8.3 Box plot of the propensity score (Relaxing the common support condition)

Chapter 9

Assessing the Causal Effect of Childbearing on Household Income in Albania

Francesca Francavilla and Alessandra Mattei

9.1 Introduction

The relationship between demographic developments and economic performance has been the subject of rather intense debate in the economics literature for nearly two centuries.

Until recently limitations on both data sources and statistical techniques have prevented clear insights into the relationship between population growth and economic wellbeing (Birdsall et al. 2001), and most of the existing studies have relied on either cross sectional or aggregate level data. Cross sectional data, no matter what techniques are applied, is unlikely to provide robust causal information about the relationship between the occurrence of life events (such as a childbearing event) and economic wellbeing. Past empirical studies concerning the relationship between economic wellbeing and fertility have consequently showed mixed results, indicating that the relationship does not appear to be unidirectional (see Schoumaker and Tabutin (1999) for further details).

In this paper we analyze to what extent births may lead to changes in economic wellbeing. In contrast to most previous studies on this issue we apply appropriate econometric techniques based on longitudinal micro data in order to identify the causal effects of child bearing events on poverty.

Fertility is measured in terms of childbearing events, and we use monthly real equivalised income as an indicator of household living standards. Childbearing might affect economic wellbeing through different channels. The most obvious one is that an additional child in the household increases the number of adult equivalence units without increasing household income. Therefore, childbearing would decrease, *ceteris paribus*, (equivalised) household income. However, as the economic theory suggests, there exist many factors that might interact with both fertility and income, generating economies and/or diseconomies of scale (Cigno 1991). One of the main factors concerns the impact of fertility on the optimal time allo-

A. Mattei (✉)

Department of Statistics, University of Florence, Viale Morgagni, 59, Firenze, Italy
e-mail: mattei@ds.unifi.it

cation within the household. According to the principle of division of labor, the birth of an additional child may require a re-allocation of specific tasks within the household (Cigno 1991). This kind of specialization is a key feature of domestic organization (Becker 1985). Private and public transfers are other possible transmission channels. If the credit market is deficient or rationed, an extended family network could substitute for a capital market by arranging loans to its young members from its middle-aged ones and enforcing repayment later when the young borrowers have become middle-aged and the middle-aged lenders have become old (Ermisch 2003). Moreover, in environments with less-developed markets, altruism or mutual “caring” among family members play an important role in facilitating risk sharing (Becker 1991). In an endogenous fertility model, public transfer could be justified if society assigned a positive welfare weight to children in their own right (Cigno 1983), or if children generated a positive externality (Cigno et al. 2000). State transfers may, totally or partially, compensate households for income loss due to the cost of children, and in turn influence fertility decisions. Other public policies, affecting quality of life, may contribute to explaining the interaction between childbearing and economic wellbeing. Service provision affects the ability of the family to deal with the reduction in equivalised income; for instance, support for child care costs will help parents to take paid work. However, some of the social transfers benefiting childbearing may be in the form of tax concessions rather than cash payments. In general, progressive income taxation would mitigate diseconomies, because any fall in earnings would reduce the marginal rate of tax, whereas progressive child subsidization could generate economies in the number of children (Cigno 1991). Since all these potential sources of economies or diseconomies could be present at the same time, it is possible that economies of scale would result for a number of children, and diseconomies for others (Cigno 1991).

The focus of this paper is primarily on the relationship between fertility and wellbeing. We perform our analysis on longitudinal data from the Albanian Living Standard Measurement Survey (ALSMS). Albania is interesting for a range of reasons. Since 1992, when democracy was re-installed in Albania, the country has experienced rapid political, social and economic changes. However, the country is by far the poorest in Europe, and in terms of the human development indicator, only ranked at 73rd out of 177 countries (see the UNDP web site: <http://hdr.undp.org>).

We take a quasi experimental approach, that is, we consider the variable of interest (the experience of a childbearing event) as the treatment variable, and our measure of wellbeing as the outcome variable. Individuals experiencing a childbearing event might be self selected, generating systematic differences in background characteristics between the treatment groups. In order to deal with this confounding factor, we first fit a multiple linear regression model that includes relevant background characteristics as well as an indicator variable for the treatment (i.e. childbearing). This estimation is then compared and contrasted with a matching approach, which is specifically designed to deal with the problem of confounding in observational studies. We apply the bias-corrected matching estimator introduced by Abadie and Imbens (2002), which allows us to regression-adjust the difference within matches for the difference in covariate values. Our analysis suggests that there is some

evidence that childbearing events can in fact increase household wellbeing in Albania, although the causal parameter estimate is not significant. In addition, the treatment effect is highly heterogeneous with respect to observable characteristics such as the woman's working status and the woman's parity. All the results appear to be robust with respect to the estimated equivalence scale: changing the equivalence scale leaves the childbearing effect on income positive and non-significant.

The structure of the paper is as follows. Section 9.2 briefly describes the Albanian context. Section 9.3 gives a short description of the ALSMS data. Section 9.4 explains how we define wellbeing putting particular emphasis on the choice of the equivalence scale. Using this wellbeing definition, Section 9.5 provides interesting descriptive patterns of wellbeing for different family types. Section 9.6 explains the methodological strategy for the causal analysis and Section 9.7 presents the results along with a dissertation on the robustness of our estimates with respect to the selected equivalence scale. Section 9.8 draws some conclusions.

9.2 The Albanian Background

Given the socialist background, Albania has a history of strong social protection. Before the collapse of communism, guaranteed employment schemes protected most families from poverty ensuring them income from earnings. Wages were low but prices and rents were controlled and the state invested extensively in maternal and early child health. Since 1992, when democracy was reinstalled, Albania has enjoyed strong economic growth and its economic progress is rapidly transforming Albania to become a middle income country. From the mid 1990s, Albania's GNP started to grow and surpassed the so-called Lower Income countries, and currently the GNP is moving toward the levels of the Middle Income countries. Despite the impressive performance of the economy over the last years, Albania continues to have one of the lowest levels of per capita income in Europe and the incidence of poverty in Albania is large compared with countries in the region.¹ According to the World Bank Poverty Assessment in 2003 one-quarter of the Albanian population – about 780,000 people – fell below the poverty line, and around 5% of the population – 150,000 people – are extremely poor. The modernization that the country experienced in the last decade has benefited Tirana and other urban areas more than rural areas. Poor individuals in rural areas comprise nearly 35% of the population and almost half of the residents in the most remote districts in the North and North-East Mountain regions are poor.

¹ According to the World Development Indicator database and the Country Poverty Assessment Reports by World Bank, Albania is the eighth poorest country among the transition economies in Europe and Central Asia. Herzegovina (19%), FYR of Macedonia (16%), Bulgaria (13%) and Croatia (8%); whereas some recent studies show that the poverty dimension in Albania is near to some countries of the Commonwealth of Independent States such as Uzbekistan and Moldova. See World Bank, *Making Transition Work for Everyone: Poverty and Inequality in Europe and Central Asia*, 2004.

Amongst the Southeast European countries, Albania performs badly in many health indicators, education attainment, and dependency ratio (see International Monetary Fund, 2005 for further details). The official statistics suggests that in 2003 Albania experienced infant and maternal mortality rates equal to 18 per 1000 births and 21 per 100,000 births, respectively, – which appear to be the highest levels of the Southeast European area. However in the last years the two indicators have reported an encouraging downward trend. Despite these pictures the life expectancy at birth, currently 74 years, is comparable with European countries.

The strong economic growth following the transformation to a market economy, obviously produced rapid and dramatic social changes. Several structural reforms have been carried out involving banking, land reforms and privatization. Almost all the small and medium enterpriser and the strategic sectors (such as telecommunication) have been privatized.

In 1993 the Social Insurance System existing since 1946 was completely reorganized. The new law introduced in 1993 and the following amendments brought substantial changes in the Albanian social assistance programme which included old age, disability and survivor pensions, sickness and maternity benefits, work injury, as well as unemployment benefits and family allowances, which were introduced for the first time. There is however, no specific child benefit, but general “economic assistance” is allocated on a means-tested basis for families with low earned income. Employees with at least twelve months of contributions are entitled to 365 days of paid maternity leave. The benefit is 80% of the average daily wage in the last calendar year for the leave period taken before childbirth and for 150 days after, whereas the benefit is paid at 50% of the average daily wage for the remainder of the entitlement period. For more children extensions are provided. Compensation is payable for changes of employment due to pregnancy. A lump-sum payment is payable to either insured parent with a minimum of 1 year’s contributions.

Moreover, the Albanian social system provides a child supplement for each dependent child under age 15. It is clear therefore, that there is still reasonably good support available for mothers with young children. Whereas, the total fertility rate has declined steadily over the years, it seems to have stabilized in recent years, and there is little indication that Albania will experience lowest low fertility as experienced in Italy and other Mediterranean countries.

9.3 The Albania Living Standards Measurement Study

Our analysis is based on data from the Albania Living Standards Measurement Survey (ALSMS), a periodic study carried out by the Albanian Institute of Statistics (INSTAT) with the technical and financial assistance of the World Bank. The first survey was conducted in 2002, and provided individual level and household level socio-economic data from 3,599 households drawn from urban and rural areas in Albania. The sample was designed to be representative of Albania as a whole, Tirana, other urban/rural locations, and the three main agro-ecological

areas (Coastal, Central, and Mountain). The 2002 ALSMS was followed by two panel surveys (in 2003 and 2004) on a sub-sample of the original households. The sample size for the panel took approximately half the ALSMS households and has re-interviewed these households annually in 2003 and 2004. The ALSMS data collected in 2002 therefore constitute “wave 1” of a three-wave panel survey.

The sample selected from the ALSMS for the panel was designed to provide a nationally representative sample of households and individuals within Albania. This differs from the original ALSMS where the sample was designed to be representative of each strata which broadly represented the main regions in Albania so that regional level statistics could be generated (Mountain, Central, Coastal, Tirana). The panel is essentially an individual level survey as individuals are followed over time regardless of the household they live in at any given interview point.

The 2002 survey contains a wealth of information collected at the individual and household levels. Information collected at the household level includes housing, subjective poverty, consumption expenditures, agriculture, non-farm enterprises, and other income. Information collected at the individual level includes demographics, migration, education, health, fertility, labor, transfers and social assistance, and anthropometrics (for children under 6 years of age). The ALSMS also collects community level information on the basic characteristics of the community, access to public services such as education, health, and transportation, community services, community organizations, community safety, migration, child labor and problems related to the environment. Finally, the ALSMS has information concerning price which can be used to adjust for regional price differences.

The two following panel waves provide updated individual level and household level socio-economic data for household members 15 years of age and older. It is important to note that we have no panel information on consumption expenditure. In addition, whereas the first wave contains complete fertility histories, waves 2 and 3 only provide additional information on any new births. All the analyses in this paper are based on a sub-sample of women of child-bearing age (15–49 years) with complete information on the relevant variables drawn from the Albanian panel survey.

9.4 A Measure of Well-Being

The focus of our study is on the extent to which childbearing events lead to changes in wellbeing. In order to address this issue we first have to define a measure of wellbeing. As a multidimensional phenomenon, wellbeing can be defined and measured in a multitude of ways. One approach is to think of one’s wellbeing as the command over commodities in general, so people are better off if they have a greater command over resources. In this view, the main focus is on whether households or individuals have enough resources to meet their needs, and wellbeing is typically seen in monetary terms. The most common welfare-monetary indicators for poverty measurement are expenditure on household consumption and household income. In

our study we use an income-based measure for poverty analysis. This choice was mainly driven by the availability of data. As previously noted, in the ALSM study information on consumption expenditure is only available for the first wave; whereas we have data on income for all the three waves of the Albanian panel survey.

Our measure of monetary wellbeing is constructed using the monthly total household income, which comprises income from dependent work (wages, in-kind salaries, bonuses) as well as non-dependent work, earnings transfer (only incoming), public transfers and other income (such as rental income, inheritance, lottery/gambling winnings and other).

When assessing economic wellbeing it is important to adjust for price variability across space and time and household heterogeneity. Microeconomic theory suggests that we may wish to account for price variability by comparing real as opposed to nominal income. Several procedures can be followed to enable such comparisons. Here we deflate the level of total nominal income by a cost-of-living index. Specifically, we convert income in 2004 to be real with respect to 2002 Leks prices, using the aggregate consumption price index reported by the International Monetary Fund (2004).

Household size and demographic composition vary across households, as do the prices they face, including wage rates. As a result, it takes different resources to make ends meet for different households. In order to adjust for household heterogeneity we use an equivalence scale, that is, we divide the real total household income by the number of adult equivalents, n_e :

$$n_e = (A + \alpha \cdot K)^\theta, \quad (9.1)$$

where A and K stand for the number of adults and children, respectively. Both α and θ take a value between 0 and 1. The parameter α is the adult-equivalence of a child, and the parameter θ reflects possible economies of scale favoring larger households, due to the allocation of fixed costs (such as heat and light) over a greater number of people.

The notion of equivalence scale is compelling. It is much less persuasive in practice, because of the problem of picking an appropriate scale. How the parameters α and θ should be calculated and whether it makes sense to even try is still subject to debate, and there is no consensus on the matter. There are two possible solutions to this problem: either pick a scale that seems reasonable on the grounds that even a bad equivalence scale is better than none at all, or try to estimate a scale typically based on observed consumption behavior from household surveys. In our study, preliminary analyses suggested that standard equivalence scales do not work very well. Looking at the cases where α and θ take values of 0.5 or 1, we found that our results were highly sensitive to both the choice of a weight of a child relative to an adult and economies of scale. Therefore, we decided to estimate the equivalence scale from the data.

Following Lanjouw and Ravallion (1995), we focus on the class of equivalence scales whereby the money metric of an individual's welfare has an elasticity θ with respect to household size. As in Lanjouw and Ravallion (1995), the parameter θ is

often termed the “size elasticity”. The welfare of a typical member of any household is then measured in monetary terms by x/n^θ , where x denotes total household consumption expenditure, and n denotes household size; n^θ can be interpreted as the equivalent number of single-persons.

It is well known that empirical data alone cannot reveal equivalence scales. Additional assumptions are needed to identify equivalence scales from observed data on household consumption patterns. The approach we follow is based on what is sometimes called Engel’s second law, which asserts that the food share is an inverse indicator of welfare across households of different sizes and compositions, namely, the higher the share of non-food spending the better off members of the household are deemed to be. Generally, an Engel curve measures the relationship between the expenditure on a particular good and the total expenditure of the household. In our study, as in Lanjouw and Ravallion (1995), we estimate size elasticity by regressing the food share on the log expenditure per person and a set of demographic variables. The basic specification is the following:

$$\begin{aligned}\omega_{ij} &= \mu + \beta \ln(x_{ij} / n_{ij}^\theta) + X'_{ij}\gamma + v_j + \varepsilon_{ij} \\ &= \mu + \beta \ln(x_{ij}) - \beta\theta \ln(n_{ij}) + X'_{ij}\gamma + v_j + \varepsilon_{ij},\end{aligned}\quad (9.2)$$

where ω_{ij} is a food share of household i in village j , x_{ij} is total household expenditure, n_{ij} denotes the number of household members, θ is the size elasticity, X_{ij} is a set of demographic variables, v_j is community specific characteristics including prices in village j , and ε_{ij} represents an error term. We consider a community fixed effect regression in order to control for relative prices across regions. The estimate of size elasticity, θ , is obtained by taking the ratio of the coefficient on log of household size to that of log of household expenditure in Equation (9.2).

Recall that, in our application information on consumption expenditure is only available for the first wave; so we estimate size elasticity using observed panel data from wave 1 of ALSM survey, and apply the estimated equivalence scale both to income in 2002 and income in 2004.

Table 9.1 shows the results. We consider different specifications of the Engel curve, both imposing the homogeneity restriction, that is $\theta = 1$, (models (5), (6), and (7)) and those which do not (models (1)–(4)). Column 1 is the simple community fixed effect regression of the food share on the logarithm of the household size. There is a slight tendency for larger households to have higher food shares, but the correlation is not strong (the correlation coefficient is 0.108). When expenditures are added (column 2) the estimated size elasticity of the money metric of welfare is 0.415. The homogeneity restriction is rejected ($t - value = -4.428$). In column 3 we give the augmented model including both household size and household composition (represented by the numbers of people in each demographic group) as independent variables. We obtain a value for θ of 0.221, with a standard error of 0.196. The homogeneity restriction is again rejected ($t - value = -3.975$). For this model the demographic composition parameters are not significant; only if the homogeneity restriction is imposed (column 6), do we observe significant even if

Table 9.1 Engel curve estimation of the size elasticity using the first wave of ALSMS. Community fixed effect regression. (Standard errors in parentheses)

Independent variables	Models						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log total expenditure		-0.064 (0.008)	-0.062 (0.008)	-0.063 (0.008)			
Log household size	0.007 (0.011)	0.026 (0.011)	-0.014 (0.037)	0.021 (0.011)			
Log expenditure per person					-0.051 (0.007)	0.060 (0.008)	0.050 (0.007)
No. of adults			0.005 (0.008)			0.010 (0.003)	
No. of children			0.013 (0.009)			0.003 (0.003)	
Proportion of children				0.022 (0.017)			0.006 (0.016)
Constant	0.631 (0.016)	1.262 (0.084)	1.268 (0.085)	1.255 (0.084)	1.092 (0.065)	1.208 (0.080)	1.084 (0.068)
Observations ^a	1301	1301	1301	1301	1301	1301	1301
No. of communities	283	283	283	283	283	283	283
R-squared	0.0117	0.0498	0.0527	0.0501	0.0568	0.0503	0.0571
Implied size elasticity (θ) ^b		0.415 (0.132)	0.221 (0.196)	0.338 (0.131)	1	1	1

^a The number of observations is given by the number of households which our 1698 panel women belonged to at the time of the first wave.

^b The estimate of size elasticity, θ , is obtained by taking the ratio of the coefficient on log of household size to that of log of household expenditure. The standard error for θ is computed using the Delta method.

not strong differences in food shares among households with a different number of adult members. As alternative, the model in column 4 includes the demographics as proportion of children in household. This specification gives an elasticity of 0.338, and leads to a rejection of the homogeneity restriction (t -value = -5.053). In addition, the model suggests that there exists a positive although not strong relationship between demographic composition and food share in the Engel curve (the regression coefficient on proportion of children appears to be significant according to a standard two-sided t -test at the 10% level). Therefore, once relaxed, the equivalence scale implied by the Engel curve appears to be approximated well by n^θ with adjustment for the proportion of children in household.

Thus, we estimate θ to be 0.338. This size elasticity implies surprisingly large falls in food spending per head for consumers. According to these estimated size

economies, ten individuals, each spending, say, 1 Lek per day in separate single-dweller households could achieve the same welfare level living as a 10-person single household with total expenditures lower than 5 Leks per day ($10^{1-0.338} = 4.6$).

9.5 Descriptive Statistics

Table 9.2 presents some descriptive statistics for the sample of 1,698 women classified by a binary variable, Z_i , equal to 1 if woman i experienced a childbearing event between the time of the first wave and 31st, December 2003, and 0 if she did not.

The upper panel of Table 9.2 shows the mean values of the components of the (real) total household income in wave 2002 and wave 2004. All the income components are real with respect to 2002 and equalised, using as number of adult equivalents $n_e = n^{\hat{\theta}}$, where $\hat{\theta} = 0.338$ is the estimated size elasticity from model

Table 9.2 Means in wave 2002 and wave 2004, and relative mean differences between waves by childbearing status for income variables and some demographic variables^a

	Means				Rel. mean	
	Wave 2002		Wave 2004		difference (%) ^b	
	Childbearing		Childbearing		Childbearing	
	Yes	No	Yes	No	Yes	No
Woman's bonuses	7	67	25	109	242.6%	62.2%
Wage	7,713	9,304	9,144	10,385	18.5%	11.6%
Income from self-employed	14,069	8,440	16,289	7,383	15.8%	-12.5%
Private transfers	661	5,562	574	1,307	-13.1%	-76.5%
Public transfers	1,807	1,736	2,296	2,087	27.1%	20.2%
Total income	21,826	21,528	28,632	21,447	31.2%	-0.4%
Maternity benefits (Yes) ^c	0.9%	0.8%	4.7%	0.5%	3.7%	-0.3%
No. of HH workers	2.20	2.01	2.07	1.98	-5.5%	-1.3%
No. of HH male workers	1.20	1.06	1.29	1.04	7.8%	-1.8%
No. of HH female workers	1.00	0.95	0.79	0.94	-21.5%	-0.7%

^a All the income variables are equalised using as equivalence scale $n_e = n^{\hat{\theta}}$, where $\hat{\theta} = 0.338$.

^b The relative mean difference is the mean difference between waves as percentage of the mean in wave 2002: $[100(\bar{x}_{2004}(z) - \bar{x}_{2002}(z))]/\bar{x}_{2002}(z)$, where for each variable $\bar{x}_{2004}(z)$ and $\bar{x}_{2002}(z)$ are the sample mean in wave 2002 and wave 2004 in the group of women with $Z = z$, $z = 0, 1$.

^c "Maternity benefits" is a binary variable equal to 1 if at least a household member received maternity benefits in the last 12 months. Therefore, the means are proportions and the relative difference in percent is the percent difference between waves for each group of women defined by childbearing status.

(4) (see Section 9.4). Table 9.2 also presents the average number of workers by household and the percentage of women belonging to a household where at least a member received maternity benefits in the last 12 months. Finally, the last two columns in the table show the mean difference between waves as percentage of the mean in wave 2002.

Table 9.2 suggests that women who experienced a new birth belong to a household with a higher number of workers with respect to women who did not. This result characterizes both wave 2002 and wave 2004. However, while the higher number of household workers in 2002 appears to be a consequence of a large number of female and male workers, in 2004 the lower number of female workers is compensated by a higher number of male workers. Looking at the trend in the period 2002–2004 Table 9.2 suggests that the number of household workers decreases for the two groups of women in the time. However, the reduction in the number of workers in households where there are women who gave birth to a new child is four percentage points greater than the reduction in the number of workers experienced by the other households. This is probably due to the fact that households who experienced a childbearing event are affected by a high reduction in the number of female workers (21.5% with respect to 2002), which is not sufficiently compensated by the increasing in the number of male workers (8% with respect to 2002). This result strongly suggests that there exists a reorganization of labor supply in households who experienced a new birth. Between 2002 and 2004 households who experienced a childbearing event tend to decrease the supply of female work, and increase the supply of male work. On the contrary, households who did not experience a childbearing event did not substantially reduce their labor supply. Table 9.2 also shows some differences in the income composition of the two groups of women defined by the childbearing status. Both in 2002 and 2004 women who experienced a childbearing event belong to households with a higher self-employed labor income and a lower wage income. This result is partially explained by the fact that women who gave birth to a new child received lower bonuses (which are a component of household income). A further explanation could be related with the higher capacity of household members who work as self-employed to react to a childbearing event modifying their labor supply.

As discussed in Section 9.2, public and private transfers are a crucial component of household income in Albania. The descriptive statistics suggest that women who experienced a new birth belong to households with a slightly higher level of public transfers but a substantial low level of private transfers. This evidence appears in both the waves. Concerning the trend of the different components of income, Table 9.2 shows an increase in the household wages for both groups of women, although we observe a higher growth, of about six percentage points, among women who experienced a new birth. On the contrary, income of self-employed workers appear to increase for household, who experienced a new birth and decrease for households who did not. Table 9.2 shows that public and private transfers move in opposite directions: the former increases while the latter decreases for both groups of women, even if public transfers increase more for women who experienced a

new birth, and private transfers have a heavier reduction for women who did not experience a new birth.

9.6 Identifying the Causal Effect of a New Birth

9.6.1 The Quasi-Experimental Approach

The aim of our analysis is to assess whether in Albania a childbearing event leads to changes in wellbeing. We address this problem using a quasi-experimental approach, that is, we consider the endogenous variable of interest as the treatment variable and a measure of wellbeing as the outcome variable. In our study, the treatment is given by the childbearing status, Z , that is, our binary treatment variable is equal to 1 if a woman experiences a childbearing event between the time of the first wave (t_0) and December 31, 2003, and 0 otherwise. The outcome of interest is the income-based measure of wellbeing at the time of the third wave (t_1) defined in Section 9.4.

More formally, consider a set of N individuals, and denote each of them by subscript i , $i = 1, \dots, N$. At time t_i ($t_0 < t_i < t_1$), subject i is “treated”, i.e., she gives birth to a new child, or “untreated”; in this latter case she will also be named “control”. The treatment indicator is $Z \in \{0, 1\}$. Interest lies in the continuous scalar outcome representing the equivalised income at the time of the third wave t_1 : $Y \in R_+$. Note that the distance between the treatment assignment – that is, the birth of a new child – and the time at which we observe the outcome variable ($t_1 - t_i$) varies among women. For each individual i , $i = 1, \dots, N$, with all units exchangeable, let $(Y_i(0), Y_i(1))$ denote the two potential outcomes, that is, $Y_i(0)$ is the income level for individual i when she is not exposed to the treatment, and $Y_i(1)$ is the income level for individual i when she is exposed to the treatment. If both $Y_i(0)$ and $Y_i(1)$ could be observed, then the effect of the treatment on i would be $Y_i(1) - Y_i(0)$. The root of the problem is that only one of the two outcomes is observed. Let the observed outcome be denoted by Y_i :

$$Y_i \equiv Y_i(Z_i) = Z_i \cdot Y_i(1) + (1 - Z_i) \cdot Y_i(0).$$

In this study, we are interested in the estimation of the average treatment effect for the subpopulation of women who experience a childbearing event, usually called, the Average effect of Treatment on the Treated (ATT):

$$\tau = E(Y_i(1) - Y_i(0) | Z = 1).$$

If we could observe both outcomes, we could estimate this causal effect using the estimator

$$\frac{1}{N_1} \sum_i Z_i \cdot (Y_i(1) - Y_i(0)),$$

where $N_1 = \sum_i Z_i$ is the number of treated units in the sample. In practice, for each treated unit i we observe only the income level under treatment, $Y_i(1)$; the untreated income level $Y_i(0)$ has to be estimated.

If the decision to give birth to a new child was “purely random”, we could expect that the background characteristics in the treatment groups to be similar, so that comparisons of the groups’ outcome variables would measure the effect of the treatment.

However, it is reasonable to believe that subjects who experience childbearing events might be self-selected, and so large differences may exist between women experiencing a new birth and those who do not on observable as well as unobservable covariates, which can lead to severe bias in the estimates of treatment effects.

Tables 9.3 and 9.4 show some descriptive statistics for the observed background variables separately for women who experience a childbearing event and women who do not. Table 9.3 presents, for each continuous covariate, the mean, the standard deviation, and the standardized percentage difference, defined as the mean difference between women who experience a childbearing event and women who do not, as a percentage of the standard deviation: $[100(\bar{x}(1) - \bar{x}(0))]/\sqrt{(s^2(1) + s^2(0))/2}$, where $\bar{x}(1)$ and $\bar{x}(0)$ are the sample means in the childbearing and no-childbearing groups, and $s^2(1)$ and $s^2(0)$ are the corresponding sample variances.

Table 9.4 shows, for categorical covariates, the proportion of women in each category in the two groups defined by the childbearing status, Z , as well as the absolute differences in percentage (third column).

As we can see in Tables 9.3 and 9.4, there exist considerable differences between women who experience a childbearing event and women who do not: sixteen of the continuous covariates have standardized differences larger than 10%; and the distributions of most of the categorical variables appear to be substantially different in the two groups of women. These differences indicate the possible extent of the bias when comparing outcomes between the two groups of women due to the different distributions of observed covariates. Therefore, before estimating the causal effect of interest we have to think clearly about the correct way to adjust for the systematic differences in background characteristics.

9.6.2 Econometric Framework

In our non-experimental context, because treatment and outcome can be endogenous, an identifying assumption is needed to consistently estimate the treatment effects of interest. We assume that assignment to treatment, Z , is independent of the outcome for untreated units, $Y(0)$, conditional on the covariates, X ; and that the probability of assignment is bounded away from one. Formally, for all x in the support of X ,

1. (*Unconfoundedness*) Z is independent of $Y(0)$ conditional on $X = x$;
2. (*Overlap*) $Pr(Z = 1 | X = x) < 1 - \eta$, for some $\eta > 0$.

Table 9.3 Means (standard deviations), and standardized differences in per cent for continuous covariate in both treatment groups before matching

Covariate	Childbearing				Standardized Difference (%) ^b
	No		Yes		
	mean	(s.d.)	mean	(s.d.)	
<i>Demographic variables</i>					
No. of adults	3.473	(1.399)	3.673	(1.484)	13.9
No. of children under 2 years	0.217	(0.477)	0.364	(0.573)	28.0
No. of children 3–6 years old	0.334	(0.576)	0.561	(0.742)	34.2
No. of children 7–10 years old	0.407	(0.612)	0.224	(0.501)	–32.6
No. of children 11–14 years old	0.504	(0.666)	0.224	(0.537)	–46.2
<i>Educational attainment</i>					
No. of household members with:					
Sub compulsory education	2.043	(1.562)	2.168	(1.850)	7.3
Compulsory education	1.559	(1.410)	1.963	(1.359)	29.1
Post compulsory education	1.332	(1.285)	0.916	(1.167)	–34.0
<i>Working status</i>					
No. of male workers	1.059	(0.700)	1.196	(0.679)	20.0
No. of female workers	0.952	(0.897)	1.000	(1.019)	5.0
No. of children workers	0.089	(0.406)	0.047	(0.253)	–12.4
<i>Measures of welfare</i>					
Deprivation index	0.358	(0.186)	0.403	(0.176)	25.1
Log of consumption expenditure ^a	9.844	(0.454)	9.841	(0.414)	–0.7
Log of income in wave 2002 ^a	9.007	(2.257)	8.827	(2.362)	–7.8
<i>Household head characteristics</i>					
Age of the household head	48.223	(11.667)	47.701	(15.252)	–3.8
Grade level of household head	9.724	(3.466)	8.643	(3.140)	–32.7
<i>Woman characteristics</i>					
Age	31.985	(10.365)	24.832	(5.509)	–86.2
Grade level	9.771	(2.686)	9.362	(2.473)	–15.8
No. of births until 2002	1.926	(1.720)	1.056	(1.204)	–58.6
Time since the last birth in months	96.998	78.526	55.548	41.011	–66.2

^a The consumption expenditure and income variables are equivalised using as equivalence scale $n_e = n^{\hat{\theta}}$, where $\hat{\theta} = 0.338$.

^b The standardized difference is the mean difference as a percentage of the average standard deviation: $[100(\bar{x}(1) - \bar{x}(0))] / \sqrt{(s^2(1) + s^2(0))/2}$, where for each covariate $\bar{x}(1)$ and $\bar{x}(0)$ are the samples means in the childbearing and no-childbearing groups and $s^2(1)$ and $s^2(0)$ are the corresponding sample variances.

The combination of these two conditions represents a relaxed form of strong ignorability introduced by Rosenbaum and Rubin 1983 (e.g., Abadie and Imbens 2002). The first assumption requires that all variables that affect the unobserved outcome and the likelihood of receiving the treatment are observed, and the second one requires that there is sufficient overlap in the probability of receiving the treatment among treated and controls. These conditions are strong, and in many cases may not

Table 9.4 Table of observed proportions and percent differences for categorical covariates

Covariate	Childbearing		Difference (%)
	No	Yes	
<i>Demographic variables</i>			
Region			
Coastal	0.280	0.187	9.3
Central	0.449	0.467	1.9
Mountain	0.129	0.168	3.9
Tirana	0.142	0.178	3.6
Area			
Urban	0.518	0.607	
Rural	0.482	0.393	9.0
No. of generations			
≤ 2	0.748	0.701	
> 2	0.252	0.299	4.7
<i>Household head characteristics</i>			
Gender			
Female	0.105	0.084	
Male	0.895	0.916	2.1
Marital status			
Unmarried	0.093	0.112	
Married	0.907	0.888	1.9
Working status			
Head does not work	0.232	0.271	
Head works	0.768	0.729	3.9
<i>Woman characteristics</i>			
Relation to household head			
Household head	0.043	0.009	3.4
Partner of the household head	0.549	0.393	15.7
Other	0.407	0.598	19.1
Religion			
No Muslim	0.232	0.121	
Muslim	0.768	0.879	11.0
Marital status			
Unmarried	0.309	0.131	
Married	0.691	0.869	17.8
Working status			
Woman does not work	0.439	0.514	
Woman works	0.561	0.486	7.5
Currently breast feeding			
No	0.940	0.869	
Yes	0.060	0.131	7.1

be satisfied. In many studies, however, researchers have found it useful to consider estimators based on these or similar conditions (see, for example, Rosenbaum and Rubin 1983; Heckman, Ichimura and Todd 1997; Dehejia and Wahba 1999; Becker and Ichino 2002).

In the present setting, the most critical assumption is the first one. Generally speaking, it may be problematic to interpret the birth of a child as a treatment of a

household because fertility may be affected by many unobservable and unobserved variables.

In our study, unconfoundedness might be violated for reasons related to the demand and supply for children and the cost of fertility control. Some characteristics of the area where the woman lives, such as the presence and the nearness of care facilities, and the possibility of reaching family planning services, which might lower the cost of fertility control, might affect fertility, making it easier for a woman to control her fertility. These same area-specific characteristics might also affect the subsequent individual living standards. Analogously, some individual unobserved characteristics (e.g., psychological cost related to contraception) might affect the propensity to give birth to a new child and, at the same time, improve (or worsen) the individual living standards.

Despite these potential confounders, in our study, we have carefully investigated which variables are most likely to confound any comparison between treated and control units, and so we believe that the assumption that all relevant variables are observed may be a reasonable approximation (e.g., Aassve, Mencarini and Mazzucco 2005 and 2006). Moreover, any alternative approach which does not rely on unconfoundedness, while allowing for consistent estimation of the causal effects of interest, must make alternative untestable assumptions, which are even more difficult to justify. Whereas the unconfoundedness assumption implies that the best matches are units that differ only in their treatment status, but otherwise are identical, alternative assumptions may implicitly match units that differ in the pretreatment characteristics. For instance, the technique of instrumental variables is sometimes considered as an alternative to assuming unconfoundedness, but in our setting the use of this approach is not particularly useful since finding valid instruments is difficult.

Under the Assumptions 1 and 2, the average treatment effect for the subpopulation with $Z = 1$ is equal to:

$$\begin{aligned} \tau &= E(Y(1) - Y(0) | Z = 1) \\ &= E[E(Y(1) - Y(0) | Z = 1, X = x)] \\ &= E[E(Y | Z = 1, X = x) - E(Y | Z = 0, X = x) | Z = 1] \\ &= E(\tau(x) | Z = 1), \end{aligned} \tag{9.3}$$

where the outer expectation is over the distribution of X conditional on $Z = 1$, and $\tau(x)$ is the average treatment effect for the subpopulation with $X = x$ and $Z = 1$. Therefore, under Assumptions 1 and 2, the ATT effect, τ , can be estimated by first estimating $\tau(x)$, for all x in the support of X for the treated (say X_1), and then averaging over the distribution of X conditional on $Z = 1$.

A usual way to control for differences in the groups' background variables is to specify a multiple regression of the outcome variable on the covariates, including an indicator variable for treatment status. When the model is well specified, the resulting estimated coefficient of the treatment indicator is a consistent estimate of the average causal effect of the treatment. Hahn (1998) showed that under the

unconfoundedness assumption the use of non-parametric series regression adjusting for all covariates can achieve efficiency bounds of the treatment effect.

However, the estimate can be badly biased when the model is not well specified as, for example, when the treatment is assumed constant, but instead it varies depending on the covariate values. In addition, when the data in the treated and comparison groups have different multivariate distributions of the covariates, the fitted regression involves extrapolations over much of the multidimensional covariate space (Rubin 1997). Such violations of model assumptions can be difficult to detect.

As an alternative to multiple linear regression, we can use matching methods to create groups of treated and control units that have similar background characteristics so that comparisons can be made within these matched groups. For each subject i , matching estimators impute the missing outcome by finding other individuals in the data whose covariates are similar but who were exposed to the other treatment. Specifically, the matching estimator we consider imputes the missing potential outcome, $Y_i(0)$, by using average outcomes for individuals with “similar” values for the covariates. We use matching with replacement, allowing each unit to be used as a match more than once.

A simple way to do this is imputing $Y_i(0)$ for a treated individual ($Z_i = 1$) with covariate values $X = x$ as the average of the outcomes we observe among controls with the similar covariate values $X = x$. When the available covariates for predicting acceptance of treatment are plentiful and/or continuous, such as in our study, the resulting matching estimator can be biased, since it may not be possible to come up with exact matches. Abadie and Imbens (2002) show that subject to some regularity assumptions, the simple matching estimator defined above is inconsistent if the number of (continuous) covariates available for matching exceeds two. In order to address this problem, they develop a bias-corrected matching estimator where the difference within the matches is regression-adjusted for the difference in covariate values.

In our study we apply their bias-corrected matching estimator. Let $J_M(i)$ be the set of indices for the matches for treated unit i that are at least as close as the M th match; i.e., for the set $\{j : Z_j = 0\}$, find the M nearest neighbors of i in the predictor space X , using a metric. The missing potential outcome, $Y_i(0)$, is then imputed as

$$\tilde{Y}_i(0) = \frac{1}{\#J_M(i)} \sum_{l \in J_M(i)} (Y_l + \hat{\mu}_0(X_i) - \hat{\mu}_0(X_l)),$$

where $\#J_M(i)$ is the number of elements of $J_M(i)$, and $\hat{\mu}_0(x)$ denotes the estimated regression function for the controls with covariate values $X = x$. The corresponding estimator for τ is

$$\tau_M^{bcm} = \frac{1}{N_T} \sum_{i:Z_i=1} (Y_i - \tilde{Y}_i(0)), \quad (9.4)$$

where *bcm* stands for bias-corrected matching.

Our motivation for using this bias-corrected matching estimator is twofold. First, it has better statistical properties than the simple matching estimator. Abadie and Imbens (2002) show that their bias-corrected matching estimator is consistent and has a sampling distribution that is asymptotically normal. In addition, they provide expressions for computing the variance of the bias-corrected estimator making it possible to test the significance of the treatment effect without relying on bootstrapping. Second, in our study, the bias-corrected matching estimator performs much better. It allows us to improve the balancing in the covariates after matching, and to obtain better results in terms of efficiency and robustness.²

9.7 Results

In this section, we apply both the regression and the Abadie–Imbens bias-corrected matching approaches to our subsample of panel women from Albania Living Standards Measurement Study (ALSMS) in the attempt to assess the impact of childbearing on economic wellbeing in Albania. Both the regression and matching approaches produce consistent estimates of the treatment effect only when we have controlled for all confounding covariates. When there are important confounding variables that have not been controlled for, either method can lead to biased estimates of treatment effects. It is important to keep in mind, however, that the two methods estimate the ATT effect under different assumptions. The simple linear regression model estimates the average treatment effect assuming that the treatment effect is constant across the subpopulation defined by the covariate values. Therefore, when the treatment effect is a non-constant function of the covariates, the regression model and the matching approach can achieve different estimates of the treatment effect even if each method produces unbiased estimates.

9.7.1 Regression Results

We first estimate the causal effect of interest using a multiple linear regression model of the form $Y | X, Z \sim N(\alpha + X\beta + \gamma Z, \sigma^2)$, where X denote the matrix of background covariates. We control for the geographic characteristics, the socio-demographic and economic variables and the pregnancy history. The regression

² The choice of estimating the causal effect of interest using the bias-corrected matching estimator proposed by Abadie and Imbens (2002) is the result of lots of preliminary analyses, concerning the selection of an appropriate set of pre-treatment matching variables, which allows us to consider the unconfoundedness assumption reasonable, and the comparison among different matching methods and matching estimators. Specifically, the goals of this preliminary work were: (1) investigating which variables were most likely to confound any comparison between treated and control units in such a way that the assumption that all relevant variables were observed might be a reasonable approximation; (2) choosing the matching method and the matching estimator which gave the best results in term of efficiency and robustness of the estimated effects.

model also contains a quadratic term for woman's age. Table 9.5 presents our regression results.

The results in Table 9.5 show that there is a statistically significant shift in the regression equation for women who give birth to a new child in comparison with those women who do not: the birth of a new child causes an increase of living standard by 8.838 Leks by month (with a standard error of 4.056 Leks).³ As a reference, note that the observed average monthly income for treated units is 28,632 Leks. Therefore, for the treated the estimated "counterfactual" average monthly income in the case of no-childbearing is 19,794 Leks (i.e., $28,632 - 8,838$). This means that having a new child would increase the average monthly income level by 44.6 percentage points (i.e., $100 \cdot 8,838 / 19,794$) with respect to the "counterfactual" situation of not having a new child. This result is surprising and puzzling. We worry about the scientific validity of the inference drawn from the regression model, which relies heavily on the correct specification of the functional form of the relationship (e.g., linearity) between the outcome and the covariates. In particular, the regression results might be driven by the specific way of extrapolating outcome values from the model (Dehejia and Wahba 1999; Rubin 1997). In our data, the observed average monthly income for controls is 21,447 Leks, which is higher than the estimated "counterfactual" average monthly income in the case of no-treatment for treated women (19,794 Leks), so that there is some sign that the regression results can be affected by the specific form of the model to extrapolate estimates of childbearing differences. In addition, the goodness of fit of our model appears to be very poor: the adjusted- R^2 is 8.5%. We could fit different specifications of the model, but we prefer to relax model assumptions by focussing on the matching approach.

9.7.2 Matching Results

The main purpose of matching is to re-establish the conditions of an experiment when no randomized control group is available. The matching method aims to construct the correct sample counterpart for the missing information on the outcome for treated individuals, had they not been treated by pairing each childbearing woman with women of the control group. Also matching estimators depend on the unconfoundedness assumption, but the diagnostics for matching analysis (checking for balance in the covariates) are much more straightforward than those for regression analysis and, enable the researcher to easily determine the range over which comparisons can be supported. Furthermore, the matching approach is more objective in the sense that the comparison group can be constructed without ever looking at the outcome variables. These two aspects of the analysis are inextricably linked in the linear regression analysis.

³ US\$1.00 equals 105.6 Leks.

Table 9.5 Regression results (Y = Equivalised household real (with respect to 2002) monthly income at the time of the third wave). Standard errors in parentheses*

	Adjusted R^2	0.085
	Overall F-statistic	5.490
	Sample size	1698
Covariates	Coef.	(s.e.)
Intercept	0.074	(16.552)
Childbearing status		
No childbearing		
Childbearing	8.838	(4.056)
<i>Household variables</i>		
Region		
Coastal		
Central	1.144	(2.278)
Mountain	-4.121	(3.223)
Tirana	9.078	(3.232)
Area		
Rural		
Urban	2.526	(2.600)
Deprivation index	-16.215	(6.090)
Income in wave 1 (per 1000)	-0.013	(0.017)
Consumption expenditure (per 1000)	0.563	(0.097)
No. of generations		
No more than 2		
More than 2	1.668	(2.787)
No. of adults	0.789	(1.571)
No. of children under 2 years	2.936	(2.556)
No. of children between 3 and 6 years	-0.030	(1.892)
No. of children between 7 and 10 years	-0.762	(1.791)
No. of children between 11 and 14 years	-0.449	(1.596)
No. of HH members with compulsory education	-0.856	(1.291)
No. of HH members with post compulsory education	-1.909	(1.683)
No. of men who work in Household	2.974	(1.942)
No. of women who work in Household	-2.198	(1.739)
No. of children who work in Household	-1.081	(2.469)
<i>Household head variables</i>		
Gender		
Female		
Male	0.472	(6.291)
Age	0.260	(0.153)
Marital status		
Unmarried		
Married	1.524	(5.498)
Grade level	0.642	(0.360)
Activity status		
Household head doesn't work		
Household head works	-1.375	(3.279)

Table 9.5 (continued)

	Adjusted R^2	0.085
	Overall F-statistic	5.490
	Sample size	1698
Covariates	Coef.	(s.e.)
<i>Woman variables</i>		
Age	-0.967	(0.899)
Square of Age	0.017	(0.013)
Relation to household head		
Head		
Partner of household head	-1.559	(6.974)
Other	-3.819	(7.996)
Religion		
No Muslim		
Muslim	1.266	(2.269)
Marital status		
Unmarried		
Married	1.786	(4.189)
Grade level		
Working status		
Woman doesn't work		
Woman works	3.086	(2.818)
Number of births until 2002	-0.912	(1.145)
Time since the last birth in months	0.002	(0.026)
Currently Breast feeding		
No		
Yes	-3.831	(4.618)

* For the categorical variables, the level which no coefficient value corresponds to, represents the baseline group.

The literature on matching methods is vast and growing. We apply the Abadie–Imbens bias-corrected matching estimator described in the previous section.⁴

Here, the biased-corrected matching estimator uses one match and the weighted Euclidean norm to measure the distance between different values for the covariates, with weights given by the inverse of the sample standard errors of the pre-treatment variables used in matching. The bias adjustment uses linear regression on all the pre-treatment covariates in Table 9.3 and 9.4, but not higher order terms or interactions. The bias correction is estimated using only the matched units in the comparison group.

⁴ All of the analysis is implemented by the use of the `nmatch` module in STATA (Abadie et al. 2001). This programme estimates the average treatment effects either for the overall sample or for the subsample of treated or control units using nearest neighbor matching estimators. The `nmatch` command implements the specific matching estimators developed in Abadie and Imbens (2002), including their bias-corrected matching estimators. The procedure `nmatch` allows individuals to be used as a match more than once. Compared to matching without replacement this generally lowers the bias but increases the variance.

9.7.2.1 Covariate Balance After Matching

To see how well the bias-corrected matching estimator performs in terms of balancing the covariates, Figs. 9.1 and 9.2 evaluate balance on observed continuous and categorical covariates, respectively, in the matched sample derived from the model.

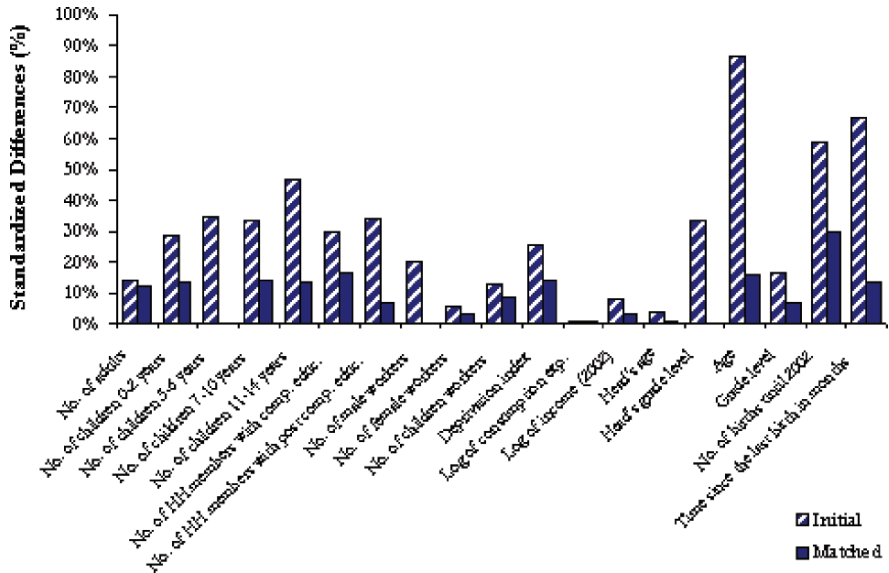


Fig. 9.1 Comparison of standardized differences (in %) for covariates between childbearing and no-childbearing women

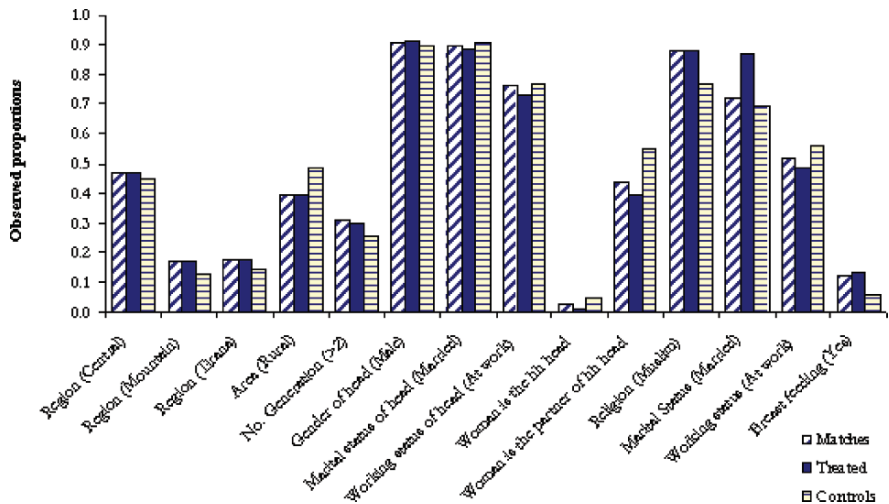


Fig. 9.2 Comparison of observed proportions for categorical covariates between childbearing and no-childbearing women

The matching performs very well in reducing the bias of the background covariates with moderate-large initial standardized differences. For instance, the initial standardized bias for “Age” is 86%, and the matching reduces it to 15%. In addition, exact matches have been obtained for three covariates, “Number of children between 3 and 6 years”, “Number of male workers”, and “Head’s grade level”, which have initial standardized differences equal to 34%, 20%, and 33%, respectively.

For the indicator variables “Region” and “Area” we specified exact matching, and for “Religion” exact matching is obtained. The other categorical variables are not matched exactly, but the quality of the matches appears very high: the average difference within the pairs is very small compared to the average difference between treated and comparison units before the matching.

These result suggest that the matched units can be considered sufficiently similar to the treated units. Therefore, provided the unconfoundedness assumption holds, one may proceed to estimate the causal effect of interest.

9.7.2.2 Estimated Causal Effects

Table 9.6 presents the estimated average causal effect of childbearing on income for the subpopulation of childbearing women using the Abadie–Imbens bias-corrected matching estimator. The estimate of the ATT effect is equal to 10,416 Leks, with a standard error of 9,441 Leks. Thus, as with the linear regression model, the matching analysis shows some evidence that giving birth to a new child increases living standards in Albania. In contrast with the regression analysis, however, the matching-based estimate of the ATT effect does not appear to be statistically significant.

There are several considerations behind the positive but negligible effect. First note that, using the Abadie–Imbens bias-corrected matching method, we estimate the “counterfactual” average monthly income for treated women in the case of no-treatment being equal to 17,658 Leks; this value is lower than the observed pre-treatment income level for the treated, which is equal to 21,826 Leks. Between wave 2002 and wave 2004 the observed average monthly income level for treated women decreases, but the difference does not appear to be relevant (see Table 9.2). On the contrary, it seems that had the treated women not experienced a childbearing event, their average monthly income level would have been much less than 21,826 Leks, being 17,658 Leks. Thus, childbearing appears to have a positive effect for women

Table 9.6 Means (standard deviations) for income (Leks) in both treatment groups after matching, and Average Causal Effect of childbearing on income (Leks) for the subpopulation of childbearing women in Albania

Estimand	Mean	(s.e.)
Income for childbearing	28,632	(90,470)
Income for matched no-childbearing women	17,658	(13,466)
Average treatment effect on childbearing women	10,416	(9,441)

who would have suffered from a stronger reduction in their monthly income level in the absence of childbearing.

Our estimated effect appears consistent with the descriptive statistics shown in Table 9.2. As we argued in Section 9.5, households where treated women reside seem to undertake a reorganization of the labor supply by increasing the number of male workers and decreasing the number of female workers. This descriptive result is in line with the positive ATT effect in the sense that treated households try to compensate the additional cost of a new member (the newly born child) and the possible loss of an active labor member (the woman who gives birth to the new child) increasing the number of male workers. In fact, it is reasonable that mothers will be completely inundated by the child bearing event whereas the other women of the family assist them with housework, while men focus on the market work. This insight is consistent with the results shown in Table 9.7, which presents the effect of childbearing on each income component. Although all the estimated effects are statistically negligible, confirming the global analysis, childbearing appears to have the highest effect on income from self-employment, suggesting that the time allocation within household is the most important means to face with income loss due to childbearing.

The positive sign of our estimated effect appears to be also fairly consistent with the Albanian welfare system. According to the Albanian Labor Code, “a woman is entitled to maternity leave provided she has been included in the social insurance scheme for the last 12 months and has been employed with an employment contract from the initial moment of pregnancy until the beginning of maternity leave. Maternity leave benefits are provided for one year, including a minimum of 35 days before delivery and 42 days after delivery. Women carrying more than one child during pregnancy are entitled to 390 days leave, including a minimum of 60 days before delivery and 42 days after delivery. Women in employment receive during maternity leave 80% of the average daily payment for the period before delivery and 50% of the average daily payment for 150 days after delivery, based on previous year’s average salary”. For women who are employers or the self-employed, the maternity benefit is equal to the basic old-age pension.

The Albanian Social Insurance System also offers birth grants to an insured person who is the mother or father of a newborn child, provided one of them has contributed for one year prior to the childbirth. The grant is however payable only once and the mothers have priority in eligibility, if insured. Birth grant is a lump sum of one-half of the minimum wage.

Table 9.7 Average Causal Effect of childbearing on equivalised income components (Leks) for the subpopulation of childbearing women in Albania

Outcome	ATT	(s.e.)
Wage	241	1515
Income from self-employed	9,729	9,460
Private transfers	-215	293
Public transfers	330	331

This system enables working mothers to make informed choices concerning the number and timing of their children. Specifically, maternity benefits and birth grants allow working mothers to recover from childbirth and to care for their newborn infants, providing them with protection against income loss due to childbirth and maternity. These law-based arguments tally with the positive effect of childbearing on public transfers shown in Table 9.7. However, this effect is much lower than the effect on income from self-employment, and it has a higher coefficient of variation (1.003 versus 0.972).

The law-based arguments, and the time re-allocation due to the birth of a new child jointly help us to explain the estimated positive effect. On the other hand, childbearing does not seem to substantially affect wage and private transfers (see Table 9.7).

In spite of the theoretical and practical positive aspects of the Albanian Social Insurance System, we have to keep in mind that parental leave and child support policies are mainly addressed to working women. In our sample, about half of the treated women worked in the week before the first interview, and the other half did not. This distribution of treated women by working status can at least partially explain the fact that the estimated effect is statistically non significant. In other words, we expect that the treatment effect is heterogeneous with respect to woman working status with a stronger and more significant effect for working women.

In addition, recent work on fertility behaviour in Albania during the nineties suggests that “traditionalism” and “norms” persist for the onset of family formation, whereas “modernity” and economic constraints impacts on the number of children, especially for third births and higher parities. For instance, using data from the Albanian Living Standard Measurement Survey, Aassve et al. (2006) show that formation in Albania is still traditional and having (at least) one child is still the norm.

These remarks suggest we should investigate the heterogeneity of the treatment effect along observable characteristics such as “woman’s working status” and “number of children”. Table 9.8 shows some sources of heterogeneity in the treatment effect.⁵ Most of the estimated effects are statistically negligible, confirming the global analysis, and the corresponding standard errors are sometimes fairly large (e.g., the estimated ATT effect for women with one child – equal to 28,480 Leks – has a standard error of 33,751 Leks). This result can be partially due to the small number of observations belonging to each subgroup.

Due to the small size of each subsample and the high sample variability, it is unlikely that we can draw robust inference on the size of the childbearing effect

⁵ All the ATT effects are estimated using the Abadie–Imbens matching estimator in its simple form. We do not regression-adjust the results because of the small size of each subsample defined by the marginal and joint values of the two covariates, “woman’s working status” and “number of children”, which we suspect being source of treatment-effect heterogeneity. For each subsample we first find one match for each treated woman using the weight Euclidean norm to measure the distance between units, with weights given by the inverse of the subsample standard errors of the matching variables. Then, we estimate the ATT effects separately in each subsample.

Table 9.8 Heterogeneity of the treatment effect (Standard errors in parentheses)

<i>Heterogeneity of the treatment effect with respect to “woman’s working status”</i>							
Covariate	Treated controls		Average income			ATT	(s.e.)
			Controls	Treated	Matched		
Woman does not work	55	698	22,008	22,307	20,500	1,807	(2,592)
Woman works	52	893	21,009	35,322	12,762	22,560	(19,758)
<i>Heterogeneity of the treatment effect with respect to “number of children”</i>							
0 children	45	481	19,388	23,689	16,139	7,550	(3,709)
1 child	32	137	22,847	50,006	21,526	28,480	(33,751)
More than 1 child	30	973	22,269	13,250	16,541	-3,291	(2,772)
<i>Heterogeneity of the treatment effect with respect to “woman’s working status” and “number of children”</i>							
Woman does not work and she has							
0 children	23	264	20,093	25,342	19,489	5,852	(5,570)
1 child	28	64	20,721	22,541	21,046	1,495	(3,631)
more than 1 child	14	370	23,597	17,022	22,371	-5,349	(4,014)
Woman works and she has							
0 children	22	217	18,530	21,960	12,363	9,597	(3,182)
1 child	14	73	24,711	85,318	24,915	60,403	(71,362)
more than 1 child	16	603	21,454	9,949	15,562	-5,613	(4,104)

in each subgroup of women from our heterogeneity analysis. However, keeping in mind this caveat, we can look at the results in order to obtain some insight on the possible presence of treatment-effect heterogeneity.

As we can see in Table 9.8, there appears to be a somewhat strong even if not much significant positive effect of a newly born child on income for working women, whereas this effect becomes small and totally negligible when we focus on women who do not work. This result appears to be consistent with the Albanian Social Insurance System.

Concerning the heterogeneity of the treatment effect with respect to the number of children, we find a significant positive childbearing effect for women who give birth to the first child. This effect appears to be larger for women who have the second child, but in this case it loses much of its significance. Finally, the effect of childbearing for women who already have at least two children is negative and much lower in absolute value. The heterogeneity of the childbearing effect with respect to the initial parity can be linked to the Albanian traditionalism in the family formation in the sense that the birth of the first infant is expected, and so the family is able to prevent income loss due to it. The birth of the second child is still quite normal in Albania, and the cost of the newly born child can be at least partially cushioned by re-using baby accessories and nursery equipment purchased for the

first child. The negative impact of a birth at higher parities (more than one) can be due to the fact that the average income level of treated women with two or more children is significantly lower than the income level for the other two groups of treated women.⁶

In order to understand better how treatment effect heterogeneity occurs, we also investigate the differences in the ATT effect across subgroups of women defined by the joint value of “woman’s working status” and “number of children”. Not unexpectedly we find a quite strong and highly significant childbearing effect for working women who give birth to the first child in the treatment spell (see Table 9.8). For women with a child, the working status seems to heavily affect the size of the positive effect, although a standard two-sided t-test suggests the two effects are not significant. For women with at least two children at the time of the first wave, there appears to be no relevant difference in the treatment effect with respect to the working status: we find a negative and barely significant effect of childbearing on living standards.

These results suggest that the treatment effect is highly heterogeneous with respect to “woman’s working status” and “number of children”. Therefore, it may be of substantive interest to investigate whether this heterogeneity in average treatment effects by “woman’s working status” and “number of children” is statistically significant or whether it is simply due to the sampling variability. We check whether the observed heterogeneity in the average treatment effects is statistically not negligible by regressing the average effect conditional on “woman’s working status” and “number of children” on the two covariates:

$$\tau(x_1, x_2) = \gamma_0 + \gamma_1 \cdot x_1 + \gamma_2 \cdot x_2 + \varepsilon,$$

where we denote with x_1 and x_2 “woman’s working status”, and “woman’s parity”, respectively. Note that we consider “number of children” as a continuous covariate in this regression model. In order to allow for heteroscedasticity of the average treatment effects, we use a variance-weighted least squares model, where the variance-weights are given by the square of the estimated standard errors of the ATT effects we computed in each subsamples. As we can see in Table 9.9, the regression model confirms that there exists relevant heterogeneity in the treatment effect along “woman’s working status” and “number of children”: all the estimated regression coefficients are statistically significant.

⁶ As an alternative, we could study the effect of a sequence of births. Lechner and Miquel (2005) developed a framework for causal analysis of sequences of treatments from a potential outcome perspective. A key feature of their framework is the identifying power of several different assumptions concerning the connection between the dynamic selection process and the outcomes of different sequences (Sequential Conditional Independence Assumptions). In our study, we prefer to focus on a static analysis, also because of the lack of information on covariates throughout the life history of each woman.

Table 9.9 Heterogeneity of the treatment effect: variance-weighted least squares results

	Goodness of fit χ^2	1539.150
	Model χ^2	2849.030
	Sample size	1698
Covariates	Coef.	(s.e.)
Intercept	4.230	(0.195)
Woman's working status*		
Not at work		
At work	1.765	(0.205)
Number of children	-3.038	(0.058)

* Women who do not work, whom no coefficient value corresponds to, represent the baseline group.

9.7.2.3 Sensitivity of the Estimated Causal Effects to the Equivalence Scale

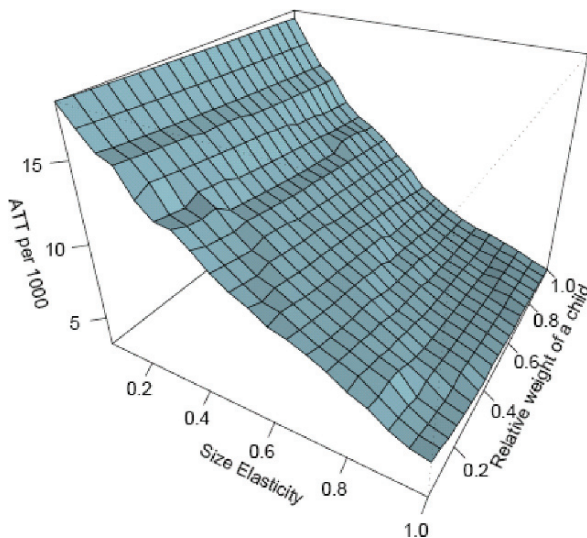
All the previous estimates rest on the plausibility of our income-based measure of wellbeing as proxy for poverty, which has been adjusted for differences in household size and composition using an equivalence scale.

We estimated the equivalent scale implied by the data using a variation of the well-known Engel method as described in Lanjouw and Ravallion (1995). Unfortunately, this method has some limitations. Gibson (2002) showed that Engel estimates of size economies are large when household expenditures are obtained by respondent recall but small when expenditures are obtained by daily recording in diaries. This results suggest that the Engel method could not give robust empirical estimates of scale economies, which should not depend on the method used to gather expenditure data. In our study, food consumption was collected by means of a 14-day diary, so we could expect that our estimate of size elasticity ($\theta = 0.338$) is biased downwards.

In addition, the assumption that the food share is an inverse welfare measure across household types, underlying the Engel method, does not always make sense. For instance, consider a larger household with the same per capita expenditures as a smaller household. If there are scale economies, the larger household is better off. Thus, according to Engel's second law, the larger household should have a lower food share. But a decline in the food share with constant per capita expenditures can occur only if there is a decline in food spending per person. It is very unlikely that people who are better off would spend less on food, especially in mid-low income countries where nutritional needs are not being met.

Given these conceptual and empirical problems with the Engel method, it seems important to carry out sensitivity analyses to see whether any conclusions reached previously using our measure of wellbeing are overturned. Our sensitivity analyses is based on Equation (9.1), trying different values of α and θ . Specifically, we approximated the continuous function (9.1) with a discrete function on a grid of points: we computed the equivalence scale (9.1) at a set of 20×20 evenly spaced values, (α_j, θ_j) , that cover the range of the parameter space of α and θ – that is $[0, 1] \times [0, 1]$. Then, for each $j = 1, \dots, 400$, we equivalised the household total

Fig. 9.3 Average treatment effect on treated by relative weight of a child and size elasticity



income using $n_{e,j} = (A + \alpha_j \cdot K)^{\theta_j}$ as equivalence scale, and re-estimated the ATT effect of interest.

As we can see in Fig. 9.3, the estimates of the average treatment effect appear to decrease almost monotonically with respect to the relative cost of a child, α , and the size elasticity, θ , ranging from 3,064 Leks (with a standard error of 3,267 Leks) – which corresponds to $\alpha = \theta = 1$ – to 18,113 (with a standard error of 15,313 Leks) – which corresponds to $\alpha = \theta = 0.05$.

This descending trend also appears looking at the two marginal functions in Fig. 9.4. Examining the trend of the ATT effects with respect to the relative cost of the child when the size elasticity is fixed at its estimated value $\hat{\theta} = 0.338$ (Fig. 9.4(a)), we see that our estimated causal effect, equal to 10,416 Leks, is the lowest. This means that if the assumption that adults and children have the same weight (equal to 1) does not hold, our estimated average treatment effect would underestimate the real treatment effect. Finally, Fig. 9.4(b) – which shows the distribution of the ATT effect as function of the size elasticity, θ , when the relative cost of the child, α , is fixed to 1 – suggests that our estimated size elasticity could be actually biased downwards, implying an enlargement of the real causal effect.

Our sensitivity analysis allows us to make clear two important remarks. First, all the estimates of the ATT effect we obtain ranging α and θ between 0 and 1 appear to be positive and statistically negligible⁷ – confirming the result reached previously; therefore, we are safe to say that our poverty estimates are not heavily affected by the adult equivalence weights that we chose. Second, the sensitivity analysis supports the conclusion that having an additional child has a non-negative effect on the living

⁷ The standard errors are omitted. However, their values along with further details are available on request from the authors.

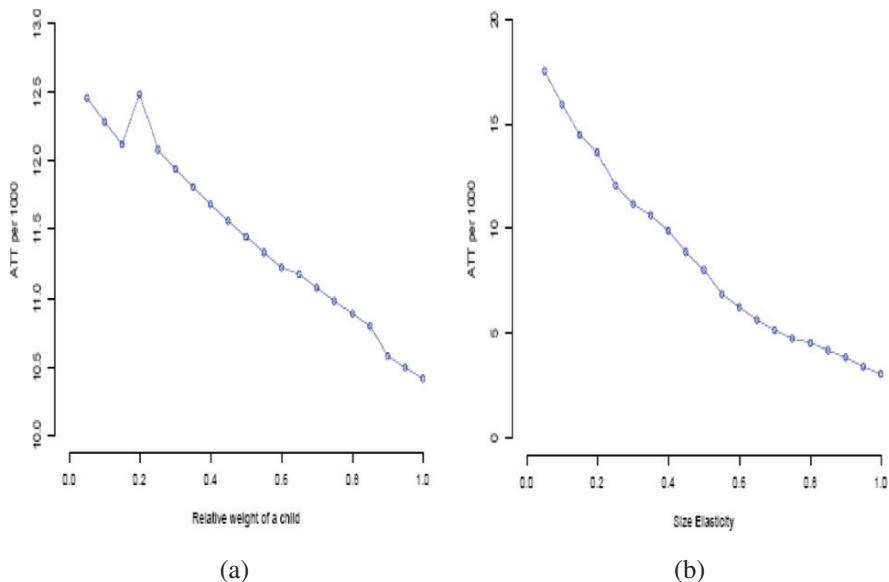


Fig. 9.4 Average childbearing effect on treated: (a) by relative weight of a child (size elasticity, θ , equals to 0.338); (b) by size elasticity (relative weight of a child, α , equals to 1)

standards in Albania, although our data seem to be unable to identify the size of this effect.

9.8 Conclusions

This paper evaluates whether and to what extent a childbearing event changes economic wellbeing for Albanian women. We use a panel sample of women drawn from the Albania Living Standard Measurement Study. Studying the causal relationships between poverty and fertility involves several crucial issues. First, a suitable measure of economic wellbeing is developed. Second, an appropriate econometric methodology is chosen, which works correctly with longitudinal information and takes into account that variation in fertility can be endogenous with respect to wellbeing. We use an income-based measure of wellbeing adjusted for household heterogeneity applying an equivalence scale. We estimate the equivalent scale from the data assuming that the number of adult equivalents in a household is given by the household size to the power of the size elasticity. Following Lanjouw and Ravallion (1995), the implied size elasticity from the Engel curve estimation in the ALSMS is 0.338. We then identify the causal effect of a childbearing event on our measure of monetary wellbeing applying both a linear regression model and the Abadie–Imbens bias-corrected matching estimator. Both approaches lead to a positive effect of childbearing on living standards, but whereas the regression model suggests that this effect is highly significant, the Abadie–Imbens bias-corrected

matching approach shows a negligible and insignificant effect. The regression results are most likely driven by the specific way of extrapolating outcome values from the model, thus preference is given to the results drawn from the Abadie–Imbens bias-corrected matching estimator, which leads to an average causal effect of 10,416 Leks (*s.e.* = 9, 441) for childbearing women. This effect seems to be mainly driven by the effect of childbearing on income from self-employment. We find that the treatment effect is fairly heterogeneous along observable characteristics such as woman’s working status and woman’s parity. Because of the high sample variability and the small number of observations of each subgroup of women defined by the marginal and/or joint values of the two covariates, it is difficult to draw clear insights on the size of the effects in each subsample. However, our heterogeneity analysis casts considerable doubt on the hypothesis that the average effect conditional on the covariates is identical for all subpopulations.

All these results rest on the plausibility of our income-based measure of well-being as a proxy for poverty, which depends on the estimated equivalence scale. In order to investigate the sensitivity of our results depending on the way in which household size and household composition is taken care of, we re-estimated the ATT effect using different equivalence scales, that is, different values of the parameters α , the weight for a child relative to an adult, and θ , the size elasticity. This sensitivity analysis finds that in Albania the estimated ATT effect is robust with respect to the estimated equivalence scale: all the estimates of the ATT effect appear to be positive and not significant.

There are two main directions for future research. The first is to extend this study by using other measures of wellbeing including multidimensional measures (such as deprivation indices) and subjective measures. Secondly, it is of considerable interest to analyze the conditional distribution of the difference between the two potential outcomes ($Y(1) - Y(0)$) given a childbearing event ($Z = 1$) as a whole, instead of focussing on its expected value as we have done in this paper.

Acknowledgements We are grateful to all the participants at the project “Poverty Dynamics and Fertility in Developing Countries” for their support and encouragement. Special thanks are due to Fabrizia Mealli and Steve Pudney for their insightful suggestions and discussions. We also thank Arnstein Aassve and Letizia Mencarini for their detailed comments which have improved the paper.

References

- Aassve, A., A. Gjonca and L. Mencarini (2006). The highest fertility in Europe for how long? The analysis of fertility change in Albania based on individual data. *European Population Conference*.
- Aassve, A., L. Mencarini and S. Mazzucco (2005). Childbearing and well-being: a comparative analysis of European welfare regimes. *Journal of European Social Policy* 15 (4): 283–299.
- Aassve, A., L. Mencarini and S. Mazzucco (2006). An empirical investigation into the effect of childbearing on economic well-being in Europe. *Statistical Methods and Applications* 15 (2): 209–227.
- Abadie, A., D. Drukker, J. Herr Leber and G.W. Imbens (2001). Implementing matching estimators for average treatment effects in Stata. *The Stata Journal* 1 (1): 1–18.

- Abadie, A. and G.W. Imbens (2002). Simple and bias-corrected matching estimators. *Mimeo*. Department of Economics, UC Berkeley.
- Becker, G. (1985). Human capital, efforts, and the sexual division of labor. *Journal of Labor Economics* 3 (1): 33–S58.
- Becker, G. (1991). *A Treatise on the Family*. Enlarged Edition. Cambridge: Harvard University Press.
- Becker, S. and A. Ichino (2002). Estimation of average treatment effects based on propensity scores. *The Stata Journal* 2 (4): 358–377.
- Birdsall, N.M., A.C. Kelley and S.W. Sinding (2001). *Population Matters: Demographic Change, Economic Growth, and Poverty in the Developing World*. Part III: Fertility, Poverty, and the Family. Oxford: Oxford University Press.
- Cigno, A. (1991). *Economics of the Family*. Oxford: Clarendon Press.
- Cigno, A. (1983). On Optimal Family Allowances, Oxford University Paper, 35, 13–22.
- Cigno, A., Luporini, A. and Pettini, A. (2000). Transfers to families with children as a principal-agent problem *Journal of Public Economics*, 87, 1165–1177.
- Dehejia, R. and S. Wahba (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association* 94: 1053–1062.
- Ermisch, F. (2003). *An Economic Analysis of the Family*. Princeton: Princeton University Press.
- Gibson, J. (2002). Why does the Engel method work? Food demand, economies of size and household survey methods. *Oxford Bulletin of Economics and Statistics* 64 (4): 341–360.
- Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica* 66 (2): 315–331.
- Heckman, J., H. Ichimura and P. Todd (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training program. *Review of Economic Studies* 64: 605–654.
- International Monetary Fund (2004). *The world economic outlook. The global demographic transition*. Tech. Rep., International Monetary Fund.
- International Monetary Fund (2005). *Albania: Selected issues and statistical appendix*. Tech. Rep., International Monetary Fund.
- Lanjouw, P. and M. Ravallion (1995). Poverty and household size. *Economic Journal* 105 (433): 1415–1434.
- Lechner, M. and R. Miquel (2005). Identification of the effects of dynamic treatments by sequential conditional independence assumptions. Working paper 2005–17, Department of Economics, University of St. Gallen.
- Rosenbaum, P.T. and D.B. Rubin (1983). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association* 79: 516–524.
- Rubin, D.B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine* 127: 757–763.
- Schoumaker, B. and D. Tabutin (1999). Relationship between poverty and fertility in southern countries. Knowledge, methodology and cases. Working paper 2, Department of science of population and development, Université catholique de Louvain.
- United Nations Development Programme (2003). *Human development report*. Report, UNDP.
- World Bank (2004). *Making transition work for everyone: Poverty and inequality in Europe and Central Asia*. Report, World Bank.

Chapter 10

Causation and Its Discontents

Herbert L. Smith

It is impossible to escape the impression that population scientists commonly use false standards in adducing causation – that they seek to make claims about the power of their research in elucidating cause and effect and admire similar claims in others, and that they mis-estimate the true values of important causal parameters. And yet, in making any general judgment of this sort, we are in danger of forgetting how variegated the human population and the mental constructs associated with its apprehension are.¹

When Preston (1993) surveyed the “contours of demography” – its role among the social sciences, its methods and orientations, and promising research areas for the future – causation figured hardly at all. The specific term shows up in but one sentence: “Because of pressures from the scientific community, it is likely that more surveys will be longitudinal in design, and thus will provide better opportunities for sorting out issues of causation” (p. 597). The reference to longitudinal data nicely anticipates the many fixed-effects models that now proliferate, where multiple observations over time are held to be the key to controlling for unobserved variation, hence causal inference.² On the other hand, the offhandedness of the observation – “*sorting out* issues of causation” [emphasis added] – suggests that the problem is (or was) not of the highest order for the field. A similar impression attaches to the use of the word “cause”, which features primarily in the demographic chestnut “causes and consequences”, of “population change”, or “variation in fertility, mortality, and

H.L. Smith (✉)

Population Studies Center, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA, USA
e-mail: hsmith@pop.upenn.edu

¹ With apologies to Freud (1961, p. 11).

² “Every empirical researcher knows that randomized experiments have major advantages over observational studies in making causal inferences. . . . It turns out . . . that with certain kinds of nonexperimental data we can get much closer to the virtues of a randomized experiment. . . . [B]y using the fixed effects methods . . . it is possible to control for all possible characteristics of the individuals in the study – even without measuring them – as long as those characteristics do not change over time. . . . [T]his is a powerful claim, . . . one that I will take pains to justify. . . . What is . . . remarkable is that fixed effects methods have been lying under our noses for many years” (Allison 2005, pp. 1–2).

migration patterns” (p. 594); there is no great sense of fundamental epistemological issues related to inferring causation.

Crimmins (1993), writing contemporaneously under the same warrant, also saw a relationship between demographers’ interest in causation and demands for new data, which had “grown increasingly complex to correspond to increasing complexity in the causal models underlying the demographic behavior we wish to understand” (p. 582). She also made a distinction between the use of causation in formal and social demography:

Formal demographers in the future will concentrate increasingly on models that incorporate the entire causal process of population change. This greater emphasis on causation, coupled with the expanding application of formal demography to new substantive areas will move the work of formal demographers increasingly closer to that of social demographers (p. 584).

The causal models of social demographers to which Crimmins (1993, p. 585) was referring were path models, or structural equation models (e.g., Duncan et al. 1972). These models had received mixed reviews when vetted from the standpoint of the counterfactual model of causation, since the involved arithmetic that transmutes correlations into path coefficients does not in and of itself render a *causal* model, absent some stipulations that lie outside the arithmetic:

The essential point . . . about [path] diagrams is that they are easily interpreted in terms of Rubin’s model when they are not causally meaningful. The causal model literature has not been careful in separating meaningful and meaningless causal statements and path diagrams. . . .

(Holland 1986, p. 958).

Freedman (1987) was less equivocal in his condemnation of the causal interpretations that putatively attached to such models (e.g., “Estimating nonexistent parameters cannot be very fruitful” [p. 125]), and Duncan himself had apparently long since recanted, labeling his work on structural equation models (Duncan 1975) “a big mistake” (Berk 2004, p. xvii).

The most prominent rehabilitation of structural equation models within a modern of theory of causation owes to Pearl (2000, esp. Chapter 7). Winship and Harding (2008) provide an excellent exegesis of these ideas in their application to the relationship between the canonical demographic accounting categories of age, period, and cohort and variation in political alienation in the United States (cf. Kahn and Mason 1987). There is a path- or graph-analytic model that offers a resolution to the well-known identification problem (Mason et al. 1973) caused by the linear relations among the constructs “age”, “period”, and “cohort”. Identification is largely with reference to the specification of the variables that intervene in a path diagram between age, period, and cohort and political alienation, the outcome variable. These are variables such as church attendance, employment, and education. I continue to wonder (Smith 2008, p. 294, 1997, pp. 333–334) whether it makes any sense to talk about variables such as age and cohort as *causes*, in any meaningful sense, since how do you change them?; and have attempted to argue (Smith 2003, pp. 464–466) that the manipulation criterion (Holland 1986, p. 959) is more blessing than curse. But

there are plenty of folks writing from both within the modern orthodoxy of potential outcomes (Winship and Sobel 2004, pp. 484–485) and without (Ní Bhrolcháin and Dyson 2007, p. 3) who can tell you why things that you cannot manipulate are causes nonetheless.

When Caldwell (1996) surveyed the relationship between demography and the social sciences, causation and causal models were not part of the equation. There were the standard anodyne references to the “causes and consequences” of population growth (p. 306) and population change (p. 307). Certainly demography was evolving under an interest in causation: “a greater emphasis on social causation” (p. 328); a move away from “‘social-bookkeeping’ . . . as a hankering for causal explanation and theory developed” (p. 329); similarly, “many demographers felt that the advances in the study of causation with regard to fertility could be, and needed to be, duplicated with regard to mortality” (p. 324). But, in general, the treatment is that of someone who knows a cause when he sees it (e.g., Caldwell et al. 1988) and it would seem to have been, for example, disciplinary predilections and prejudices, not formal understandings of causation, that had to date hindered “the use of anthropological approaches and concepts to study the nature and causes of demographic [behavior] . . .” (Caldwell 1996, p. 327).

A companion paper surveying the history of formal demographic models (Coale and Trussell 1996) makes no reference to their causal epistemological basis and, indeed, the search for causes is explicitly located outside of these models: The models are descriptive and were never intended to be anything else: “No deep theory, or even shallow theory, underlies the search for empirical regularities. In contrast, the discovery of empirical regularities can simulate the search for underlying causes” (p. 483).

Morgan and Lynch (2001) surveyed demography with an emphasis on data and methods. They make a strong case for the scientific status of demography less on the sophistication of data and methods than empirically, based on the cumulative salience of research papers. In this respect demography looks more like physics and chemistry, less like the (other) social sciences (*pace* Caldwell 1996). Causal modeling and causal thinking again factor only glancingly in the assessment of the field and its scientific status. This is the first survey of demography to highlight explicitly the type of causal models that feature so prominently in the current volume:

[T]he new wave of demography is more ambitious, aiming toward causal modeling at the individual level. The primary microlevel theories are social–psychological or microeconomic, and the method of choice is based in econometrics

(Morgan and Lynch 2001, p. 46).

The specific ideas underlying these methods – the potential outcomes framework and the concomitant primitive definition of a causal effect – are not mentioned, which is a shame, since this was a lost opportunity to reinterpret the role of such thinking in the history of demography (if not the role of demography in the history of thinking about causation). A search for the word “cause” in Morgan and Lynch (2001) turns up a number of references to cause-deleted life tables. Is this a matter of an atavistic use of a term that has now evolved in the direction of a

more precise, more developed scientific meaning? Hardly. “The construction of [a cause-deleted life] table . . . involves a thought experiment in which we ask ‘what would happen if . . .’” (Preston et al. 2001, p. 80). This is a clear counterfactual: life expectancy as observed in the U.S. in 1964 (for example) and life expectancy *if there were no mortality due to cancer* (for example); Keyfitz (1977) estimated the latter as 3% greater than the former. The comparison is at the population level, the effect of a cause (Holland 1986, p. 945), or the difference in life expectancy outcomes under two alternative treatment values {current rates of cancer, no cancer}.

At the individual level, the parallels are more complicated. In the simplest model for mortality, each death is ascribed to a single cause. This is closer akin to the cause of an effect: We observe a death and we ascribe it uniquely to a cause. Sadly, death is a constant; everyone will die. Thus the effect – the variation in response – is in *time to death*. In the theoretical framework for *competing risks*:

It is imagined that each individual at birth is endowed with a set of cause-specific ‘times due to die.’ The actual observed time, of course, is the minimum of these, because a person who dies from one cause cannot later die from another

(Namboodiri 1991, p. 120).

This is precisely the potential outcomes framework (Holland 1986, p. 946) and a nice statement of what Holland (1986, p. 947) called the Fundamental Problem of Causal Inference: that causal effects are defined in terms of alternative outcomes for a given unit, but that for any unit, only one outcome can be observed. The standard computational methods for cause-deleted life tables are quite recondite (e.g., Preston et al. 2001, pp. 80–84) and can be read as alternative assumptions for the identification of otherwise unobservable potential mortality outcomes. And this is without a full consideration of the problem of population heterogeneity, since “it is now widely recognized that eliminating one disease may yield an increase in deaths from another, due to comorbidity” (Morgan and Lynch 2001, p. 46). People who are spared death from cancer are not only exposed to the hazard of mortality from other causes, but the causes are likely correlated in their effects. This has been modeled with reference to manifest data on the secondary causes of death that appear on death certificates (Manton and Poss 1979) and to unobserved heterogeneity, the posited but unobserved distribution of frailty, or individual-specific probability of dying (Keyfitz and Littman 1979; Vaupel et al. 1979).

This modest claim on behalf of the causal thinking behind cause-deleted life-tables does cast a different light on what might be called the evolutionary progression of demography: that demography first mastered descriptive treatments of population data, life tables chief among them, and is now moving on toward scientifically valid models of cause and effect (e.g. Moffitt 2003, p. 448). But for me to dispute this view on the basis of a minor piece of scholarship – scholasticism, in truth – would be disingenuous. My discontent with this view is deeper: I think that there is no logical scientific hierarchy with description, however valuable, on the bottom, and causal analysis on the top. The weak version of this claim is empirical, that demography has long functioned well as a science (Morgan and Lynch 2001) absent anything other than a casual approach to the description/causation divide (which is, in essence, my reading of Caldwell’s [1996] synthesis). The strong version is

theoretical and methodological (Smith 1990, 2003, 2005, pp. 258–268). In brief: The formal treatment of causation has had the useful effect of undermining the received wisdom (e.g., Campbell and Stanley 1963), that experiments are the “gold standard” and inherently superior to observational studies. The emphasis on the heterogeneity of treatment effects (e.g., Heckman 2001, pp. 712–732) makes the very definition of a treatment effect, or a policy intervention, contingent on the distribution of traits within a population. This places demography, a field that has long had an interest in the accurate representation of populations and their heterogeneity, in a useful place. Ultimately, one wants a balance between emphases on representation and randomization (Kish 1987, Chapter 1).

This said, I can hardly gainsay the fact that there is a specter haunting demography, the specter of causal modeling (e.g., Bachrach and McNicoll 2003).³ This derives from some primitive but powerful thinking about the definition of causation in both statistics and economics. Holland (1986) was instrumental in calling attention to Rubin’s (1974) model and the statistical view of causation; Gelman and Meng (2004, especially Part I) is a good example of its substantial influence. Heckman (2005a) provides an exhaustive summary of what might be termed the “strong program” in economics. Morgan and Winship (2007) provide an excellent practical synthesis that reflects a close reading of both literatures, which developed in large measure in studied ignorance of one another.

Some future historian of science will have a field day with this topic, less in expositing the commonalities between the points of view, more in regaling readers with the whiggish efforts of each camp to document disciplinary priorities (e.g., Rubin 1990; *cf.* Heckman 2005a, p. 1, 2001, pp. 686–690). That the econometric and statistical approaches have so very much in common can be inferred, even before careful study of definitional and notational equivalences, by the massive amount of energy devoted – and spleen vented – in the pursuit of “boundary maintenance” (Gieryn 1999), of ruling science “in” and non-science “out”. How else to explain, for example, the intensity of Heckman’s (2005b) rebuke to a few points of reinterpretation scattered within an otherwise semblable tract (Sobel 2005)? It is well-established within sociological theory that dissent is more threatening to orthodoxy than is outright heresy (Coser 1956, pp. 86–93). The proto-sociological functionalist interpretation of the narcissism of minor differences is also apposite: “a convenient and relatively harmless satisfaction of the inclination to aggression, by means of which cohesion between the members of the community is made easier” (Freud 1961, p. 72). Although Freud’s pretensions to science have taken a beating (e.g., Webster 1995), it is hard not to marvel at his anticipation of the anthropology of early twenty-first century economics!⁴

Demographers are fortunate, therefore, to have Moffitt’s (2009, 2005, 2003) papers on causal analysis, which are orthodox in the presentation of the economic

³ With apologies to Marx and Engels (1976), all the more so since, as is evident from the papers in this volume, the powers of European social and statistical science are wholly receptive to the power of causal modeling.

⁴ But where to stand in throwing stones? Freud’s penchant for reading out dissenters, and the parallels between the psychoanalytic edifice and religion, are well attested (Webster 1995, pp. 359–362).

framework, with its emphasis on exclusion restrictions (as, for example, via the method of instrumentation – the search for variables that affect the assignment of subjects to treatments, but not the response subsequent to treatment), but tolerant with respect to parallel formulations within statistics, and catholic with respect to the role of this form of analysis within demography.

A very important point that emerges clearly (Moffitt 2005, pp. 94–96) is that causal inference in regression-type models is compromised when assignment of subjects to treatments is a function of unobserved differences between individuals *and* when there is variability across subjects in the effects of a treatment that is associated with who gets, or opts for, a treatment. Experienced researchers who are first exposed to the reinterpretation of regression analysis through the lens of the potential outcomes model for causal inference can be forgiven for imagining that it is a case of old wine in new models, since the problem of omitted variable bias is in general well apprehended. It is the emphasis on heterogeneity in treatment effects that is new, subtle, and powerful in its implications for inference:

For example, as cigarette prices vary across areas, the fraction of individuals who smoke will change as some individuals who would have smoked if prices were low instead choose not to smoke because prices are higher. With this Z_i , one can estimate the average effect of smoking of these “switchers”. Suppose that the variation in cigarette prices in the data induces a variation in the fraction who smoke from 30% of the population to 40%. The price variation allows the estimation of the average β for the 10% of the population who were affected by this variation. What cannot be estimated is the average β in the entire population because doing so would require having a Z_i that moved the fraction of smokers from 0% to 100%, thereby permitting the researcher to observe how Y changes as the entire population goes from not smoking to smoking, or vice versa

(Moffitt 2005, p. 96).

This is the nub of the possible disjuncture between the interests of population scientists and the predilections of those working within the formal structures of micro-econometrics. The emphasis on modeling and conceptualizing causal processes at the individual level is not without costs. Moffitt (2005) is admirably clear on this point: “[T]here is an important trade-off between the validity of a particular estimated causal effect and its generalizability, for pursuit of the former often leads to the loss of the latter, and vice versa” (p. 91). My own efforts (Smith 2003, especially pp. 462–463, 2005, pp. 258–268) have emphasized causal inference at the population level. There is no reason to imagine that one perspective need dominate the other *a priori*, but I do think that a pall mall rush toward micro modeling of causal processes may distract from the proper specification of estimands. What *can* be estimated is not necessarily what *should* be estimated. It depends on the way the problem is posed. A similar disquiet would appear to underlie the discontent manifest on behalf of demography by Ní Bhrolcháin and Dyson (2007), although I continue to find the potential outcomes framework congenial, even at the macro level.

I close with some comments that follow from Moffitt’s (pp. 96–97) discussion of structural versus reduced forms. The government increases the availability of contraceptives in an area. Fertility declines. This is the reduced form result. A structural mechanism concentrates on the pathways from more contraceptives to lower

fertility. Was lower fertility the result of more couples using contraceptives? Or did the policy effect some sort of attitudinal change that resulted in less sexual activity?

[S]ome economists have taken the view that if the only aim of the researcher is to know the effect of Z_i on Y_i . . . it does not matter what the mechanism or the channel of effect is . . . One only needs to assume that Z_i is not itself endogenous. This makes unbiased estimation much easier, but at the cost of not learning as much about the social process being studied”
(Moffitt 2005, p. 97).

The canonical method in support of the assumption that “ Z_i is not itself endogenous” is to assign subjects to treatments at random – to do an experiment! Discussions of the problem of causal inference from observational data often start with the premise that everything would be easier were an experiment feasible (e.g., Winship and Morgan 1999, p. 659), but the desire to understand *causal mechanisms* undermines the experimental model pretty quickly. My own interest in causation began (Smith 1990) with an attempt to make sense of a debate (Zeisel 1982a, 1982b; Rossi et al. 1982) over the proper interpretation of results from a randomized experiment (Berk et al. 1980; Rossi et al. 1980). Since then, there have been great advances, largely by greater precision in the definition of estimands. These include the causal interpretation of instrumental variables, by analogy with the experimental notion of “intention to treat” (e.g., Angrist et al. 1996; Gennetian et al. 2005). The generalization of these ideas, under the rubric of “principal stratification” (Frangakis and Donald Rubin 2002; Frangakis 2004; Rubin 2004), also has great promise for at least some problems in demography.

For a wide variety of reasons, it is very far from my intention to express an opinion upon the value of causal thinking in the field of demography. I have endeavored to guard myself against the enthusiastic prejudice which holds that causation is the most precious thing that we possess or could acquire and that its application puts us on a path that will necessarily lead to heights of unimagined perfection. One thing only do I know for certain and that is that our judgments of value follow directly our wishes for greater insight – that, accordingly, they are an attempt to support our illusions with arguments. Social scientists, including demographers, have gained control over the forces of calculation and computation to such an extent that with their help they would have no difficulty in estimating virtually any parameter. But who can foresee with what success and with what result?

References

- Allison, P.D. (2005). *Fixed Effects Regression Methods for Longitudinal Data Using SAS*[®]. Cary, NC: SAS Institute Inc.
- Angrist, J.D., G.W. Imbens and D.B. Rubin (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association* 91(434): 444–455.
- Bachrach, C. and G. McNicoll (2003). Introduction [to Causal Analysis in the Population Sciences: A Symposium. *Population and Development Review* 29(3): 443–447.
- Berk, R.A., K.J. Lenihan and P.H. Rossi (1980). Crime and Poverty: Some Experimental Evidence from Ex-Offenders. *American Sociological Review* 45(5): 766–786.

- Berk, R.A. (2004). *Regression Analysis: A Constructive Critique*. Thousand Oaks, CA: Sage Publications.
- Caldwell, J.C. (1996). Demography and Social Science. *Population Studies* 50(3): 305–333.
- Caldwell, J.C., P.H. Reddy and P. Caldwell (1988). *The Causes of Demographic Change: Experimental Research in South India*. Madison, WI: University of Wisconsin Press.
- Campbell, D.T. and J.C. Stanley (1963). *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand-McNally.
- Coale, A. and J. Trussell (1996). The Development and Use of Demographic Models. *Population Studies* 50(3): 469–484.
- Coser, L.A. (1956). *The Functions of Social Conflict*. New York: The Free Press.
- Crimmins, E.M. (1993). Demography: The Past Thirty Years, the Present, and the Future. *Demography* 30(4): 579–591.
- Duncan, O.D. (1975). *Introduction to Structural Equation Modeling*. New York: Academic Press.
- Duncan, O.D., D.L. Featherman and B. Duncan (1972). *Socioeconomic Background and Achievement*. New York: Seminar Press.
- Frangakis, C.E. (2004). Principal Stratification. In: *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family*, eds. Andrew Gelman and Xiao-Li Meng. Chichester, England: John Wiley & Sons Ltd.
- Frangakis, C.E. and Donald B. Rubin (2002). Principal Stratification in Causal Inference. *Biometrics* 58(1): 21–29.
- Freedman, D. (1987). As Others See Us: A Case Study in Path Analysis. *Journal of Educational and Behavioral Statistics* 12(2): 101–128.
- Freud, S. (1961). *Civilization and Its Discontents*. Translated and edited by James Strachey. New York: W. W. Norton & Company, Inc.
- Gelman, A. and X.-L. Meng (eds.) (2004). *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family*. Chichester, England: John Wiley & Sons Ltd.
- Gennettian, L.A., P.A. Morris, J.M. Bos and H.S. Bloom (2005). Constructing Instrumental Variables from Experimental Data to Explore How Treatments Produce Effects. In: *Learning More from Social Experiments*, ed. Howard S. Bloom. New York: Russell Sage Foundation.
- Gieryn, T.F. (1999). *Cultural Boundaries of Science: Credibility on the Line*. Chicago: The University of Chicago Press.
- Heckman, J.J. (2001). Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture. *Journal of Political Economy* 109(4): 673–748.
- Heckman, J.J. (2005a). The Scientific Model of Causality. In: *Sociological Methodology 2005*, ed. Ross M. Stolzenberg. Boston, MA: Blackwell Publishing.
- Heckman, J.J. (2005b). Rejoinder: Response to Sobel. In: *Sociological Methodology 2005*, ed. Ross M. Stolzenberg. Boston, MA: Blackwell Publishing.
- Holland, P.W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association* 81(396): 945–960.
- Kahn, J.R. and W.M. Mason. (1987). Political Alienation, Cohort Size, and the Easterlin Hypothesis. *American Sociological Review* 52(1): 155–69.
- Keyfitz, N. (1977). What Difference Would It Make if Cancer Were Eradicated? An Examination of the Taeuber Paradox. *Demography* 14(4): 411–418.
- Keyfitz, N. and G.S. Littman (1979). Mortality in a Heterogeneous Population. *Population Studies* 33(2): 333–342.
- Kish, L. (1987). *Statistical Design for Research*. New York: John Wiley & Sons.
- Manton, K.G. and S.S. Poss (1979). Effects of Dependency Among Causes of Death for Cause Elimination Life Table Strategies. *Demography* 16(2): 313–327.
- Marx, K. and F. Engels (1976) [1888, 1848]. *Manifesto of the Communist Party*, pp. 476–519 in their Collected Works, Volume 6, translated by Samuel Moore. New York: International Publishers.
- Mason, K.O., W.M. Mason, H.H. Winsborough and W.K. Poole (1973). Some Methodological Issues in Cohort Analysis of Archival Data. *American Sociological Review* 38(2): 242–258.

- Moffitt, R. (2003). Causal Analysis in Population Research: An Economist's Perspective. *Population and Development Review* 29(3): 448–458.
- Moffitt, R. (2005). Remarks on the Analysis of Causal Relationships in Population Research. *Demography* 42(1): 91–108.
- Moffitt, R. (2009). Issues in the Estimation of Causal Effects in Population Research, with an Application to the Effects of Teenage Childbearing. In: *Causal Analysis in Population Studies: Concepts, Methods, Applications*, eds. H. Engelhardt, A. Prskawetz and H.-P. Kohler. Ort: Verlag.
- Morgan, S.P. and S.M. Lynch (2001). Success and Future of Demography: The Role of Data and Methods. *Annals of the New York Academy of Sciences* 954: 35–51.
- Morgan, S.L. and C. Winship (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.
- Namboomdir, K. (1991). *Demographic Analysis: A Stochastic Approach*. San Diego, CA: Academic Press, Inc.
- Ní Bhrolcháin, M. and T. Dyson (2007). On Causation in Demography: Issues and Illustrations. *Population and Development Review* 33(1): 1–36.
- Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.
- Preston, S.H. (1993). The Contours of Demography: Estimates and Projections. *Demography* 30(4): 593–606.
- Preston, S.H., P. Heuveline and M. Guillot (2001). *Demography: Measuring and Modeling Population Processes*. Malden, MA: Blackwell Publishers Inc.
- Rossi, P.H., R.A. Berk and K.J. Lenihan (1980). *Money, Work, and Crime: Experimental Evidence*. New York: Academic Press.
- Rossi, P.H., R.A. Berk and K.J. Lenihan (1982). Saying It Wrong with Figures: A Comment on Zeisel. *American Journal of Sociology* 88(2): 390–393.
- Rubin, D.B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology* 66(5): 688–701.
- Rubin, D.B. (1990). Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies. *Statistical Science* 5(4): 472–480.
- Rubin, D.B. (2004). Direct and Indirect Causal Effects via Potential Outcomes. *Scandinavian Journal of Statistics* 31(2): 161–170.
- Smith, H.L. (1990). Specification Problems in Experimental and Nonexperimental Social Research. In: *Sociological Methodology 1990*, ed. Clifford C. Clogg. Cambridge, MA: Basil Blackwell.
- Smith, H.L. (1997). Matching with Multiple Controls to Estimate Treatment Effects in Observational Studies. In: *Sociological Methodology 1997*, ed. Adrian E. Raftery. Oxford, England: Basil Blackwell.
- Smith, H.L. (2003). Some Thoughts on Causation as It Relates to Demography and Population Studies. *Population and Development Review* 29(3): 459–469.
- Smith, H.L. (2005). Introducing New Contraceptives in Rural China: A Field Experiment. *Annals of the American Academy of Political and Social Science* 599: 246–271.
- Smith, H.L. (2008). Advances in Age-Period-Cohort Analysis. *Sociological Methods & Research* 36(3): 287–296.
- Sobel, M.E. (2005). Discussion: 'The Scientific Model of Causality'. In: *Sociological Methodology 2005*, ed. R.M. Stolzenberg. Boston, MA: Blackwell Publishing.
- Vaupel, J.W., K.G. Manton and E. Stallard (1979). The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography* 16(3): 439–454.
- Webster, R. (1995). *Why Freud Was Wrong: Sin, Science, and Psychoanalysis*. New York: Basic Books.
- Winship, C. and D.J. Harding (2008). A Mechanism-Based Approach to the Identification of Age-Period-Cohort Models. *Sociological Methods & Research* 36(3): 362–401.
- Winship, C. and S.L. Morgan (1999). The Estimation of Causal Effects from Observational Data. *Annual Review of Sociology* 25: 659–706.

- Winship, C. and M. Sobel (2004). Causal Inference in Sociological Studies. In: *Handbook of Data Analysis*, eds. M.A. Hardy and A. Bryman. Thousand Oaks, CA: Sage Publications Inc.
- Zeisel, H. (1982a). Disagreement over the Evaluation of a Controlled Experiment. *American Journal of Sociology* 88(2): 378–389.
- Zeisel, H. (1982b). Hans Zeisel Concludes the Debate. *American Journal of Sociology* 88(2): 394–396.

Index

Note: Locators indicating tables and figures are followed by *t* and *f* respectively.

A

- Aassve, A., 215, 224
- Abadie, A., 6, 202, 213, 216, 217, 220, 222, 224, 229, 230
- Abadie–Imbens matching estimator, 224
bias-corrected, 220
- Abbott, A., 2
- Abbring, J.H., 34
- Accelerated failure time (AFT) model, 114
- Adsera, A., 32
- Agresti, A., 86
- Albania, causal effect of childbearing on household income
background, 203–204
descriptive statistics, 209–211, 209*t*
Engel curve estimation of size elasticity using first wave of ALSMS, 208*t*
identifying causal effect of new birth
econometric framework, 212–217, 213*t*, 214*t*
quasi-experimental approach, 211–212
living standards measurement study, 204–205
measure of well-being, 205–209
Engel’s second law, 207
“size elasticity,” 207
microeconomic theory, 206
results
ATT effects, 223, 226
childbearing/no-childbearing women, 221*f*
covariate balance after matching, 221–222
estimated causal effects, 222–227, 222*t*, 223*t*
heterogeneity of treatment effect, 225*t*, 227*t*
matching results, 218–229
sensitivity of causal effects, 227–229, 228*f*, 229*f*
- Albanian Institute of Statistics (INSTAT), 204
- Albanian Labor Code, 223
- Albanian Living Standard Measurement Survey (ALSMS), 202, 204
- Albanian Social Insurance System, 223
- Allard, M.A., 193
- Allison, P.D., 96
- ALSMS, *see* Albanian Living Standard Measurement Survey (ALSMS)
- Ancillary time-dependent covariates, 89
- Andersen, P.K., 138, 139
- Anderson, S., 72
- Angrist, J.D., 14–15, 18, 21–22, 24, 26, 32, 111, 151–152, 239
- Arroyo, C., 31
- Ashenfelter, O., 112
- Associational models, *see* Statistical model
- ATE, *see* Average treatment effect (ATE)
- ATET, *see* Average treatment effect for treated (ATET)
- ATT, *see* Average effect of treatment on treated (ATT)
- Average effect of treatment on treated (ATT), 211
- Average treatment effect (ATE), 11, 16
- Average treatment effect for treated (ATET), 176–177

B

- Bachrach, C., 2, 237
- Back-door criterion (Pearl), 73
- Background knowledge, 62
- Baker, R.M., 159
- Barndorff-Nielsen, O., 70
- Barnett, M., 183
- Barnow, B., 27
- Basic causal model, 10–13

- Becker, S.O., 178, 214
 Becker G.S., 150, 157, 169, 171, 202
 Bedard, K., 150
 Berk, R.A., 234, 239
 Bijwaard, G.E., 5, 111, 113,
 127, 134
 Biliyas, Y., 134
 Birdsall, N.M., 201
 Björklund, A., 11, 17
 Blalock, H.M., 64, 84
 Blau, D.M., 170
 Blossfeld, H.-P., 5, 33, 83, 88–91, 94, 96–98,
 100–101, 106
 Blossfeld-Klijzing-Pohl-Rohwer study, 98–99
 estimates of transition from consensual
 union to marriage, 99*t*
 Blossfeld-Manting-Rohwer study, 97–98
 Bollen, K.A., 73
 Bound, J., 21, 159
 Brännäs, K., 115
 Bratti, M., 151
 Brooks-Gunn, J., 183
 Brown, M., 169
 Brown, S., 170
 Browning, M., 149
 Bumpass, L., 168
- C**
- Caldwell, J.C., 235, 236
 Caldwell, P., 235
 Cameron, A.C., 22, 26
 Campbell, D.T., 237
 Card, D., 23
 Carlson, M., 167, 171, 182, 193
 Carneiro, P.J., 17
 Causal analysis
 in population studies, 1–7
 in social sciences
 background knowledge, 61–63
 goals of causal analysis, 59–60
 probabilistic modeling, 63–64
 variation and regularity, 60–61
 Causal effects, 36, 134
 and econometrics, 33
 to estimate, *see* Instrumental variables (IV)
 Causal inference, models of, 84–88
 Causal mechanisms generating trends and
 variation, 1–2
 Causal model
 basic, 10–13
 constraints, *see* Background knowledge;
 Structural stability
see also Statistical model
- Causal relationships, empirical analysis of, 83
 consequences for explanations of, 101
 Causation, 83–84
 discontents of, 233–240
 as generative process, 83–84
 Blossfeld-Klijzing-Pohl-Rohwer study,
 98–99
 Blossfeld-Manting-Rohwer study,
 97–98
 Mills-Trovato study, 100
 models of causal inference, 84–88
 parallel and interdependent processes,
 88–90
 interdependent processes: system
 approach, 91–96
 substantial explanations, 100
 abortion/miscarriage/problem of
 conditioning on future events,
 104–105
 actors, probabilistic causal relations and
 hazard rate, 101–102
 diffuse marriage preferences and
 negotiation process, 102–103
 unobserved marriage deci-
 sions/observed rate of entry
 into marriage, 103–104
 “Causation as generative process,” 83
 Cherlin, A.J., 167, 170, 193
 Chevalier, A., 22
 Childbearing in teenage, causal effects in
 population research, *see* Teenage
 childbearing, causal effects in
 population research
 Childbearing status (*Z*), 212
 Child wellbeing, relationship dissolution on,
 168
 CIA, *see* Conditional Independence
 Assumption (CIA)
 Cigno, A., 150, 201, 202
 Clark, D., 183
 Cléroux, R., 72
 Coale, A., 235
 Cogswell, J.J., 181
 Coleman, J., 170
 Completely recursive system, 76
 complex systems and, 74–75
 3-component, 76
 with marginal independence, 77
 first 3 components, 75
 first 4 components, 75
 Conditional independence, 94–95
 Conditional Independence Assumption (CIA),
 34, 175–176

- Conditional model, 67–69
 essential features of, 68–69
 and exogeneity, 69–71
- Confounding, complex systems and completely recursive systems
 complex systems and completely recursive systems, 74–75
 confounders and confounding, 71–74
- Confounding variable, 71–72, 73
 SES as, 72*f*
- Connelly, R., 32
- Consequential manipulation approach,
 causation as, 85, 86–87
 drawbacks, 88
- “Contours of demography,” 233
- “Control variable” approach, 84
- Coser, L.A., 237
- Counterfactual approach, 3
- Courgeau, D., 88–91, 106
- Covariance structure model, 79
- Cox, D.R., 5, 83, 87, 100, 105, 113, 115, 120
- Cox, M., 183
- Crimmins, E.M., 234
- Currie, J., 21
- D**
- Dankmeyer, B., 150
- Data mining, *see* Exploratory data analysis
- Dawid, A.P., 37
- Defiers, 26
- Defined time-dependent covariates, 89
- De Finetti, B., 66
- Dehejia, R., 177, 184, 214, 218
- Demography
 and labour market choices, evaluation of causal relationship, 150
 and social sciences, relationship between (Caldwell, 1996), 235
- Deschênes, O., 150
- Descriptive models, *see* Statistical model
- Descriptive statistics for selected variables and selected subsamples, 48*t*
- Ding, W., 34
- Di Pino, A., 5, 149, 157
- Di Tommaso, M.L., 32
- Drewianka, S., 169
- Drobnìè, S., 89, 90
- Duncan, B., 234
- Duncan, O.D., 84, 234
- Durkheim, E., 59, 79
- Dynamic average treatment effect (DATE), 36, 40
- Dynamic causal model, 34
 basic structure of model, 35–36
 defining estimand: average causal effects, 36
 identification, 37–39
- Dyson, T., 235, 238
- E**
- Econometrics, 33
- Edin, K., 183
- Education-elasticity of Italian women’s labour market participation, 163*f*
- Eells, E., 92
- Elder, G.H., 87, 89
- Elster, J., 101
- Elwood, J.M., 72
- Engels, F., 237
- Engel’s second law, 207
- Engle, R.F., 62
- Ermisch, F., 202
- Evans, W.N., 22, 32
- Exogeneity
 and causality, 71
 in structural conditional model, 71
- Explicit causal perspective, 59
- Exploratory data analysis, 66
- Extrapolation, 16–17
- F**
- Fagan, J., 183
- Father involvement in childrearing,
 benefits of, 171
- Feichtinger, G., 1
- Fertility, measurement of, 201
- FFCWS, *see* Fragile families and child wellbeing study (FFCWS)
- Fisher, R.A., 86
- Fixed effect models, 86
- Florens, J.-P., 62, 70, 76
- Fragile families and child wellbeing study (FFCWS), 168
- Francesconi, M., 31
- Frangakis, C.E., 239
- Franklin, C., 86
- Freedman, D., 234
- Freedman, R.A., 100
- Freud, S., 233, 237
- Fricke, T., 2
- Front-door criterion, 74
- Fundamental problem of causal inference, 236
- Furstenberg, F.F., 193

G

- Gallagher, M., 169
 Garfinkel, I., 167
 Geling, O., 134
 Gelman, A., 237
 Generalized Accelerated Failure Time (GAFT)
 model, 112, 115–116
 endogenous covariates in, 116–118
 with endogenous variables is, 117
 Generative process, causation as, 87–88
 advantages, 88
 important aspects of, 92–94
 temporal lags and effect shapes, 93*t*
 Gennettian, L.A., 239
 Gergen, P.J., 181
 Geronimus, A., 21
 Geweke, J., 33
 Gibson, J., 227
 Gieryn, T.F., 237
 Gill, R.D., 34
 Goldfeld, S.M., 152
 Goldstein, J.R., 182
 Goldthorpe, J.H., 2, 84–88, 101
 Granger, C.W.J., 33, 84
Granger causation, 84
 Gruber, J., 21

H

- Haaga, J.G., 169
 Hahn, J., 215
 Han, A.K., 115
 Hannan, M.T., 89–91, 96
 Harden, B., 193
 Harding, D.J., 234
 Harknett, K., 182
 Heckman, J.J., 3, 10–11, 17–18, 26–27, 32–33,
 87, 111, 150, 156–157, 168, 173,
 176, 214, 237
 Hedström, P., 87, 101
 Heiland, F., 6, 167, 170, 173, 182
 Hendry, D.F., 62
 Hofferth, S.L., 170
 Hoffman, S., 21
 Holland, P.W., 2, 84–87, 91, 234, 236–237
 Horney, M.J., 169
 Hotz, J.V., 31, 32
 Hotz, V.J., 22, 150, 168
 Huinink, J., 88, 89, 91
 Hume, D., 60, 62, 79

I

- Ichimura, H., 214
 Ichino, A., 178, 214

- Illinois re-employment bonus experiment,
 126–133
 average unemployment durations: control
 group and (non-)compliers, 127*t*
 effect of bonus on length of unemployment,
 130*t*
 effect of bonus on quantiles of
 unemployment, 131*t*
 duration of BLACKS, 132*t*
 instrumental variable linear rank estimates
 for effect of bonus, 129*t*
 IVLR of
 descriptive statistics for
 control/claimant/ employer bonus
 group, 145*t*
 estimated $\hat{\epsilon}$ in GAFT model for bonus
 data, 148*t*
 regression coefficients (constant bonus
 effect), 146*t*
 regression coefficients (time-varying
 bonus effect), 147*t*
 Imbens, G.W., 6, 12, 14, 18, 22, 24, 34, 38–39,
 45, 151–152, 172, 192, 202, 213,
 216–217, 220, 222, 224, 229–230
 Instrumental variable approach, 32
 Instrumental variable estimation for duration
 data, 111–114
 endogenous covariates in duration models,
 114–115
 endogenous covariates in GAFT
 models, 116–118
 generalized accelerated failure time
 model, 115–116
 intuition for instrumental variable
 estimation, 118–120
 Illinois re-employment bonus experiment,
 application to, 126–133
 instrumental variable linear rank
 estimation, 120
 efficiency of IVLR estimator, 123–124
 estimation in practice, 125–126
 IVLR estimator, 120–123
 Instrumental Variable Linear Rank estimator
 (IVLR), 112
 Instrumental Variable method, 120
 Instrumental variables (IV), 10, 13–18, 86
 in econometrics, 117
 types of, 18–23
 Interdependent processes
 causal approach, 91–96
 causes and time-dependent
 covariates, 91

- principle of conditional independence, 94
 - time and casual effects, 92
 - unobserved heterogeneity, 92–93
 - joint determination of, 95
 - system approach, 90–91
 - disadvantages, 90–91
 - Internal time-dependent covariates, 89–90
 - International Monetary Fund (2004), 206
 - Italian women, labour market participation
 - age-elasticity of, 163*f*
 - probability of, 162*t*
 - Italy, female labour participation, demographic processes, 160*t*
 - age-elasticity of Italian women labour market participation, 163*f*
 - background, 149–151
 - complexity, reasons, 149
 - education-elasticity of, 163*f*
 - labour market participation probability of Italian women, 153, 162*t*
 - model specification, theoretical/methodological issues, 151
 - data, 158–159
 - discussion, 161–164
 - estimation results of reduced-form equations, 155*t*–156*t*
 - labour force participation, 159
 - model, 153–158, 154*t*
 - probability of marital dissolution, 157
 - results, 159–161
 - Survey on Household Income and Wealth, 158
 - “treatment effect estimate,” 153
 - “weak instruments,” use of, 153
 - Italy, geographical mobility in, 151
 - IVLR, *see* Instrumental Variable Linear Rank estimator (IVLR)
 - IVLR estimator, 120–123
 - asymptotic properties of, 122
 - defined, 121
 - efficiency of, 123–124
 - optimal weight function in, 123
- J**
- Jaeger, D.A., 159
 - Jaffee, S.R., 183
 - Jekielek, S.M., 171
 - Jenicek, M., 72
 - Johnson, W.E., 193
 - Johnston, J., 84
 - Joshi, H., 150
- K**
- Kahn, J.R., 234
 - Kalbfleisch, J.D., 89, 115
 - Katus, K., 100
 - Kelly, J.R., 92, 93
 - Kerlinger, F.N., 84
 - Kernel matching estimator, 177
 - Uniform/Epanechnikov/Gaussian, 177
 - see also* Parental separation on child health, new estimates on effects
 - Keyfitz, N., 236
 - Kilpeläinen, M., 182
 - Kish, L., 237
 - Klein, J.P., 138
 - Klempinger, D., 18, 31
 - Klunnert, M.D., 181–182
 - Koenker, R., 134
 - Korenman, S., 21
 - Krueger, A.B., 18, 21–22, 111
- L**
- L04, 34, 44
 - Labor market effects of birth sequences, 51*t*, 53*t*–54*t*
 - Lalive, R., 112
 - Lamb, M.E., 171
 - Landes, E.M., 150, 157
 - Lanjouw, P., 206–207, 227, 229
 - LATE estimate, 15–16, 23
 - Lechner, M., 4, 31–35, 38, 39–40, 45, 172, 192, 226
 - Lee, L.F., 10
 - Lehrer, S.F., 34
 - Lelièvre, E., 88–91, 106
 - Lemieux, T., 22
 - Leridon, H., 72
 - Levine, D.I., 194
 - Levitan, H., 182
 - Lieberson, S., 96
 - Lillard, L.A., 95
 - Linear regression model, 11
 - LISREL type models, 65
 - Littman, G.S., 236
 - Liu, S.H., 6, 167, 170, 171, 173, 182
 - LM01, 34, 38
 - and DATE, 36
 - “Local autonomy,” 94
 - Lu, H.H., 168
 - Lundberg, S., 150
 - Lutz, W., 1
 - Lynch, S.M., 235–236

M

Maccoby, E.E., 170
 Macurdy, T.E., 150
 Macy, M.W., 87
 Maddala, G.S., 152, 153
 Mahalanobis metric, 41
 sequential matching estimator, 42*t*–43*t*
 Manning, W., 168
 Manser, M., 169
 Manski, C.F., 3, 27
 Manting, D., 98
 Manton, K.G., 236
 Marginal-conditional decomposition, 69, 74
 “Marginal treatment effect” (MTE), 17, 18*f*
 Marini, M.M., 2, 92
 Marital dissolution, probability of, 157
 Marriage preferences and negotiation process, 102–103
 partial likelihood estimates of transition
 from consensual union to marriage, 103
 Marriage/unobserved marriage decisions,
 observed rate of entry, 103–104
 marriage rates and pregnancy, 104*f*
 Martin, J.A., 167, 170
 Marx, K., 237
 Mason, K.O., 234
 Mason, W.M., 234
 Mayer, K.U., 89
 Mazzucco, S., 215
 McElroy, M.B., 169
 McElroy, S., 169
 McGrath, J.E., 92–93
 McLanahan, S.S., 167, 169–171, 190, 193
 McNamee, R., 72
 McNicoll, G., 2, 237
 Mencarini, L., 215
 Meng, X.-L., 237
 Meyer, B.D., 111–113, 127
 Michael, R.T., 150, 157, 169
 Microeconomics, 33
 Mill, J.S., 79
 Miller, R.A., 32, 150
 Mills, M., 94, 97, 99, 100
 Mills-Trovato study, 100
 Mincer, J., 150
 Miquel, R., 33, 34, 38, 226
 Mixed proportional hazard (MPH) model, 114
 Moeschberger, M.L., 138
 Moffitt, R., 2, 4, 9, 11, 17, 20, 31, 133, 236,
 238–239
 Morgan, S.L., 2, 74, 76, 237, 239
 Morgan, S.P., 235–236

Morrison, D., 167
 Mouchart, M., 4, 59, 62, 68–69, 75

N

Nagin, D.S., 3
 Namboodiri, K., 236
 “Natural natural experiments,” 21
 Neidell, M.J., 181
 Neighborhood, dimension of, 178
 Nelson, F.D., 152–153
 Nepomnyaschy, L., 192
 Neyman, J., 33, 76
 Ní Bhrolcháin, M., 235, 238
 Non-experimental designs, 3

O

Oakes, D., 115
 Observational designs, *see* Non-experimental
 designs
 Ordinary Least Squares (OLS), 167
 Osborne, C., 167
 Oulhaj, A., 68, 69

P

Painter, G., 194
 Parallel and interdependent processes, 88–90
 dynamic system, 90
 interdependent processes: system approach,
 91–96
 levels of, 88–89
 Parental separation, definition, 173
 Parental separation on child health, new
 estimates on effects
 consequences of, 169–170
 estimation results
 assessing conditional independence
 assumption, 192
 box plot of propensity score overlap,
 190*f*
 choosing bandwidth, 190–191
 main findings, 184–190
 probit estimates of propensity score,
 185*t*–186*t*
 propensity score of parental relationship
 dissolution, 184
 relaxing common support condition,
 191–192
 sensitivity analysis, 190–192
 test of balancing properties between
 control and treatment group,
 187*t*–189*t*
 measure of child health, 181–182

- sample means by relationship status three years after an out-of-wedlock birth, 180r–181r
 - sample selection, 179–181
 - separation, victims of, 182–183
 - separation and selection, 170–171
 - statistical framework and estimation strategy
 - average treatment effect for treated (ATET), 176–177
 - conceptual model, 171–173
 - Conditional Independence Assumption (CIA), 175–176
 - matching, 175–177
 - matching estimators, 177–178
 - potential outcome approach, 174–175
 - Parents' propensity, to separate, 184
 - "Partialling" approach, 84
 - Path models/structural equation models (Crimmins), 234
 - Pearl, J., 61, 64, 73–74, 234
 - Pearl, Judea, 73
 - Pedhazer, E., 84
 - "Policy relevant treatment effects," 26
 - Pollak, R.A., 150
 - Popper, K., 64
 - Poss, S.S., 236
 - Pötter, U., 33
 - Powell, M.J.D., 125
 - Powell-method, 125
 - Prein, G., 101
 - Prentice, R.L., 89, 115, 120
 - Press, W.H., 125
 - Preston, S.H., 1, 233, 236
 - Propensity score matching, 86
- Q**
- Quandt, R.E., 152
 - Quantitative causal analysis, 59
 - Quetelet, A., 59, 79
- R**
- Rabe-Hesketh, S., 76
 - Raphael, J., 193
 - Rationale, 61
 - Ravallion, M., 206–207, 227, 229
 - Reduced form modeling, 32
 - "Reference couple," 159
 - see also* Italy, female labour participation, demographic processes
 - Regression discontinuity designs, 86
 - Regularity, 62
 - Reichmann, N.E., 178, 192
 - Reinhold, S., 24
 - Ribar, D.C., 19, 21–22, 169, 170
 - Richard, J.-F., 62
 - Ridder, G., 112, 113, 115, 127, 134, 136
 - Ritualo, A., 167
 - Robb, R., 11, 173
 - Robins, J.M., 27, 33, 34, 121, 125, 134
 - Robust dependence, causation as, 84–85
 - drawbacks, 88
 - Rohwer, G., 86, 88, 91, 94, 106
 - Romantic relationship, dissolution of, 116, 168
 - Rosenbaum, P.R., 39, 44, 168, 173, 176, 213
 - Rosenbaum, P.T., 213–214
 - Rosenfeld, R.A., 32
 - Rosenzweig, M., 21, 25
 - Rossi, P.H., 239
 - Rubin, D.B., 3, 9–10, 33, 37, 39, 44, 84–86, 151, 168, 173, 176, 213–214, 216, 218, 234, 237, 239
 - Rubin Causal Model, 9
 - Russo, F., 4, 59, 61–62, 64, 71–72, 81
- S**
- Sampling distribution, 65
 - Sampling theory framework, 74
 - Sandefur, G., 170–171, 190
 - Sauer, R., 31
 - Savage, L.J., 66
 - Schlesselman, J.J., 72
 - Schmid, R., 150
 - Schneider, B., 84, 86
 - Schoumaker, B., 201
 - Scott, E., 76
 - S-DCIA*, *see* Strong conditional dynamic independence assumption (*S-DCIA*)
 - Sequential potential outcome models, fertility on labor market outcomes, 31–33
 - data, 47–50
 - dynamic causal model, 34
 - basic structure of model, 35–36
 - defining estimand: average causal effects, 36
 - identification, 37–39
 - estimation results
 - multiple treatments and many periods, 44–45
 - quantity, 50–52
 - sequential matching estimators (SM), 40–44
 - structure of sequential estimators, 39–40
 - timing, 52
 - specifying causal parameters of interest

- Sequential potential outcome models, fertility
 on labor market outcomes (*cont.*)
 effects of timing and spacing, 46
 issues, 45–46
 number of children, 46
 role of confounding variables, 46–47
- Sequential probit models, estimated
 coefficients of, 49*t*
- Shadish, W.R., 84
- Shaw, K., 169
- SHIW, *see* Survey on Household Income and Wealth (SHIW)
- SHIW dataset (2002), 158
- Sigle-Rushton, W., 171, 173, 193
- Silverman, B.W., 190
- Sims, C.A., 33
- “Simultaneous equation models,” 75, 76
- Singer, B., 2, 92
- “Size elasticity,” 207
- Skrondal, A., 76
- Smith, H.L., 2, 233, 234, 237–239
- Sobel, M.E., 2, 235, 237
- Socio-economic status (SES)
 relation between tobacco and asbestos, 73*f*
 smoking, asbestos exposure and cancer of respiratory system, 72*f*
- Solon, G., 21
- Sorenson, E., 167
- “*Sorting out* issues of causation,” 233
- Sousa-Poza, A., 150
- Spiegelman, R.G., 113, 126–127
- Sporik, R., 181
- Staiger, D., 153, 159
- Stanley, J.C., 237
- Statistical inference and structural models,
 66–67
- Statistical model, 65–66
 of complex systems, 75
- Steele, F., 96, 171, 173
- Stinchcombe, A.L., 101
- Stock, J.H., 26, 153, 159
- Stress, psychological, 182
- Strong conditional dynamic independence
 assumption (*S-DCIA*), 38, 39
- “Strong program” (economics), 237
- Structural approach, 31
 advantages and disadvantages, 32
- Structural model, 67
- Structural modelling, exogeneity/causality
 causal analysis in social sciences
 background knowledge, 61–63
 goals of, 59–60
 probabilistic modelling in, 63–64
 variation and regularity in, 60–61
- conditional models, 67–69
 and exogeneity, 69–71
- confounding
 complex/completely recursive systems,
 74–75
 confounders and confounding, 71–74
- exogeneity and causality, 71
- partial observability and latent variables
 general case, 78–79
 three-component system, 76–78
- structural modelling
 meaning of structurality, 64–65
 statistical inference and structural
 models, 66–67
 statistical model, 65–66
- Structural stability, 62
- Suppes, P., 61
- Survey designs, *see* Non-experimental designs
- Survey on Household Income and Wealth
 (SHIW), 158–159
- Swisher, R.R., 171, 190
- Switching regression model, 11
- T**
- Tabutin, D., 201
- Teenage childbearing, causal effects in
 population research, 9–10
 basic causal model, 10–13
 instrumental variables, 13–18
 issues, 23
 differences in estimates across
 instruments, 23–25
 heterogeneity, 23
 instruments available, 26–27
 reduced form *versus* structural form,
 25–26
 relevance of instruments to policies of
 interest, 25
 weak instruments, 26
 types of instrumental variables, 18–23
- Teiramaa, E., 182
- Temporal lags and effect shapes, 93*f*
- Terza, J.V., 152, 153
- Thomas, R.L., 62
- Thomson, E., 170
- Three-component system, 76–78
 completely recursive system, 76
 with marginal independence, 77
- Time-dependent covariates, 92
 probability of change in dependent
 variable, 101
 see also individual time-dependent
 covariates

- Todd, P., 214
 Tolman, R.M., 193
 Toulemon, L., 72
 Transition rate (for population), 96
 from consensual union to marriage, 99*t*
 partial likelihood estimates, 103*f*
Treatise (Hume), 60
 Treatment effects, 16*f*
 Trivedi, P., 22, 26
 Troske, K.R., 32
 Trovato, F., 97, 100
 Trussell, J., 235
 Tsiatis, A.A., 113, 120–123, 125 134, 144
 Tuma, N.B., 89–91, 96
 Two-component system, 77
- V**
- Van den Berg, G.J., 112, 114
 Vandresse, M., 75
 Variation, 61–62
 rationale of, 61
 Vaupel, J.W., 1, 236
 Ver Ploeg, M., 20
 Viitanen, T., 22
 Voicu, A., 32
 Vytlačil, E., 17–18, 26, 87
- W**
- Wahba, S., 177, 184, 214, 218
 Waite, L.J., 95, 169
 Walker, J.J., 32
 Waller, M.R., 171, 190
 W-DCIA, *see* Weak dynamic conditional
 independence assumption
 (W-DCIA)
- Weak dynamic conditional independence
 assumption (W-DCIA), 37–38
 quantities identified under, 39
 Webster, R., 237
 Weisch, D., 181
 Weiss, Y., 169, 171
 Wellbeing, definition, 205
 Whyte, M.K., 183
 Widmer, H., 150
 Wiehler, S., 34
 Wiener, N., 33
 Willekens, F.J., 88
 Willer, R., 87
 Williamson, J., 64
 Willis, R.J., 169, 171
 Wilson, M., 183
 Winship, C., 2, 74, 76, 234–235, 237, 239
 Wolpin, K., 21, 25
 Woodbury, S.A., 113, 126–127
 World Bank Poverty Assessment (2003), 203
 Wright, R.J., 182, 190
 Wright, S., 59, 61, 64
 Wu, L.L., 170
 Wunsch, G., 4, 59, 73–74
- Y**
- Yamaguchi, K., 89, 96
 Ying, Z., 144
 Yogo, M., 26
- Z**
- Zeisel, H., 239
 Zhang, J., 31
 Zhang, W., 96
 Zhao, Z., 44