OXFORD

A HISTORY OF ECONOMETRICS

The Reformation from the 1970s

Duo Qin

A History of Econometrics

This page intentionally left blank

A History of Econometrics

The Reformation from the 1970s

Duo Qin

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2013
Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

ISBN 978–0–19–967934–8

Printed and bound in Great Britain by
CPI Group (UK) Ltd, Croydon, CR0 4YY

Preface

It was in sultry Manila, the Philippines, in 2005 when I set my mind seriously to this book project. At the time I was busy building macro-econometric models of major member countries for the Asian Development Bank (ADB). The decision seems to be as unrelated to the occasion and the location as one can possibly imagine. When I was first engaged in history research at Oxford and wrote 'The Formation of Econometrics: A Historical Perspective' as my doctorial thesis, I had never touched the murky water of applied modelling. The experience at ADB somehow revived in me an interest which had been numbed by years of working at the academic periphery within the orthodox economics circle of a British university. The Manila experience made me aware of the potential advantage of being an informed spectator with distance.

I was greatly inspired by T. Haavelmo's 1944 monograph when I undertook my doctorial research project in 1986. Returning to the history twenty years later, I find myself now critical of his empirical works. It has certainly taken me a long time to learn to appreciate real gems in empirical studies and to realize that the life force of econometrics lies fundamentally in its practical function rather than its methodological assertions. In fact, the two chapters of applied case study in the present book were the most time-consuming to research and write on, except for the designing and building of the citation database for Chapter 10. The latter has taught me how difficult the task of planning and collecting raw data for research is.

Nevertheless, methodology and philosophy are indispensible sources of stimulation for a history project like the present one. From a pragmatic perspective, I draw the philosophical support of this project from F. Nietzsche: 'History, conceived as pure knowledge, once it becomes sovereign, would be a kind of conclusion to living and a final reckoning of humanity. Only when historical culture is ruled and led by a higher force and does not itself govern and lead does it bring with it a powerful new stream of life, a developing culture for example, something healthy with future promises' (1873). Moreover, Nietzsche provides me with an expedient excuse for not attempting to discover the immutable truth of the history, but to develop an interpretation of it in the hope that the interpretation may benefit the econometrics teaching and research in the present and onwards. This is, I hope, not merely my own wishful thinking. Years of following an academic career have convinced me of the truth of Huxley's observation, especially in economics, 'that men do not learn much from the lessons of history is the most important of lessons history has to teach' (1959).

The writing of this book has taken much longer than I ever anticipated in Manila, almost comparable to the eight-year long Chinese Anti-Japanese War. During the period, I have been fortunate to get help and support from friends, collaborators, and colleagues. I feel greatly indebted to John Aldrich, Olav Bjerkholt, Marcel Boumans, Margaret Buckley, Stella Cavanagh, Diana Chan, Bill Farebrother, Christopher Gilbert, Xinhua He, David Hendry, Kevin Hoover, Martins Kazaks, Sreekala Kochugovindan, Ruben Lee, Terry Mills, Mary Morgan, Machiko Nissanke, Pedro Perez, Stephen Pollock, Aris Spanos, Tao Tan, Alf Vanags, Jun Wang, and Lifong Zou. But the list is far from exhaustive. I am also grateful to two particularly helpful interruptions during this book project. The first is the co-edition with Olav Bjerkholt of R. Frisch's 1930 Yale lecture notes; the second is the edition of a four-volume collection, '*The Rise of Econometrics*,' in Routledge's Critical Concepts in Economics series. Both editorial jobs have extended my knowledge of and revitalized my interest in the history, even though they caused an inevitable delay to the completion of the present book.

Finally, I extend my due thanks to the three institutions where I have worked during the last eight years: the ADB, Queen Mary, and SOAS of the University of London. These institutions have not only enabled me to carry out this project but also helped sustain my privileged position as a lone observer of econometric practices at various hierarchical levels and from widely different positions.

October, 2012, London

Acknowledgements

The author has reused materials from her articles originally published in various journals and hence would like to thank the pertinent publishers for granting permission. The details are as follows.

1. ‘Bayesian Econometrics: the first twenty years’, *Econometric Theory*, 1996, vol. 12, 500–16, Cambridge University Press.
2. ‘The error term in the history of time series econometrics’ (co-authored with C. L. Gilbert), *Econometric Theory*, 2001, vol. 17, 424–50, Cambridge University Press.
3. ‘The Phillips curve from the perspective of the history of econometrics’, *History of Political Economy*, 2011, vol. 43 (suppl 1), 283–308, Duke University Press.
4. ‘Rise of VAR Modelling Approach’, *Journal of Economic Surveys*, 2011, vol. 25, 156–74, Wiley.
5. ‘Consolidation of the Haavelmo-Cowles Commission Research Programme’, *Econometric Theory*, forthcoming in the special issue: Haavelmo Memorial issue, Cambridge University Press.

The author is also grateful to C. L. Gilbert for his generosity in allowing her to reuse a large part of their joint work listed in item 2 above.

This page intentionally left blank

Contents

This page intentionally left blank

List of Figures

This page intentionally left blank

List of Tables

This page intentionally left blank

List of Abbreviations

2SLS	two-stage least squares
3SLS	three-stage least squares
ADL	autoregressive distributed lag
AIC	Akaike information criterion
AR	autoregressive
ARIMA	autoregressive integrated moving average
ARMA	autoregressive-moving average
BIC	Bayesian information criterion
BVAR	Bayesian VAR
CC	Cowles Commission
CGE	computable general equilibrium
COMFAC	common factor (model)
CORC	Cochrane–Orcutt (model)
CORE	Centre for Operations Research and Econometrics (Belgium)
DAE	Department of Applied Economics (Cambridge University)
DFM	dynamic factor model
DGP	data generation process
DSGE	dynamic stochastic general equilibrium
D-W	Durbin–Watson (test)
EC	error correction
FIML	full-information maximum likelihood
GDP	gross domestic product
GIVE	Generalised Instrumental Variable Estimator
GMM	general method of moments
GNP	gross national product
IID	independent and identically distributed
ILS	indirect least squares
ITT	index of topic transfer
IV	instrumental variable
JEL	Journal of Economic Literature
JSTOR	Journal Storage
LIML	limited-information maximum likelihood

LSE	London School of Economics
MA	moving average
MAR	moving average representation
MDL	minimum description length
ML	maximum likelihood
NAIRU	non-accelerating inflation rate of unemployment
NBER	National Bureau of Economic Research
OLS	ordinary least squares
PIC	posterior information criterion
RBC	real business cycle
RE	rational expectations
SEM	simultaneous-equation model
SUR	seemingly unrelated regression
SURE	seemingly uncorrelated regression estimation
SVAR	structural VAR
VAR	Vector AutoRegression

Introduction

> The science hangs like a gathering fog in a valley, a fog which begins nowhere and goes nowhere, an incidental, unmeaning inconvenience to passers-by. H.G. Wells (1908), *A Modern Utopia* (from Section 3.3 'Utopian Economics')

This book is a sequel to *The Formation of Econometrics: A Historical Perspective*, my first book written mostly in the late 1980s. Although it was unclear to me then, my nascent research into the history of econometrics arose largely in the wake of reformative movements of the textbook econometrics built on the Cowles Commission (henceforth, CC) paradigm. The time elapsed since then is enough to turn those movements into history.

The present investigation is focused on various schools of thought and practices which attempted a 'paradigm shift' of the CC tradition during the 1970s and the 1980s, and principally the two decades after the 1973 oil crisis. Note that the time period overlaps with the transition of macroeconomics, as described in Blanchard (2000), from an epoch of formalization of an 'integrated framework' during the post-WWII period to an new epoch of exploration into the 'role of imperfections in macroeconomics' after 1980. Indeed, the history reveals a great deal of interaction between the reformative attempts in econometrics and the paradigm shift in macroeconomics around 1980. For that reason, the current study is somewhat biased towards macro-econometrics and modelling methods using time-series data at the expense of micro-econometrics.

The base materials of the current study are predominantly original research documents, since there are almost no previous historical studies of this kind. Except for a very few works, such as two chapters in Epstein's 1987 book (one on exogeneity and the other on VAR models) and another paper on exogeneity by Aldrich (1993), the field of the current study has remained unchartered territory. The available derivative materials are largely in the form of literature surveys, interviews, and methodological studies.

Two survey studies of three methodological schools carried out by Pagan (1987, 1995) are particularly helpful in setting the scene of the present book. The three schools form the subject matters of Chapters 2, 3, and 4 respectively, after an introductory chapter, that is Chapter 1, on the formation and consolidation of the CC paradigm. In comparison to Pagan's studies, these three chapters cover a much wider scope in terms of both the historical perspective and the relationship between econometrics and economics. This is particularly reflected in the focal discussion on the issue of 'model choice', an issue virtually left out by the CC group when they formalized econometrics. The complicated methodological problems involved in model choice are further illustrated in Chapters 5 and 6 through two case studies. The cases are chosen as representative of two common and somewhat distinct uses of econometric models. The first case, the Phillips curve, represents the use of models essentially for the purpose of theoretical verification and policy related debates. The second case, business-cycle modelling, embodies the type of research activities which ultimately have a more pragmatic and often tougher purpose—real-time forecasting. These chapters make it evident how unavoidable and vitally important the issue of model choice is for any econometric models to attain a satisfactory degree of theoretical soundness and empirical relevance.

The subsequent three chapters return to the old style of my first book, namely to examine the history by themes which are selected from the perspective of econometrics rather than from the history of economic thought. The object of investigation in Chapter 7 is structural parameters. Instead of focusing on the estimation aspect as textbook econometrics does, the chapter is devoted to the implicitly shifting specifications of structural parameters and the problems these give rise to. The problems are further exposed in Chapter 8 from the perspective of the error terms, the mirror reflection of model choice. The discussion culminates in Chapter 9 where the vexing subject matter of model selection and design procedures is examined directly. Due to the nature of the thematic arrangement, there is unavoidably a certain degree of repetition, across different chapters, of some topics and studies. However, by presenting them via different approaches or themes, I believe that the repetition helps to deepen and extend our understanding of the essence of econometrics.

The last chapter explores the use of citation analyses for historical research. In particular, the impact of the CC paradigm over three and half decades—1970–2005—is evaluated via a number of citation statistics. Interestingly, the statistics suggest that the impact of the CC paradigm has largely withstood reformative attempts for a 'paradigm shift', implying that the rise of various alternative schools during the 1970s and the 1980s has not yet resulted in a methodological revolution in econometrics, nor has it led to a shifting of epochs as described in macroeconomics by Blanchard (2000). Some tentative explanations of why the impact has been limited are offered in the epilogue.

Overall, the history considered in the current study reflects a gradual 'externalization' of research issues, to use Carnap's (1950) terminology. The focal issues upon which econometric tools and methods are developed have gradually diverged from '*internal questions*' towards more '*external questions*'. For example, parameters are internal to models, whereas the existence of models is external with respect to parameters. Econometric research has moved from the issue of how best to statistically estimate a priori given structural parameters, central under the CC paradigm, to harder issues concerning how to choose, design, and specify models and also how to evaluate which models are relatively the most efficient and simple while being substantively useful or relevant. Such externalization should provide a breeding ground where ideas and approaches which may eventually revolutionalize econometrics are mostly likely to grow.

It is obvious that the current book is on the internal history of econometrics, similar to my last book. Here, a summary of the external history does not seem to be necessary because the history is so recent and the research materials involved are also relatively diversely distributed. Nevertheless, many of the background stories can easily be found from various available sources, such as interviews and surveys, and are referred to whenever necessary in the subsequent chapters.

1

Consolidation of the Cowles Commission Programme

Modern econometrics is commonly regarded as being laid out by Trygve Haavelmo's *The Probability Approach in Econometrics* (1944) and formalized by researchers at the Cowles Commission (henceforth, CC) during the 1940s (see the CC Monograph 10 edited by Koopmans, 1950; and Monograph 14 edited by Hood and Koopmans, 1953). The formalization has been relatively well recorded and studied, for example see Christ (1952a, 1994), Hildreth (1986), Epstein (1987), Qin (1993), Gilbert and Qin (2006), and Spanos (2006).

This chapter examines the historical process through which the CC research programme became consolidated into mainstream econometrics.[1] The process went quite smoothly during the two decades 1950–70. These decades can therefore be viewed as a 'normal science' period by Kuhn's (1962) terminology. While the period serves the general purpose of preparing the background for the post-1970 reformative movements, it is particularly interesting to delve into because of a puzzling observation—the smooth consolidation took place in spite of the fact that the research community was not unaware that the programme was 'an intellectual success but an empirical failure' (Heckman, 2000), as shown from the papers presented at a panel discussion on 'Simultaneous Equation Estimation: Any Verdict Yet?' at the 1958 meeting of the Econometric Society, for example see Christ (1960), Liu (1960), and Waugh (1961). We are thus keen to find out what factors, or 'methodological rules' in Lakatos' (1977) terminology, have played the crucial role of promoting the CC paradigm despite its lack of empirical support.

Our investigation starts from a summary of the CC programme (Section1.1). The consolidation process is subsequently examined respectively through three

[1] The CC programme or paradigm is also referred to as the Cowles methodology or the Haavelmo-CC approach. The latter emphasizes the important role that Haavelmo has played in the CC research orientation and its historical continuity to some of the schemes proposed by Frisch.

areas: the development of standard econometrics textbooks (Section 1.2), the evolving themes and trends of research in econometric methods (Section 1.3), and the contribution of the CC programme to applied modelling (Section 1.4). The investigation reveals that the programme gained dominance primarily through its adherence to the scientific banner and style rather than its empirical relevance. What has attracted the academic community most comes from the hard science methodology.[2] Internal consistency of the arguments and mathematic elegance appear to rank at the top of the list of ingredients, serving as the key positive heuristic during the consolidation process. The possibility of further division of the econometric discipline into compartmentalized research and self-contained teaching topics has also played an important role. The most vital is probably the division of research tasks between economics and econometrics. The division helps to convert the aesthetics of econometric research into a means of producing mathematically elegant measurement tools for available theoretical models. Empirical evidence or applied relevance becomes comparatively depreciated. These findings provide a historical explanation of how 'the scientific illusion' criticized by Summers (1991) came to dominance in economics. To a large extent, the findings also reinforce the observation that advocacy leads to the dominance of a research programme more efficiently than objectivity (e.g. Mitroff, 1972, 1974 and Armstrong, 1979).

Several historical factors have probably contributed to this dominance.[3] First of all, academic acceptance was the prime driver of the CC enterprise; scientific rigour thus served as a trump card. Secondly, the poor availability of quality data and the primitive computing facilities of the time made it far easier to gain elegant theoretical results than robust empirical ones. Furthermore, the part of applied issues which have been virtually left out of the formalized econometrics often demands the use of comprehensive social knowledge or tacit experience, which is hard to trim or narrow down and is somewhat incompatible not only with the hard scientific ethos but also with pedagogical needs. Finally, the maturity of an academic discipline entails the establishment of a hierarchy where status descends in an orderly way from pure theoreticians to general practitioners.

Once econometrics became academically established, concerns and dissatisfaction were brewing from the increasingly noticeable gap between formalized econometrics and reality. The most crucial trigger was probably the widespread failure of macro-econometric models to predict and track the global recession following the 1973 oil crisis. In fact, 1970 had already seen

[2] This can also be seen as a part of what McCloskey (1983) describes as the pervasive movement of 'modernism'.

[3] Other psychological or related factors, such as selective perception, avoidance of disconfirming evidence, and group reinforcement of biased preconceptions, should not be ruled out, e.g. as described in Armstrong (1978), though they are not discussed here.

a mood for change, with the community critically questioning the practical value of the formalization (see Section 1.5).

1.1 Cowles commission methodology

A major impetus for the formalization of modern econometrics was the famous Tinbergen debate, triggered by Keynes' sceptical evaluation of Tinbergen's pioneering work *Statistical Testing of Business-Cycle Theories* (1939).[4] The debate highlighted the need for an academic base for the type of macro-econometric models explored by Tinbergen. In response to this need, Haavelmo and the Cowles Commission essentially formalized a new discipline, which still defines the core of present-day mainstream econometrics. As mentioned earlier, this part of the history has been relatively well researched; what follows is a highly condensed summary.

Methodologically, the CC enterprise is closely associated with the 'structural' approach. The approach is genealogically attributable to R. Frisch (e.g. Phelps Brown 1937;[5] Frisch 1938; Bjerkholt and Qin 2010). As an assistant to Frisch, Haavelmo played a pivotal role in adapting Frisch's proposal to probability-based statistical theories around the turn of 1940. The adaptation stemmed from Haavelmo's captivation by the Neyman–Pearson theory of hypothesis testing, as illustrated by his vehement argument for taking economic theories as testable statistical hypotheses (1939). However, Haavelmo's attention soon turned to the identification problem in connection with a simultaneous-equation model (SEM) and the associated estimation problem, which led to his discovery of the 'simultaneity bias' of the ordinary least squares (OLS) estimator (Haavelmo 1943) and his wholehearted embracing of the probability approach as the foundation of econometrics (Haavelmo 1944).[6]

Much of the CC research programme under Marschak's directorship during 1943–48 followed immediately on Haavelmo's work. In retrospect, the CC programme may be summarized from three perspectives. At a broad methodological level, it attempts to bridge theory and empirical research in a logically rigorous manner. Specifically, the CC research principle is to make all assumptions explicit in order to facilitate the discovery of problems and revision of the assumptions in the light of problems that might subsequently emerge. Moreover, the assumptions should be as consistent

[4] This debate is well documented in Hendry and Morgan (1995; Part VI and also the Introduction).

[5] Phelps Brown (1937: 356–6) reports a summary of 'an ideal programme for macrodynamic studies' presented by Frisch at the 1936 Oxford Meeting of the Econometric Society.

[6] The initial version of this work was released in 1941; see Bjerkholt (2007a) for a detailed historical account.

as possible with knowledge of human behaviour and are classified into two types: the first are those assumptions which are statistically testable and the second are provisional working hypotheses (see Marschak, 1946). At the level of the discipline of economics, demarcation is made between economists and econometricians. The job of postulating theoretic models is delegated to economists while econometricians are to specify and estimate structural models relayed to them by economists. Accepting on faith that those structural models are correct and autonomous,[7] the econometricians' job is further confined to devising statistically the best estimators for the parameters of the structural models. The importance of working with a priori constructed structural models is derived from the belief that only such models would satisfy the need for quantitative policy analyses. This structural approach also differentiates econometrics from statistical economics (see Gilbert and Qin, 2007). At the technical level, the CC researchers formalized the econometric procedure into three steps: model specification, identification, and estimation. They chose an SEM as the most general form of structural model and the starting point for the three steps. This choice is essentially based on the conceptual adequacy of SEMs in representing the Walrasian general equilibrium system.

Denote a linear SEM where sets of endogenous and exogenous variables are denoted by y_t and z_t respectively:

$$\sum_{i=0}^{p} B_i y_{t-i} = \sum_{i=0}^{p} \Gamma_i z_{t-i} + \varepsilon_t \, , \tag{1.1}$$

where B_i and Γ_i are matrices of structural parameters of interest, and p denotes the lag length. Model specification amounts to adopting a jointly normal distribution for the error term, ε_t, following Haavelmo's (1944) arguments that all the variables, that is $x_t = (y_t, z_t)$, should be specified as a set of jointly distributed stochastic variables. Identification is designated to the examination of the conditions under which the structural parameters of interest, especially those in the non-diagonal matrix B_0, are uniquely estimable.[8] The issue is demonstrated via a transformation of the structural model (1.1) into what is now commonly called the 'reduced-form':

$$y_t = -\left(\sum_{i=1}^{p} B_0^{-1} B_i y_{t-i} \right) + \sum_{i=1}^{p} B_0^{-1} \Gamma_i z_{t-i} + B_0^{-1} \varepsilon_t = \sum_{i=1}^{p} A_i y_{t-i} + \sum_{i=1}^{p} \Pi_i z_{t-i} + u_t \tag{1.2}$$

[7] For the concept of autonomy in connection to structural models, see Frisch (1938), Haavelmo (1944: section 8), and Aldrich (1989); also Chapter 7.

[8] Note that 'identification' carried far wider connotation prior to this formalization, e.g. Hendry and Morgan (1989) and Qin (1989).

Identification requires that structural parameters, (B_i, Γ_i), should be uniquely deducible from the reduced-form parameter estimates, (A_i, Π_i), and that the data sample in use should contain adequate variability to enable the estimation of (A_i, Π_i). The role of structural estimation is to deal with the nonlinear nature of the transformation of $(B_i, \Gamma_i) \rightarrow (A_i, \Pi_i)$. The principle method adopted is maximum likelihood (ML), because the OLS is an inconsistent estimator with the specification of an SEM, as shown by Haavelmo (1943). Ideally, the full-information maximum likelihood (FIML) estimator is to be used but, given the primitive computing facility of the time, a computationally more convenient method, known as limited-information maximum likelihood (LIML) estimator, was devised (e.g. see Epstein, 1987, 1989).

The CC's technical advance initiated a new standard for research. The standard embraced not just rigorous workmanship but also the task of measuring assumed known structural models. The CC group believed it crucial that the measurement was focused on structural models, because these models were the only type valid for simulating policy alternatives, for example see Marschak (1946). Reduced-form models were only useful for forecasting, a task somehow implicated as being inferior to that of policy analyses. It should be noted that the anchor of econometric measurement on structural models effectively granted the models the status of the maintained hypothesis, making them immune from hypothesis testing. Although the principle of hypothesis testing was highly commended, it became peripheral to the CC enterprise. Indeed, neither Haavelmo nor the CC group endeavoured much to transform the principle into operational devices for model diagnostic purposes. In CC parlance, the issue was referred to as making 'model choice' among multiple hypotheses and was intentionally left out of their research agenda, as acknowledged by Marschak (1950: section 2.6). But the CC group was not unaware that this issue would become unavoidable as soon as their formalized procedure was put into practice (see e.g. Koopmans 1957). The SEM of equation (1.1) was indeed general enough to represent any economy in the abstract but it was also unidentifiable by definition without further restrictions. The CC group therefore saw as the next stumbling block the lack of more specifically formulated theoretical models than the general SEM and, around 1950, re-oriented its research direction from measurement issues to theoretical model formulation (Christ, 1952a).

1.2 Programme consolidation: textbook standardization

Although CC Monograph 14 was produced for the purpose of expounding the earlier technical developments made in Monograph 10, the CC's works served the research community rather than education. Meanwhile, the first generation of post-war econometrics textbooks had essentially been distilled out of the modelling experience of practitioners and styled as workman's manuals covering a wide range of measurement issues, as can be seen from Tinbergen (1951) and Klein (1953). It took over a decade before econometrics textbooks gradually converged towards a regression-based statistics manual centred on estimation methods, that is, a setting which has been widely recognized as standard econometrics (see also Gilbert and Qin, 2006: section 4.4.3). It was actually a decade during which the CC methodology met with doubt and criticism from several directions (see the next section and also Qin, 1993: chapter 6). However, none of these were discussed in the textbooks subsequently published.

In order to better present the textbook convergence process, key features of the major textbooks produced during the two decades of 1950–70 are summarized in Table 1.1. In particular, the thematic layout of the textbooks is represented in four categories: 'applied analysis', 'statistical methods and estimators', 'SEM techniques', and 'pre-1940 techniques'. The first, 'applied analysis', covers chapters or sections which address squarely and relatively fully economic topics instead of mere and simple illustrations of econometric methods; the middle two categories are self-explanatory and the last, 'pre-1940 techniques', refers to those statistical methods which fell out of

Table 1.1 Thematic layout of major textbooks, 1950–1970

			Percentage of pages on themes of:			
Author (year)	Number of chapters	Pages	Applied analysis	Statistical methods & estimators	SEM techniques	Pre-1940 techniques*
Tinbergen (1951)	8	258	38	21	1	1
Tintner (1952)	11	370	27	50	8	41
Klein (1953)	7	355	14	69	15	0
Valavanis (1959)	12	223	0	85	32	8
Johnston (1963)	10	300	0	83	21	0
Goldberger (1964)	7	399	6	79	23	0
Malinvaud (1966) (French ed.: 1964)	20	631	5	76	18	3
Christ (1966)	11	705	11	48	19	0
Leser (1966)	7	119	41	46	19	0
Fox (1968)	14	568	20	60	17	6
Dhrymes (1970)	12	592	18	73	40	7
Wonnacott & Wonnacott (1970)	18	480	0	83	22	0

* 'Pre-1940 techniques' means those techniques experimented on by early econometricians but which largely fell into disuse during the 1950s and the 1960s, such as 'bunch map analysis', principal components and factor analysis. The selection of a page range is based on the unit of a section or a chapter; simple numerical illustrations of a technique within a section are not counted as 'applied analysis'. There is a certain overlap among the last three thematic categories.

fashion during the consolidation era, such as 'bunch map analysis', principal components and factor analysis. The categories are shown in terms of percentage of page coverage. As can easily be seen from Table 1.1, there was, during the 1960s, a substantial increase in the proportion of statistical content at the expense of 'applied analysis'; and this increase was accompanied by a rising proportion devoted to SEM. Let us now take a closer look into the contents of some of these books.

The main part of the pioneering textbook by Tinbergen (1951)—Part II—is a working description of how to practise econometrics by starting from available theories and then applying suitable statistical methods to measure and test the theories so as to enable model-based policy analyses. The remaining two parts contain ample discussion of applied cases and policy analyses, such as demand for agricultural products and technical development. The relevant statistical methods, such as correlation analysis, and SEMs versus the reduced-form models, are packed into the appendix.

Contrary to Tinbergen's practitioner style, Tinter's *Econometrics* (1952) is essentially a manual of the statistical techniques which have been tried on economic data, with an emphasis on multivariate analysis and time-series topics. It is nevertheless discernible from the book that discussions of special time-series features of economic data dominate those on techniques, a characteristic which has almost totally disappeared in the present-day textbooks. Technical issues pertaining to SEMs are sketchily addressed in chapter 7 under the title 'Stochastic Models with Errors in the Equations'. It was with the publication of Klein's (1953) *A Textbook of Econometrics* that the SEM techniques began to move to be more central.[9] That move was further helped by Valavanis' (1959) textbook. Focusing on ML methods, Valavanis implicitly promotes the SEM techniques to the cream of econometrics. One noticeable aspect of the SEM instruction in these textbooks is a notational simplification of the original CC models. The simplification amounts to reducing equations (1.1) and (1.2) into a pair of static models:

$$\begin{aligned} By_t &= \Gamma z_t + \varepsilon_t \\ y_t &= B^{-1}\Gamma z_t + B^{-1}\varepsilon_t = \Pi z_t + u_t \end{aligned} \tag{1.3}$$

Since the dynamic aspect of structural model (1.1) is not directly relevant to the technical gist of the identification conditions and the LIML method, the simplification helped to endorse the static SEM of (1.3) in neglect of the time-series features of economic data, for example those discussed at length

[9] Notice that the CC Monograph 14, which aimed at popularizing the SEM techniques, was only published in 1953. However, Klein had first-hand knowledge of the CC's work since he joined the group in 1944. On the other hand, Tintner was not involved.

in Tintner (1952). It was not until the rational expectations movement in macroeconomics and the rise of the VAR (Vector AutoRegression) approach and the LSE (London School of Economics) dynamic specification approach in econometrics from the mid-1970s onwards that the importance of dynamic modelling became gradually reinstated (see Chapters 3 and 4).

Clear signs of textbook standardization came with Johnston (1963) and Goldberger (1964). The two textbooks have quite similar structures, probably because the two authors shared their drafts and teaching experiences at the University of Wisconsin (see Gilbert and Qin, 2006). Statistical techniques and regression-based estimation methods occupy the greatest part of the two books. The SEM techniques are central, while the disfavoured pre-1940 techniques are dropped completely. Meanwhile, empirical cases merely function as illustrations of statistical techniques, and features particular to economic data are portrayed as 'miscellaneous' data 'problems' or 'complications' from the viewpoint of standard regression models. This toolbox style was significantly strengthened by the publication of an English version of Malinvaud's (1964) textbook in 1966. While retaining the layout of placing regression techniques in the elementary part and SEM techniques in the final part, Malinvaud's textbook was more comprehensive in terms of coverage and more rigorous in terms of mathematical treatment. It thus helped reinforce the Haavelmo–CC SEM approach as the central edifice of econometrics.

A few of the late 1960s textbook writers endeavoured to maintain the applied style led by Tinbergen (1951), see, for example, Leser (1966) and Fox (1968). For example, Leser (1966) devotes one chapter to production functions and another to demand analysis. Unfortunately, the endeavour somehow failed to compete with the growing dominance in university curricula of a statistics-menu-based econometrics. This growing dominance was accompanied by an expansion of the menu. For instance, an increasing number of post-CC techniques appeared in later textbooks such as Dhrymes (1970) and Wonnacott and Wonnacott (1970), for example spectral analysis and various cross-sectional data processing methods. The style was secured and settled in the 1970s, as demonstrated by the books of Theil (1971), Koutsoyiannis (1973), and Maddala (1977). It deserves to be mentioned here that textbooks of the 1970s began to branch into elementary/introductory and advanced levels to facilitate multi-semester teaching. Applied topics on their own merits were marginalized in the core teaching, which became largely a statistics subject in economics departments. In some universities, applied topics were organized into optional courses. Separate textbooks on applied econometrics appeared around the turn of 1970, for example Cramer (1969), Wallis (1973), and Wynn and Holden (1974).

Consequently, a simplified version of the CC approach came to be popularized through textbook standardization. Econometrics was taught simply as a set of universal statistical tools, useful mainly for estimating the parameters of a priori formulated theories. The scientific rigour of such a structural approach promoted the research styles of selecting and adapting applied issues and data to fit theories—see for example the case of the Lucas–Rapping inverse Phillips curve in Chapter 5—and of hiding data exploratory modelling experiments as 'sinful' activities—see Leamer (1978a: Preface).

1.3 Programme consolidation: emulative research

The previous section shows that the establishment of textbook econometrics took nearly two decades. During those two decades and actually not long after the publication of the CC Monographs 10 and 14, the CC approach was met by serious scepticism and opposition. The opposition arose mainly over three issues.[10] The first, and possibly the most heated, was on identification, or the arbitrary imposition of identification restrictions on SEMs needed in practice. This is shown by the reservation of Orcutt (1952) about the classification of 'endogenous' versus 'exogenous' variables; the objection of Wold to SEM as a valid structural model due to its lack of clearly specified causality (see Bentzel and Wold, 1946; Wold, 1954, 1956; Wold and Juréen, 1953); and the critique by Liu (1960) demonstrating the absence of correspondence of those restrictions to reality. The second issue was concerned with the CC's choice of research focus, which was criticized for being too narrow to allow for 'discovery' or 'hypothesis seeking', as shown from the 'measurement without theory' dispute (see Koopmans, 1947, 1949a; Vining, 1949) between the two rival groups—the CC and the National Bureau of Economic Research (NBER). The last issue was related to the restoration of the least squares method as a versatile and respectable estimation method in practice, demonstrated by (i) Fox's (1956) comparison of the OLS and ML estimates of the Klein–Goldberger (1955) model; (ii) Monte Carlo experiments carried out by Wagner (1958) and Christ (1960) respectively to compare the two estimation methods; and (iii) Waugh's (1961) summary verdict in favour of the OLS method.

Somehow, those disputes and disagreements were short-lived. They hardly stifled the inspiration that the CC's rigorous technical exposition excited, especially among young academics. To many of them, the CC's monographs opened a vast territory for a new style of scientific research. As researches following the CC's technical path accrued, the critical stage was reached in the 1960s whereby econometrics could be established around the core task of devising estimators for given structural parameters from a priori postulated

[10] For more description of these debates, see (Qin, 1993: chapter 6) and (Gilbert and Qin, 2006).

theoretical models. Inventions such as the two-stage least squares (2SLS) procedure made independently by Theil (1953) and Basmann (1957), and the instrumental variable (IV) estimation procedure developed by Sargan (1958, 1959) for SEMs were among the leading works. An arguably more innovative avenue was explored by Drèze (1962) to reformulate the CC's SEM methods using the Bayesian statistical approach. Drèze's initiative inspired a number of researchers and their endeavours during the 1960s are best represented by the first Bayesian textbook by Zellner (1971a), the layout of which was essentially patterned on Johnston (1963) and Goldberger (1964), see Chapter 2 for more details.

However, the area where the consolidation has experienced the most enduring success is probably micro-econometrics. Pioneering instances include the Tobit estimation for models of limited dependent variables (see Tobin, 1955, 1958), and the seemingly unrelated regression (SUR) procedure for models containing a set of individual regressions with correlated cross-regression error terms (see Zellner, 1962). Both methods were devised for data features particular to microeconomic sample surveys. Since the wide availability of such data and the computing capacity to handle them came only gradually, the technical diffusion of those methods was slow, albeit steady. Most of the major advances were made well after the turn of 1970, when the CC approach encountered a new wave of methodological criticisms in macro-econometrics (see Section 1.5). Remarkably, these criticisms caused little deterrent to the consolidation of the CC paradigm in micro-econometrics.[11] The very nature of cross-section survey data placed the emphasis of empirical studies on one-off explanations, that is explanations of particular surveys at the individual case level without explicit predictive targets, an emphasis which directly stimulated the construction of more detailed structural models. Newly constructed structural models in turn helped consolidate the measurement role of applied researches. The combination resulted in a further strengthening of the CC paradigm.

One interesting example was a joint study by Chamberlain and Griliches (1975) on modelling the economic effect of education. Technically, their 'novel' contribution was the device of an ML estimation procedure for an SEM extended with the error-in-variable component, an extension aimed at representing the unobservable-variable phenomenon. However, they concluded, after all the effort spent on deriving the ML estimation procedure, that this new procedure 'did not produce results which differed greatly from those based on simpler methods' and that this 'elaborate procedure, designed to detect possible sources of bias, yielded little evidence of such bias'

[11] The enduring effect of the consolidation in micro-econometrics is discernible from Heckman's (2000) retrospection of the CC tradition. More evidence is presented in Chapter 10.

(Chamberlain and Griliches, 1975: 436). This finding was virtually a rediscovery of the OLS versus ML verdict from over fifteen years earlier—as mentioned in the first paragraph of this section—and also of what was implied in Haavelmo's empirical studies nearly two decades before—as will be described in the next section.

In fact, evidence of the consolidation in micro-econometrics can also be found quite easily in the post-1980 literature. One of the best examples is three chapters, designated to the topics of demand analysis, producer behaviour, and labour economics respectively, in Part 8 'Selected Applications and Uses of Econometrics' in Volume III of the *Handbook of Econometrics* edited by Griliches and Intriligator (1986). The following quote gives another example. The quote is from a chapter in Volume II of the *Handbook of Applied Econometrics* which is reviewing the topic of dynamic optimisation with microeconomic data: 'Structural modelling imposed a discipline on empirical research that led to a rude awaking: theories which appear plausible and seem to agree with broad empirical regularities receive considerably less support when subjected to the rigors of structural modelling. Scaling back bold claims about the data to accommodate the findings of dynamic structural models is a healthy tonic for the profession, encouraging us to search for bigger, more informative data sets and to develop more subtle hypotheses' (Miller, 1997: 292).

Let us, however, return to the pre-1970 consolidation process for more evidence from bibliographic data. In particular, a sample of journal publications for the period 1950–70 have been drawn from the archive of economic and statistical journals in JSTOR, making use of its function of multiple-topic bibliographic search. The sample was collected by searching research papers under the topics of 'econometrics' or 'econometric model'. A subsample was then drawn by adding the topic of 'simultaneous equation model' (SEM) to the above sample. The resulting two time series of numbers of publications are plotted in the left panel of Figure 1.1. Two further subsamples were also drawn within the subsample of the publications under 'simultaneous equation model'—one for 'identification' and the other 'estimation', the two core topics of the CC approach. Once the publications under these two subsamples were collected, they were manually classified to tag those which are devoted to methodological issues. The right panel of Figure 1.1 illustrates the proportions of publications on estimation, identification, and methodology respectively in the SEM subsample.

The left panel Figure 1.1 shows a rising trend in SEM-related research publications, especially during the early to mid-1960s. The right panel shows that most of the publications were on the topics of identification and estimation, with those focused on methodological issues totally fading out by the mid-1960s; the growth of research papers on identification and estimation

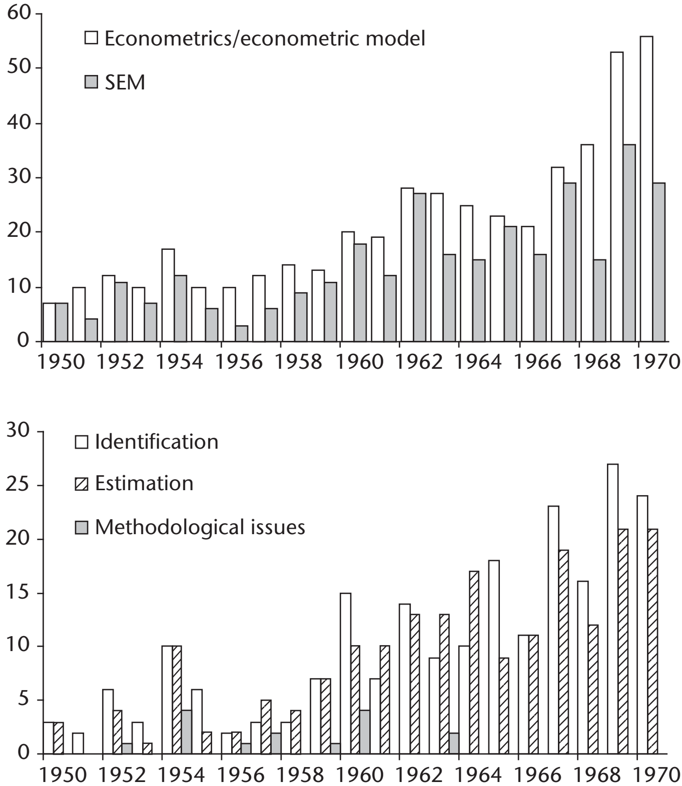

Figure 1.1. Numbers of research papers on SEMs, 1950–1970

Note: Since the JSTOR archive allows searches of multiple texts in combination, the 'identification' subsample is obtained by searching the combination of 'identification' and 'identifiable', and the 'estimation' subsample is by 'estimation' and 'estimator'.

techniques was particularly pronounced for the post-1960 period. Research papers on identification led the trend, probably due to the fact that identification was considered conceptually a prerequisite for parameter estimation of SEMs and also partly due to the constraint of the computing facilities of the time. Notice from the left panel, however, that the trend of the SEM-led research stagnated towards the end of the 1960s. This observation prefigures the mood for change to be described in Section 1.5.

1.4 Programme consolidation: the role of applied modelling

Econometrics was largely part of applied economic research prior to its formalization through the CC enterprise during the 1940s. Among the CC community of the time, only two members, Haavelmo and Klein, have ever engaged themselves in empirical studies. Noticeably, their engagement shows a substantial difference in motivation; and the consolidation is reflected by Haavelmo's motivation gradually gaining an upper hand over Klein's within the academic community.

Haavelmo's empirical studies were essentially motivated by the need to demonstrate the inferiority of the OLS estimation method in SEMs. Aggregated consumption was chosen for this purpose in two successive studies (Haavelmo, 1947 and Girshick and Haavelmo, 1947). In the first study, the structural parameter of interest was the marginal propensity to consume with respect to income, which was estimated by two methods, OLS versus ILS (indirect least squares), the latter being conditional upon the assumption that consumption and income were mutually endogenous in a theoretical model where investment was the driving independent variable. The US national account data of the period 1922–41 were used and two sets of estimates were obtained, one from the full sample and the other the subsample of 1929–41. The full-sample OLS estimate was reported as 0.732, which fell just outside the upper limit of the reported 95 per cent confidence interval (0.57, 0.73) of the ILS estimate 0.672, and the subsample point estimates by the two methods were shown to be much closer. What is peculiar, however, is the absence of the corresponding confidence intervals of the OLS estimates, when they are compared to the ILS estimates. To find explanations, a re-estimation of Haavelmo's model has been carried out and the results are given in Table 1.2, along with Haavelmo's original results. As can be seen from Table 1.2, it would have been obvious that the OLS inconsistency under small and finite samples lacked statistical significance, had the missing intervals or the standard deviations of the OLS been reported.

Somehow, Haavelmo's (1947) empirical study strengthened his conviction of the inferiority of the OLS. In his joint work with Girshick (1947), simple OLS estimates were totally absent and the OLS method was dismissed from the outset for being logically inconsistent. A five-equation SEM was set up

Table 1.2 Estimates of the marginal propensity to consume (Haavelmo, 1947)

Sample periods		1922–1941	1929–1941
ILS estimates via the income-investment equation (95% confidence interval)	Haavelmo's estimates	**0.672** **(0.57, 0.73)**	**0.712** (N/A)
OLS estimates of the consumption equation (95% confidence interval)	Haavelmo's estimates	**0.732** (N/A)	**0.723** (N/A)
	Re-estimates	0.732 (0.670, 0.795)	0.724 (0.656, 0.791)

Note: Haavelmo's subsample estimate seems to be using the full-sample fitted income from the first-stage LS in the second-stage LS estimation. A re-estimation following that approach yields 0.714, which is close to 0.712.

to study the demand for food; but the real purpose was to illustrate, using annual time-series data of the period 1922–41, how to estimate the structural coefficients consistently with the a priori SEM specification. The calculation task was much heavier and more complicated than that in Haavelmo (1947). Lessening the computational burden formed an important drive for the invention of the LIML estimator (see Epstein, 1989). Sadly, the empirical significance of undertaking this burden was not evaluated. Again, the re-estimation of two key parameters in the food demand equation of their model shows clearly that the OLS inconsistency is statistically insignificant (see Table 1.3).

The lack of significant improvement by those elaborate SEM-consistent estimators over the OLS in applied studies was soon exposed explicitly by Christ (1952b), as described below, and subsequently verified by others, which led to the practical rehabilitation of the OLS around 1960, as mentioned earlier. Unfortunately, a general awareness of this lack of improvement was somehow absent in the academic community and it was to be repeatedly 'rediscovered' from applied modelling experiments for decades to come. Aside from the case of the Chamberlain–Griliches (1975) study described in the previous section, two other cases are given in Chapter 5—one of modelling the wage–price relation by Sargan (1964) and the other of modelling the unemployment–inflation trade-off by Lucas (1972b, 1973). It was as if there was a conscious choice among the academic community to ignore the lack of significant improvement. Certainly, this lack was hardly mentioned in those textbooks listed in Table 1.1.

With the benefit of hindsight, it is apparently Haavelmo's method-illustrative style which has touched the right chord with the academic community. To many, applied studies have become evaluated primarily by the illustrative capacity of their methods and techniques rather than empirical significance. What really counts is the scientific rigour of Haavelmo's

Table 1.3 Structural parameter estimates of the demand for food equation in the SEM of Girshick and Haavelmo (1947)

Parameters of interest: coefficient for	Relative prices	Income
Girshick and Haavelmo's ILS estimates	**–0.246**	**0.247**
(standard deviation)	(N/A)	(N/A)
OLS re-estimates	–0.346	0.286
(standard deviation)	(0.068)	(0.038)
2SLS re-estimates	–0.423	0.287
(standard deviation)	(0.107)	(0.043)
LIML re-estimates	–0.607	0.324
(standard deviation)	(0.238)	(0.072)

Note: These results are based on equation (4.1) in the original paper. The 2SLS re-estimates are found to be noticeably different from the original estimates, because there are several numeric differences in the variance–covariance matrices between the calculation of the original paper and the present re-estimation.

arguments. For example, Haavelmo's (1947) paper is regarded as 'the first direct contribution to the development of exact finite sample distribution theory' (Basmann, 1974: 210). Scanning through the bibliographic sample collected for Figure 1.1, one can easily see that Haavelmo's method-illustrative style had become widely adopted in econometric journal publications towards the end of the 1960s. The goal of dealing with real economic issues by means of appropriately selected statistical techniques, the original goal which had promoted the birth of econometrics in the first place, was relegated largely to applied economists. According to a survey paper by Wallis (1969), applied econometrics was defined as including primarily those 'methods and techniques which have seen successful application in a number of areas of economic investigation' (Wallis, 1969: 771). The methods and techniques covered in Wallis's survey dealt mainly either with estimation issues associated with different model formulations, such as partial adjustment models or expectations models, or with forecasting and simulation issues of SEMs.

Noticeably, Klein's applied modelling career shows a reverse pattern. It gradually diverged from method-instigated topics towards reality-instigated topics. Klein's first applied endeavour was a 16-equation macro-econometric model of the USA covering the sample period of 1921–41 (known as Klein's Model III, see Klein, 1947, 1950). The model was inspired by both Tinbergen's macro-econometric model-building works and the subsequent CC theoretical contributions. Although considerably smaller than the 48-equation Tinbergen model of the USA (1939),[12] Klein's Model III was the first fully fledged macro-econometric model based on the CC structural approach. It followed the SEM structural → reduced-form procedure estimated by both the OLS and the LIML. The practical accomplishment of his endeavour was closely scrutinized and severely challenged by Christ (1952b).[13] Among other things, Christ found that the OLS estimates did not differ significantly from the complicated LIML estimates on the whole, that the model predictions using OLS estimates produced smaller errors than those using LIML estimates, and that the predictions of the structural model were no better than those obtained by naïve models of data extrapolation. Essentially, Christ's investigation revealed that the most vulnerable place in applied modelling lay in the equation and model design rather than the choice of estimation methods. This finding also implied that model design was actually where the greatest potential lay in terms of raising the signal-to-noise ratio and thus the value-added of the modelling research.

[12] Tinbergen's model used 52 time series of the period 1919–32.

[13] Christ's test of Klein's model was descended from Marshall's (1950) test and jointly discussed by M. Friedman, Klein, G. H. Moore, and Tinbergen, as appended to Christ (1952b).

Christ's investigation had a discernible impact on the subsequent Klein–Goldberger model of the USA, which contained 20 equations covering the period 1929–52. That is most clearly seen from the discussions that Klein and Goldberger (1955) devoted to issues relating to equation and model design. For example, they discussed at length whether additional lag terms were desirable and what level of aggregation was appropriate; the choices were often made from a combined consideration of both available theories and data. Such a heavy use of Tinbergen's 'kitchen work' style demonstrated the severe practical limit of the CC's structural modelling approach. Nevertheless, the Klein–Goldberger model adhered to the CC methodology when it came to estimation. It was estimated by LIML only. The alternative OLS estimates were subsequently supplied by Fox (1956). The difference between those estimates and the LIML estimates was smaller than expected, illustrating once more that there was relatively little to be gained from more sophisticated estimation methods than the OLS.

The OLS received its formal rehabilitation towards the end of the 1950s, as mentioned in the previous section. Ambivalent about the rehabilitation and especially its impact on single-equation versus simultaneous-equation modelling approaches, Klein nevertheless acknowledged, 'users of econometric models are often not really interested in particular structural parameters by themselves. They are interested in the solution to the system, under alternative sets of conditions. . . . It is the difference between partial and general analysis that is involved. It is conceivable that partial analysis is an end, in itself, for some problems—possibly those of a purely pedagogical nature—but most problems call for a more complete analysis of the system' (Klein 1960: 217).

Interestingly, the Klein–Goldberger model did not derive its empirical appeal from adherence to the CC techniques. Rather, the appeal was mainly from the dynamic properties of the model, that is properties which would enable quantitative studies of the time effects of different economic factors (see Goldberger, 1959). Scrutinized independently by Adelman and Adelman (1959), the capacity of the Klein–Goldberger model to generate dynamic properties that were interesting and desirable for business cycle studies was confirmed. Modelling dynamics became the central theme in two immediate progenies of the Klein–Goldberger model, one built by Liu (1955, 1963) and the other by Suits (1962), see also Bodkin et al. (1991: ch. 3). Liu's modelling experience made him acutely critical of the CC's approach; Suits abandoned the conventional structural model formulation and went for a growth-rate model. Both modellers simply used OLS and made forecasting accuracy their primary criterion in model evaluation.

As his empirical experiences accrued, Klein's modelling approach also became more eclectic, synthetic, and increasingly detached from the CC's approach. Realizing that it was of primary importance to have as much data information as available, he soon ventured into models using quarterly time series. In his first quarterly model (see Barger and Klein, 1954), he experimented with Wold's recursive model, a rival model to the CC's SEM, and tried to hybridize the two in order to deal with the time-series problems newly emerged in association with quarterly data. An extension of this research resulted in a quarterly model of the UK (see Klein et al., 1961), and subsequently the massive Brookings model (see Duesenberry et al., 1965, 1969).

It was evident from both the documentation and comments on these models (see, e.g., the comments by Griliches, 1968; Gordon, 1970), that a great deal of attention and debate was devoted to the choice of variables, equation forms, and specifications, as well as cross-equation linkage within a model. In comparison, execution of the CC's SEM techniques yielded marginally improved results, since the actual modelling process was far more complicated and entangled than the CC's formalized stepwise procedure. Nevertheless, these repeatedly negative findings did not shake Klein's faith in the CC paradigm, nor that of his fellow modellers. It seemed that the strong scientific image of the paradigm mattered much more than its empirical usefulness. As observed by Bodkin et al. (1991: chapter 4), a team-work approach and institutionalized modelling maintenance were among the major factors driving the rapid growth of macro-econometric modelling activities in the 1960s. Both factors entailed funding, which would be impossible to secure without the backing of a strong scientific image that the CC paradigm readily projected.[14] In contrast, the macro modelling activities led by Klein attracted dwindling interest from the academic research community, as shown from the citation analysis in Chapter 10.

Nevertheless, applied macro-econometric modelling had approached its heyday (Klein, 1971), and econometrics had been established as a fully respectable sub-discipline within economics (see e.g. Schachter, 1973) by the end of the 1960s. Worries of the pioneers over the status of econometrics, especially those aroused by the Tinbergen debates, were over and almost forgotten.

[14] Evidence provided by Amstrong (1978, 1979) suggests that econometricians carry out research more on belief than on fact, that the majority believe in the superiority of method complexity, and that advocacy and group conformity tend to dominate objectivity in their research strategy.

1.5 Mood changes: 'econometrics' or 'playometrics'

As described in previous sections, the CC approach was consolidated and standardized through research and higher education. Meanwhile, the process was accompanied by a further division of research topics, interests, objectives, and especially a split in the research field between theoretical and applied econometrics within econometrics, as well as between econometrics and applied economics within economics. As a result, applied modelling activities were marginalized in the econometrics community. The mainstream research strategy became settled on narrowly defined topics which would allow for the vigorous exhibition of the skills of mathematical statistics with increasing complexity.

This strategy, however, bred a paradoxical phenomenon. While the original objective of developing econometrics after the hard science paradigm was to improve the understanding and forecasting of the economic reality, the pursuit for scientification led to the discipline growing further away from reality. This phenomenon in turn bred a shift of sentiment within the profession—a mood of critical introspection over the credibility of econometrics.

A strong disappointment was voiced by Frisch, the founding father of econometrics. In a short essay in honour of Roy Harrod, Frisch called mainstream econometrics 'playometrics' to express his bitter aversion to those research works which fed the 'self-admiration' of the econometrics society but showed little economic relevance or interest in 'social and scientific responsibility' (Frisch 1970). Frisch's criticism found an unexpected exponent: W. Leontief.[15] Similar to Frisch, Leontief's criticism was blunt and scathing; and it was delivered on a far more influential occasion—his presidential address at the 83rd meeting of the American Economic Association in 1970. Being largely an onlooker of the consolidation process, Leontief expressed 'an uneasy feeling about the present state of our discipline' and extended substantially the earlier warning of his predecessor, F. Hahn.[16] Leontief observed:

> Much of current academic teaching and research has been criticized for its lack of relevance, that is, of immediate practical impact. . . . The trouble is caused, however, not by an inadequate selection of targets, but rather by our inability to hit squarely any one of them. The uneasiness of which I spoke before is caused

[15] Frisch disputed with Leontief over identification issues in the early 1930s (see Morgan, 1990: chapter 6).

[16] Hahn warned in his presidential address to the Econometric Society (1970): 'The achievements of economic theory in the last two decades are both impressive and in many ways beautiful. But it cannot be denied that there is something scandalous in the spectacle of so many people refining the analyses of economic states which they give no reason to suppose will ever, or have ever, come about. It probably is also dangerous'.

> not by the *irrelevance* of the practical problems to which present day economists address their efforts, but rather by the palpable *inadequacy* of the scientific means with which they try to solve them....
>
> ...I submit that the consistently indifferent performance in practical applications is in fact a symptom of a fundamental imbalance in the present state of our discipline. The weak and all too slowly growing empirical foundation clearly cannot support the proliferating superstructure of pure, or should I say, speculative economic theory....
>
> ...The mathematical model-building industry has grown into one of the most prestigious, possibly the most prestigious branch of economics.... Alongside the mounting pile of elaborate theoretical models we see a fast-growing stock of equally intricate statistical tools. These are intended to stretch to the limit the meagre supply of facts.
>
> ...In no other field of empirical inquiry has so massive and sophisticated a statistical machinery been used with such indifferent results. Nevertheless, theorists continue to turn out model after model and mathematical statisticians to devise complicated procedures one after another. Most of these are relegated to the stockpile without any practical application or after only a perfunctory demonstration exercise....
>
> Continued preoccupation with imaginary, hypothetical, rather than with observable reality has gradually led to a distortion of the informal valuation scale used in our academic community to assess and to rank the scientific performance of its members. Empirical analysis, according to this scale, gets a lower rating than formal mathematical reasoning. Devising a new statistical procedure, however tenuous, that makes it possible to squeeze out one more unknown parameter from a given set of data, is judged a greater scientific achievement than the successful search for additional information that would permit us to measure the magnitude of the same parameter in a less ingenious, but more reliable way.... It is not surprising that the younger economists, particularly those engaged in teaching and in academic research, seem by now quite content with a situation in which they can demonstrate their prowess (and incidentally, advance their careers) by building more and more complicated mathematical models and devising more and more sophisticated methods of statistical inference without ever engaging in empirical research.... This complacent feeling, as I said before, discourages venturesome attempts to widen and to deepen the empirical foundations of economic analysis, particularly those attempts that would involve crossing the conventional lines separating ours from the adjoining fields. (Leontief, 1971: 1–5).

To a great extent, the criticisms of these eminent figures endorsed forcefully what Vining (1949) had objected to in the CC approach on behalf of the NBER approach twenty years earlier, as well as those subsequently and periodically voiced criticisms in relation to business cycle modelling research (see Chapter 6). The criticisms were soon echoed by a wave of condemnation of the formalist movement in economics, for example see Phelps Brown

(1972); Worswick (1972). There was even a description in caricature of the life of the 'Econ' tribe by Leijonhufvud (1974). A 'priestly caste' was found among the academic economist community, with 'Math-Econ' assuming the highest status while 'Develops' ranked lowest. The caste encouraged 'fields' segregation within the tribe and promoted the manufacture of models for 'ceremonial' use to the outsiders, whereas within the community, there was 'alienation, disorientation, and a general loss of spiritual values'. However, more serious attacks on the CC-based structural model approach were targeted at forecasting records and policy simulation issues. These became of increasing concern as the world economy entered a turbulent period in the wake of the 1973 oil crisis. A number of forecasting comparisons demonstrated that econometric forecasts were no better than extrapolations using models without economic content, for example see Kosobud (1970); Cooper (1972); Nelson (1972); Elliott (1973); Granger and Newbold (1974b); Narasimham et al. (1974); and Levenbach et al. (1974). The especially poor forecasting records of the post-oil crisis period by macro-econometric models were deemed as 'econometric failure on a grand scale' by Lucas and Sargent (1979). The validity of conducting model-based policy simulations was challenged by Lucas (1976), a paper now widely referred to as the Lucas critique.[17]

It was in this climate that critical appraisals of the methodology of the established CC structural approach revived within the econometrics community. Among them, three strands of reformative ideas are particularly noticeable for their systematic exploration into alternative model specification routes (see Pagan, 1987, 1995). The first is the Bayesian specification search pioneered by E. Leamer (1978a), in an effort to rescue econometricians from the schizophrenia of preaching econometric theory on the top floor while 'sinning' in the basement with applied model mining (see Chapter 2). The second is the VAR approach led by C. Sims (1980a), as a dynamically general way of tackling the issue of 'model choice' in an effort to bring macro-econometrics closer to reality (see Chapter 3). The third is the LSE dynamic specification approach, led mainly by D. Hendry (1980), to restructure applied macro model building on a more systematic and data-coherent basis so as to rid econometrics of those elements associated with 'alchemy', 'econo-mystics', 'economic-tricks', or 'icon-ometrics' (see Chapter 4). Paradigm reform was looming.

[17] The paper first circulated as a Working Paper at Carnegie-Mellon University in 1973.

2

Rise of Bayesian Econometrics

Bayesian econometrics is seen by many as a rival to mainstream econometrics using classical statistical methods. Unlike the classical methods, Bayesian econometrics embarks on the task of parameter estimation from the starting point of an explicit specification of prior distributions to the parameters under question. The Bayesian parameter estimators are thus derived from the posterior distributions which are the result of combining 'the priors' (abbreviation for the prior distributions) with the data-based likelihood functions representing the sampling distributions. This chapter looks at the history of Bayesian econometrics from the 1960s up to the early 1990s. It investigates how Bayesian inference was adopted in econometrics, the main causes of the growth of Bayesian econometrics, and what major achievements Bayesian econometrics has attained. In order to keep our focus on evaluating Bayesian research in the light of the development of econometric methodology, the following aspects are put aside: (i) the influence of the developments of external factors such as Bayesian statistics and computing technology; (ii) the detailed technical advances; (iii) the development of the Bayesian approach in economics; and (iv) the relevant philosophical debate. Descriptions of these aspects can be found in various literature surveys and books (e.g. Zellner, 1971a, 1971b, 1984, 2008; Poirier, 1988; Howson and Urbach, 1989; Florens et al., 1990; Koop, 1994; Koop et al., 2007).

Our investigation shows that Bayesian econometrics rose from a desire to emulate the CC paradigm. The subsequent emergence of a Bayesian model specification approach, led by Edward E. Leamer, was a methodological reaction to the lack of an explicit treatment for model specification in the paradigm. Although Leamer's works highlighted the empirical fragility in most of the a priori formulated models, he fell short of providing an alternative modelling strategy which would increase the robustness of econometric models systematically. The later fusion of Bayesian methods with time-series econometrics has blurred the ideological demarcation between Bayesian econometrics and econometrics based on classical methods. The Bayesian route has repeatedly proved capable of generating tools equivalent to or comparable with those generated by the classical approach. However, none of these developments presents new methodological challenges to the CC paradigm.

Section 2.1 describes the entry of Bayesian inference into econometrics, mainly during the 1960s; Section 2.2 turns to the emergence of Leamer's Bayesian specification search;[1] Section 2.3 continues the Bayesian-based model selection research; Section 2.4 sketches the Bayesian fusion with time-series econometrics. The last section concludes with some retrospective assessments.

2.1 Bayesian entry and reformulation of the structural approach

The potential of Bayesian inference in econometrics was recognized by J. Marschak as early as 1950 (see Marschak, 1954).[2] To answer the question 'Is there any relation between subjective probabilities and the statistical method?', Marschak used simple examples to illustrate how the ratio of two of Bayes' formulae could be used to compare and modify degrees of prior belief through combining them with the likelihood functions. But Marschak's paper did not inspire any econometrician disciples to try the Bayesian route.[3] It took a decade before the route was seriously explored in econometrics, mostly motivated by Bayesian research in mathematical statistics. The pioneering works included W. D. Fisher (1962), Drèze (1962), Hildreth (1963), Rothenberg (1963), Tiao and Zellner (1964a, 1964b), Zellner and Tiao (1964).

The works of Fisher (1962) and Hildreth (1963) were exploratory at a purely theoretical level. Fisher examined the different effects on model estimation induced by different purposes of model use, for example for prediction or for policy simulation. Starting from two different loss functions corresponding to the two different purposes, Fisher derived two sets of coefficient estimates that would minimize the two different loss functions respectively, using Bayes' theorem, that is, in terms of assumed prior densities. The procedure effectively linked the step of estimation with the desired welfare function for policy control purposes. Hildreth's discussion also considered this issue from the decision-making viewpoint. He showed how to obtain different coefficient point estimators by replacing the standard statistical criteria with other criteria according to the requirements of the problem concerned.

[1] Sections 2.1 and 2.2 are basically a summary version of sections 3–5 of Qin (1996).

[2] Although his paper was published in 1954, the content was delivered in lectures at the end of 1950.

[3] Citation searches of Marschak's paper in JSTOR and Web of Science yield zero results; in Google Scholar, the few articles citing the paper during the historical period of our investigation are unrelated to econometrics.

Tiao and Zellner were among the first to try Bayesian methods for regression models. In their (1964a) paper (mimeo 1963), Tiao and Zellner started from the standard Bayesian procedure of a simple regression model, that is:

$$y_t = \beta x_t + u_t; \qquad u_t \sim IID\left(0, \sigma^2\right) \tag{2.1}$$

and proposed a sequential Bayesian method to overcome the difficulty of specifying the prior of the regression coefficients, $p(\beta, \sigma)$. The idea was to split a data sample into two, and take one of the subsample estimates as the parameters of the prior, which was assumed to follow the locally uniform distribution function. The posterior coefficient estimates were then obtained by combining the specified prior with the likelihood function, $l(\beta, \sigma)$, based on the data of the other subsample. The method was applied to an investment equation with the time-series data of two companies treated as two subsamples. Around the same time, they proposed another Bayesian method for a regression model with first-order autocorrelated errors, that is extending (2.1) by $u_t = \rho u_{t-1} + \varepsilon_t$ (see Zellner and Tiao, 1964). The method was designed—by choosing priors for ρ—to circumvent the difficulty of deriving the desired asymptotic property of classical estimators in that situation.

Drèze's (1962) memorandum was probably the first to emulate the CC programme within the SEM framework. Instead of discussing the estimation issue, Drèze devoted his attention to the identification issue. From a Bayesian perspective, Drèze classified the a priori information of an SEM into two parts: one consisting of the split between endogenous and exogenous variables and the other consisting of all the assumptions concerning the signs and magnitudes of the structural parameters and the parameters of the covariance matrix of the error term. It was mainly in the latter part that Drèze felt dissatisfied with the classical methods, feeling that they either ignored this piece of information or inserted it in the form of exact restrictions, an insertion incompatible with the Cowles Commission probability approach of econometric modelling. Bayes' principle was a promising method of improvement, in his view, because 'the revision of prior probabilities in the light of observations generated by a known process is a straightforward application of Bayes' theorem' (Drèze, 1962: 24). Starting from the assumed known split between endogenous and exogenous variables, he formalized the latter piece of information into a joint prior density of the parameters concerned and derived the Bayesian formulae of an SEM and its corresponding reduced form. On the basis of this pair of Bayesian structural SEM and its reduced form, Drèze reformulated the Cowles identification conditions. He showed that the limitation of sample information, included in the likelihood function, was to provide estimates of the reduced form coefficients on the basis of endogenous versus exogenous specification, that the identifiability of simultaneous equation coefficients normally required additional information, which was embodied in the prior density, and that the distribution type of the prior density usually determined whether an SEM was over-, or just, or under-identified.

Inspired by the works of Fisher (1962) and Drèze (1962), Rothenberg (1963) combined an SEM with a loss function representing policy-making behaviour in order to study the effects of different priors on the posterior parameter estimates under the alternative assumptions of the known and unknown error variance of the model. He used Haavelmo's (1947) consumption model as an illustration of his Bayesian device and concluded that 'the Bayesian and traditional solutions are quite similar in exactly identified models with "weak" prior information' (Rothenberg, 1963: 19).

Emulation of the CC programme culminated in the publication of Zellner's *An Introduction to Bayesian Inference in Econometrics* (1971a). The book was essentially a Bayesian reformulation of standard econometrics textbooks (e.g. Johnston, 1963; Goldberger, 1964), proceeding from simple linear regression to special problems in regression analysis (e.g. models with autocorrelated errors or with errors in variables) and then moving on to SEMs, with a substantial section on issues of structural model estimation and identification.

On the research front, Bayesian econometricians shared the measurement commitment with non-Bayesian econometricians of the time and focused their attention on two technical difficulties in devising Bayesian estimators: the difficulty of specifying prior distributions that were both economically interpretable and mathematically tractable, and the difficulty of integral calculation pertinent to the derivation of the posterior distributions from joining the prior with likelihood functions. In the context of a simple/multivariate regression model, various Bayesian estimation analogues were worked out for different models already developed by the classical camp, such as the autoregressive model and the common-factor model (e.g. see Thornber, 1967; Zellner and Geisel, 1970; Lempers, 1971; Zellner, 1971a; Shiller, 1973). Their empirical findings often confirmed those of the classical camp, although Bayesian priors were believed to have the power to cure the symptom, often observed in classical regressions, of having economically wrong signs or magnitudes among coefficient estimates. Developments of Bayesian estimators for SEMs followed more closely the CC strategy underlying the full-information versus limited-information maximum estimation methods. The main research was carried out by Drèze (1968), Harkema (1971), Morales (1971), Rothenberg (1973), and Drèze and Morales (1976) (see also the survey by Rothenberg, 1975).[4] Technically, the need to impose high dimensional priors in SEMs ran into the difficulty of finding analytical solutions for the

[4] Rothenberg's (1973) monograph was an extension of his doctoral thesis submitted in 1966; the joint paper by Drèze and Morales (1976) first came out as a CORE discussion paper in 1970.

resulting integral calculation. Considerable efforts were therefore focused on developing methods of numerical integration as well as Bayesian computer software. An important breakthrough in numerical integration was made by Kloek and van Dijk (1978), who exploited numerical Monte Carlo integration procedures to enable integral calculation for a wider range of priors than were analytically soluble.[5] Meanwhile, many leading Bayesians also involved themselves in developing Bayesian computer programs (see Press, 1980).

These efforts to surmount the technical difficulties were vitally sustained by the philosophical appeal of the Bayesian inference—it was the route to make econometrics rigorously consistent with economics. But merely reformulating textbook econometrics by the Bayesian route was inadequate to deliver the claimed benefit. As far as empirical studies were concerned, nothing really different from what had been obtained by the classical route was presented by the few empirical Bayesian results, such as the study of the residual autocorrelation effect by Chetty (1968) and studies of the specification effect of the distributed-lag model by Zellner and Geisel (1970). Significantly different results occurred only with the imposition of tightly restrictive priors and/or small sample sizes. The situation plainly indicated disagreement between data information and highly restrictive a priori beliefs. This disagreement highlighted the possibility of model misspecification and, consequently, the role of the priors in examining the possibility, as Rothenberg (1973: 2) observed, 'either one uses prior information to improve sample estimates or one uses the sample to test the validity of the information'. Exploration of the latter route led to the Bayesian specification approach.

2.2 Emergence of the Bayesian specification search

The first Bayesian econometrician whose research was focused on the effect of model specification was E. E. Leamer. As a graduate student at the University of Michigan in the late 1960s, Leamer was perplexed by the enormous gap between textbook econometrics and applied econometric research, and especially by much of the ad hoc data mining activities involved (see 1978a: Preface). Searching for ways to patch up the gap with great sympathy toward applied modellers, Leamer was drawn to Bayesian inference by its acceptance of 'a basketful of viewpoints' on the basis of 'a single principle—Bayes Rule'. He became convinced that the rule provided flexible means to economists for 'using uncertain prior information' and 'combining disparate evidence' (Leamer, 1972a).

[5] The technique was subsequently extended by Geweke (1988, 1989) and evolved into powerful computing algorithms in the 1990s.

Leamer began his investigation with the distributed-lag model (Leamer 1972b). Unlike those who took the model for granted and who concentrated on deriving Bayesian estimators, Leamer delved into the nature of collinearity. He found that the problem had never been 'rigorously defined' and set about filling the gap (see also Chapter 7). From a Bayesian viewpoint, Leamer perceived collinearity as a problem arising from the modellers' attempts at 'interpreting multidimensional evidence' in the light of some 'undominated uncertain prior information', rather than the 'weak evidence problem' prescribed in textbook econometrics. He then narrowed it down from the problem of 'characterizing and interpreting a multidimensional likelihood function into a problem of characterizing and interpreting a multidimensional prior distribution' (Leamer, 1973: 378) and attributed the problem to inadequate specification of prior information to help 'allocate' data information 'to individual coefficients'. Leamer thus proposed measuring the degree of collinearity by 'the sensitivity of the posterior distribution to changes in the prior distribution' (Leamer, 1973). Such sensitivity analysis taught him the lesson that 'collinearity thus creates an incentive to use prior information more carefully' (Leamer, 1978a: 174). Leamer's conclusion implied that the cure for collinearity should lie with more careful parameterisation of theoretical relations in model specification.

Subsequently, Leamer grouped collinearity together with problems involving the interpretation of data evidence approximately in line with certain theoretical postulates and labelled modellers' efforts to handle these problems as 'the interpretive searches' (1978a: ch. 5). Thinking about these interpretive searches, Leamer observed that applied modellers' searches often went beyond interpretation to extend new hypotheses based on data evidence, and labelled these efforts 'post-data model construction' (Leamer1974). He likened such practice to 'Sherlock Holmes inference', that is 'weaving together all the bits of evidence into a plausible story' and ascribed the practice to the existence of too many 'viable alternatives' such that their explicit formulation was not feasible. To Leamer, the situation set inference with economic data apart from that with scientific data. In his words, 'inference with economic data is more accurately described as a prejudiced search for an acceptable model than an unprejudiced selection of one from among a well-specified set of models' (Leamer, 1974: 126) because of (i) the uncontrollable and complex nature of economic processes, (ii) economists' wide use of non-sample information, and (iii) the inclusion in models of only a small fraction of the large set of economic variables. Leamer pointed out that his Bayesian research was not intended to deal with the issue of how the search for new economic models should proceed but to provide a method for making coherent post-data model inferences.

In his attempt to make post-data model inference coherent, Leamer observed two major difficulties. One was caused by implicit prior judgements or decisions upon which data evidence was selected and presented, as the selection was therefore bound to bias the search for new hypotheses. Leamer described this bias as producing 'discounted' or 'contaminated' (i.e. not absolutely objective) data. The other difficulty was concerned with the risk of 'double-counting the data evidence' (1978a: 285), that is, the erroneous use of the same data evidence for both constructing new hypotheses and testing existing hypotheses. Contemplating ways out of these difficulties, Leamer found that neither classical nor Bayesian methods would render any ready solutions (see Hendry et al., 1990). This finding was actually an expression of Koopmans's (1949a) pessimism over the possibility of hypothesis seeking under the structural approach and confirmed Lempers' (1971) finding about the lack of straightforward Bayesian methods for choosing among different explanatory variables.

Nevertheless, Leamer (1974) reduced post-data model searches to two types: those 'adding new variables to an existing model' and those 'adding an entirely new model to the model space'. Either way, the searches were meant to start with very simple theoretical models and extend those until various econometric and economic criteria were apparently met. Leamer named the initial models 'presimplified' models and the extended models 'unsimplified' models. He showed that the prior of a presimplified model tended to prejudice inference on the model parameters, unless data grew strong enough to outweigh the prejudice. He also showed that, in cases where such strong evidence was present, the only legitimate post-data model search was to specify the additional prior, required by the model extension, in consistency with the initial prior. Leamer observed that this consistency seldom held in practice and concluded that 'post-data model construction may thus be interpreted as the data dependent decision that pre-simplification is undesirable' (Leamer, 1974: 122). Since only presimplified models were available from economics for most of the applied cases, he relegated the specification of unsimplified models to 'a supermind'. Interestingly, Leamer's characterization of the search post-data models recapitulates Theil's (1957) description of shifting 'maintained' hypotheses. But the Bayesian vein probably makes the pitfalls more transparent in pursuing such an ad hoc simple-to-general modelling strategy.

Further delving into specification problems led Leamer to attempt the taxonomy of Bayesian specification searches, in contrast to what was taught in econometrics textbooks. Specifically, the taxonomy was illustrated in a flow chart, as reproduced in Figure 2.1 (see Leamer, 1978a: 16–17), where the inputs were represented by ovals, major elements in data analysis by rectangles, specification search decisions by diamonds, and problems of statistical inference by solid directional lines as opposed to dotted lines which indicated philosophical problems beyond statistical theory. Textbook econometrics or traditional specification searches, in his view, only covered the bold components of (1), (2), (3), (6), and (7) in the flow chart, which were referred to as 'simple statistical inference', because of the 'axiom of correct specification',

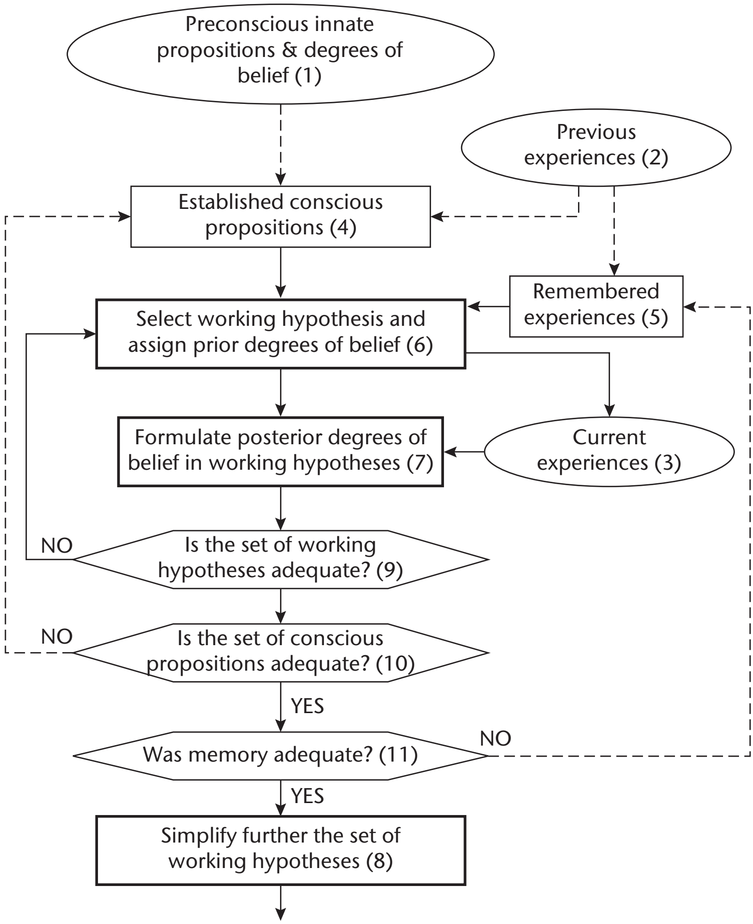

Figure 2.1. Leamer's schematic diagram of model inference
Source: Adapted from Leamer 1978a: 17, fig. 1.1.

that is there was a 'unique' and 'complete' model of the explained variable, in which all variables were 'observable', the explanatory variables were 'small in number' and the unknown parameters 'must be constant' (Leamer, 1978a: 4). Since the axiom failed to hold unambiguously in practice, empirical specification searches amounted to extending the areas marked with a bold outline by components (4) (formulating the priors), (9) (examining the adequacy of working hypotheses), and (8) (simplifying the working hypotheses), which was referred to as 'complete statistical inference' (Leamer, 1978a: 18–19). Interestingly, Leamer classified the other two decision components—(10) and (11)—as philosophical problems, although he carried out a Bayesian analysis of the link from (11) to (5) and (6) under the title '"Explaining Your Results" as

Access-Biased Memory' (Leamer 1975; see also 1978a: ch. 10). On the whole, Leamer's Bayesian specification searches were focused on formulating the priors and especially on revising them during an examination of hypothesis adequacy. These searches were further classified as 'hypothesis-testing', 'interpretive', 'proxy', and 'data-selection' searches.

It might be worth reiterating that Leamer considered the post-data model construction search as a borderline case for statistical inference (see Leamer, 1978a: 20). Note also that he placed the 'simplification searches' at the very end of his flow chart. Since the starting point models were usually of the 'presimplified' type, following textbook econometrics, Leamer was doubtful of the practical relevance of simplification searches (Leamer, 1978a: ch.6). Nevertheless, the discussion here was centred on how to use a Bayesian cost–benefit measure to evaluate the impact of dropping an apparently insignificant, or marginally significant, explanatory variable on the parameter estimates of other retaining explanatory variables, an issue which had been examined under the label of 'omitted variable bias' twenty years previously by Griliches (1957).

Leamer's taxonomy of specification searches helped to focus his methodological quest on those empirical modelling problems which fell into the gap between the 'simple statistical inference' and 'complete statistical inference'. Accepting the stance that it was the job of economists to supply theoretical models, Leamer subsequently narrowed the problems down to 'model selection' and set to building a 'conceptual framework' to assist model selection based on Bayesian statistical decision theory (Leamer 1983b).

2.3 Model selection based on statistical theory

Leamer started his model selection research from a Bayesian theorization of 'regression selection strategies' (Leamer 1978b). He saw the common practice of fitting many different regression models in order to come up with a preferred model as an informal selection process and tried to formalize it by seeking to explicitly specify the priors corresponding to those selection strategies. In the regression model context, the strategies were classified into several types, such as selection by principal component regression, by stepwise regression, or by a subjectively imposed constrained regression. These were translated into different types of priors. For instance, the common practice of deleting variables whose coefficient estimates had insignificant t values was shown to correspond to specifying uniform priors on hyperbolas. The purpose of Leamer's characterization was to reveal and identify as explicitly as possible the prior structures of modellers' working rules in model selection.

Leamer's (1983a) investigation into regression selection strategies reinforced his view that 'fragile' empirical findings were frequently the result of 'whimsical assumptions' or vague theories. Resigning himself to the fact that economists could only provide ambiguous and no one unique model specification a priori, he again resorted to the Bayesian route to establish a formal and explicit statistical procedure for model selection. Specifically, Leamer proposed dividing all the possible explanatory variables for a certain explained variable into two groups: one labelled 'free variables', that is those which were a priori assumed indispensible in economic theories, and the other 'doubtful variables'. The simplest case would be to extend the regression model (2.1) into:

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + u_t \tag{2.2}$$

where x_{1t} represents the free variable and x_{2t} the doubtful one. The division was transformed into the specification of the priors for the coefficients associated with the two types of variables. In particular, the priors of those doubtful coefficients, for example $p(\beta_2)$, should have zero means with small standard deviations. The primary goal was to study the sensitivity of those free coefficients, which was carried out by varying the priors imposed on the two coefficient sets and examining the resulting confidence intervals derived from the posteriors of the first set, for example β_1. The largest of the intervals, referred to as the 'extreme bounds', were reported and used as a key model selection criterion—models with relatively narrow and insensitive extreme bounds with respect to a wide selection of the priors should be selected while those with wide and sensitive bounds were deemed too fragile to be useful (see Leamer, 1983b, 1985a; Leamer and Leonard, 1983).

In order to keep the extreme-bounds based selection criterion viable, Leamer realized the need to start from a horizon that was as wide as possible in terms of variable choice. In his view, the horizon of conventional models was too narrow to cover the discovery of 'anomalies' from data. Including all the possible variables 'from the beginning' thus became a prerequisite for the execution of 'global sensitivity analysis' (1983a, 1985a). Note that what Leamer essentially promoted here is a general-to-specific modelling strategy. Leamer was aware of the highly probable occurrence of collinearity as the number of explanatory variables increased. But he maintained that collinearity was not a problem from the Bayesian perspective as it could be avoided by carefully selected priors.

On the other hand, the new starting point of a general model required more clarification, as a modeller had to start from a certain specific model depending on the available data and the issue of interest at hand. Leamer described such a starting model as 'pre-simplification' and stated, 'there is no formal way to assure that a given pre-simplification is optimal, and a data analysis must therefore remain an art' (1983b: 305). Simplification of presimplified models was termed as 'post-simplification' and the relevant statistical theories were viewed from the perspective of Bayesian loss functions. Here, Leamer saw the information-based Akaike (1973) criterion as essentially applying a quadratic loss function to model selection. But he was not keen on the criterion, because methodologically its underlying principle of parsimony rejected the existence of any 'true' models and practically it lacked sufficiently specific criteria on those free coefficients, criteria he regarded as the most necessary in econometric modelling (1983b). Interestingly, Leamer had used virtually the same statistics as the Schwarz (1978) criterion (also known as Bayesian information criterion) as a selection criterion under what he classified as hypothesis-testing search (1978b: ch. 4), because he saw the criterion as aiming to select a posteriori the most probable models. The link between the Schwarz criterion and the Akaike criterion was not discussed in Leamer's (1983b) work. The disregard might be due to Leamer's opinion on the rarity of hypothesis-testing search in practice at the time, as described in the previous section. He was probably too deeply absorbed in establishing selection criteria centred on those theory-based free coefficients of only 'approximately true' structural models.

However, the desire to select approximately true models by means of global sensitivity analysis was somehow unfulfilled when it came to empirical experiments. Results of the extreme bounds on free coefficients of interest often unearthed more fragile models than approximately true ones. For instance, the money demand model experiment by Cooley and LeRoy (1981) left them 'unpersuaded' by any existing models on the one hand and unable to come up with better alternatives on the other; Leamer's own efforts to apply global sensitivity analysis to modelling the inflation–unemployment trade-off resulted in extreme bounds that were too wide to help differentiate the rivalry between the monetarist and the Keynesian theories (1986, 1991)[6] (see also Chapter 5 of this book). Leamer's model selection device thus met with severe criticisms and doubts as to its usefulness (e.g. McAleer et al., 1985; Yitzhaki, 1991).

Nevertheless, empirical under-achievement had little negative influence on the Bayesian pursuit on the methodological front. One issue that Leamer left aside in his search for a systematic model selection strategy was exogeneity. A Bayesian exogeneity test was developed by Lubrano et al. (1986) in the context of a single-equation model and a test procedure in an SEM context was

[6] The explanatory variables in Leamer's (1986) inflation–unemployment trade-off model include money, GNP, oil price, labour force, private sector expenditures, government expenditure and deficit; his (1991) unemployment model had over ten explanatory variables other than dynamic terms.

subsequently devised by Zellner et al. (1988). These tests effectively extended the focus of Leamer's research from the probability distributions of structural coefficients—the βs in equation (2.2)—to the distributions of parameters relating to the properties of the residual term. The extension brought Bayesian model selection research closer to that of the non-Bayesian. Fusion of the two routes gave rise to new model selection tools and criteria, such as rival-model-comparison-based 'encompassing' (e.g. Hendry and Richard, 1989; Richard, 1995) and prediction-based 'posterior information criterion' (e.g. Phillips, 1995a). An important catalyst of the fusion was the rise of the VAR approach (see Chapter 3) and especially the Bayesian VAR as a useful tool for time-series specification search.

2.4 Bayesian fusion with time-series econometrics

Similar to Leamer's specification search and model selection, the VAR approach emerged as a remedy for the lack of a priori adequately and credibly formulated theoretical models. But while Leamer focused on the selection issues concerning multiple explanatory variables in a single-equation context, the VAR approach was primarily concerned with the dynamic specification of SEMs. More precisely, the approach advocated the use of a dynamically general VAR as the starting point, instead of any specific given SEMs, in order to overcome the inadequately specified dynamics of these SEMs as judged by data-based time-series properties of the variables involved. Nevertheless, it promoted a similar general-to-simple modelling strategy as Leamer, albeit from a different angle—that of dynamics. Technically, an immediate difficulty that the strategy encountered was the curse of dimensionality, that is the number of parameters increased drastically with the number of variables included in the VARs and the maximum lag terms needed at the initial stage. The problem of dimensionality also had an adverse effect on the forecasting performance of VARs, as improved within-sample fit via more general VARs tended to give rise to larger out-of-sample mean squared errors. Here, the Bayesian method appeared as a possible candidate to help resolve the problem (see Sims, 1980a).

Experiments applying Bayesian methods to VARs were first explored by R. Litterman in his doctoral research, which was completed in 1980 (see Litterman, 1986a). In building a monthly macro forecasting VAR model, Litterman employed the Bayesian principle of a mean squared error loss function to control the forecasting errors along with Bayesian priors to tackle the 'overparameterization' problem of general VARs. His trial imposition of priors on the lagged terms of VARs was motivated mainly by Leamer's (1972b) study of the distributed-lag model and Theil's mixed estimation technique

(1963) (see Litterman, 1979). Litterman's experiments were carried out in joint research by Doan, Litterman, and Sims (1984), which played a vital role in promoting the BVAR (Bayesian VAR) approach. The priors that Doan et al. (1984) used on the lagged terms amounted to imposing certain common restrictions on the variances or the lag shapes of the coefficients concerned. The relationship between the priors and the one-step-ahead forecasting errors was monitored and the priors which would minimize the forecasting errors were selected. Note that the priors used in BVARs had no structural underpinning, that is, they represented no a priori economic knowledge, quite different from the normal position assumed by mainstream Bayesian econometricians. Moreover and distinctly different from Leamer's strategy, the priors used by Doan et al. (1984) became the object of sensitivity analysis and effectively served as a means to achieve dynamic simplification. The model estimation was carried out recursively to study the constancy of coefficient estimates over time. Their recursive results showed that the posterior parameter estimates were insensitive to varying priors as the sample size increased, suggesting that it was possible to obtain relatively data-congruent models when they were dynamically adequately specified.

Bayesian research further deviated from mainstream econometrics when it became involved in the debate on the statistical inference of unit roots and stochastic trends in single time-series analysis (see, e.g., Koop, 1994). There was a surge of interest on the inference from the mid 1970s onwards, awakened by Granger and Newbold's (1974a) warning on nonsense regression between nonstationary variables, facilitated by the Dickey–Fuller (1979) unit-root test procedure and stimulated by Nelson and Plosser's (1982) finding of widely exhibited unit-root properties in macro time series. Technically, unit-root inference posed a serious challenge to classical asymptotic distribution theory. The BVAR experiments, however, led Sims to view the issue critically. Using Bayesian flat priors on the lagged terms of an autoregression to allow for unit roots, Sims (1988) showed that the Bayesian inference was unaffected by unit roots and could generate finite sample results different from those generated by the classical route (see also Sims and Uhlig, 1991). The difference made Sims doubt the validity of unit-root tests based on the classical approach. Moreover, he disapproved of the unit-root properties because of the lack of a solid foundation for economic behaviour.

Sims's Bayesian illustration of the unit-root issue was criticized by Phillips (1991), who showed how the choice of different priors would affect the Bayesian inference about unit roots and that flat priors were improper for the unit-root discussion while more appropriate choices of the priors would generate Bayesian results comparable to those from classical inference. However, the criticism left untouched Sims's argument about the economic foundation. The Bayesian unit-root discussion led Phillips on to explore the use of Bayesian methods for data-based dynamic model selection. He proposed a posterior information criterion (PIC) to help determine the lag order and trend properties of the variables in dynamic models and devised a forecast-encompassing test to assist a data-based search for parsimonious dynamic models (see Phillips, 1995a; Phillips and Ploberger, 1991, 1994).[7] Interestingly, the test was shown to provide a Bayesian alternative to the forecasting encompassing principle via the classical route developed by the LSE dynamic specification modelling approach, to be described in Chapter 4.

Meanwhile, the idea of applying Bayesian methods to VAR models was extended to dynamic stochastic general equilibrium (DSGE) models. A commonly recognized weakness of DSGE models was the arbitrary use of 'calibrated' parameters. A handy way to tackle this weakness was to assign Bayesian prior distributions to those parameters and simulate the model outcomes in probabilistic terms. The idea was initially explored by Canova (1994) and also DeJong et al. (1996). Within the VAR model framework, the Bayesian methods were also extended to the production of confidence bands for impulse responses (e.g. Sims and Zha, 1998). These extensions helped to revitalize Bayesian econometrics (e.g. van Dijk, 2003; and also Geweke et al., 2011 for more detailed surveys of subsequent developments). Noticeably in the revitalization, the subjectivist image of the Bayesian approach was weakened as econometricians' understanding of the versatility of Bayesian inference in model estimation broadened, and Bayesian econometrics reverted to an alternative technical division rather than a methodological one.

2.5 Methodological reflection

The history of Bayesian econometrics serves as a good illustration of the evolution of econometric research—moving from paradigm consolidation towards methodological reforms and then towards more versatile applications. The Bayesian entry was primarily motivated by the CC programme. Its research agenda was focused on the measurement of structural parameters of a priori given models. It sought to improve the measurement by explicitly specifying, into prior distributions, the available uncertain information concerning the structural parameters. It took pride in strengthening the rigour of parameter estimation via its probability foundation. In that respect, the Bayesian movement could be seen as the faithful and fundamental extension of Haavelmo's (1944) probability revolution. However, early Bayesian efforts to emulate and reformulate textbook econometrics have not generated

[7] Phillips's (1995a) paper came out first in 1992 as a Cowles Foundation Discussion Paper.

extensive impact, as shown in Section 2.1. Technically, Bayesian methods were hampered by the computational difficulty of solving high dimensional integrals of posterior distributions. Empirically, the limited Bayesian applications failed to present distinctively better results than those achieved by the classical route. Methodologically, many modellers were put off by the explicit subjectivist philosophy of the Bayesian approach.

It is, however, Bayesian subjectivism which helped to alienate Bayesian econometricians from textbook econometrics and eventually led them to challenge it, as shown by Leamer's specification searches and model selection investigation described in Sections 2.2–2.3. By digging into the gap between textbook econometrics and empirical modelling and trying to amend it in a systematic manner, Leamer has distinguished Bayesian research as being at the forefront of developing econometric methodologies (see Pagan, 1987). Interestingly, the Bayesian subjectivism provided Leamer with a convenient, overt, and formal explanation for arbitrariness in model choice, a key issue left un-formalized in the CC programme. The Bayesian perspective enabled him to deliver a sharp exposition of the consequence of this incomplete task—that is the widespread fragility of structural parameter estimates from directly estimating a priori formulated structural models.

On the other hand, much of Leamer's discussion on model mis-specification issues had a precedent in non-Bayesian econometricians' writings, for example, Theil (1957; 1961). Presenting the discussion in a Bayesian vein was probably the most innovative part of Leamer's analysis. The prevalent textbook method of econometric practice might be too strong to make Leamer break with the theory-dominant tradition of model formulation. But Bayesian subjectivism—over-subjective theories that led inevitably to fragile structural parameter estimates—undoubtedly provided him with a handy excuse not to leap into a heavily data-driven modelling approach. Hence, there is a discernible ambivalence in Leamer's approach toward the CC paradigm. This approach critically exposed the major drawback of the paradigm but, at the same time, it tried to defend and rigorously amend the measurement task of given structural parameters, that is, the parameters of 'free variables', a task which forms the core of the paradigm. Such ambivalence limits the methodological distinctiveness of the Bayesian approach.

The subsequent fusion of Bayesian methods with time-series econometrics, particularly the VAR approach, has apparently further distanced Bayesian econometrics from the traditional paradigm. But ironically, the demarcation of Bayesian econometrics as opposed to classical econometrics has largely been blurred during the fusion, and with it its methodological distinctiveness. Bayesian priors have become more widely accepted as the means to produce 'more useful, honest and objective reporting' of empirical results in explicit probabilistic terms (Sims, 2000). They have also been extended into the toolkit of macroeconomists with the result that the specification of the likely distributions of structural parameters has come to be considered part of the theoretical model formulation, for example in the case of dynamic CG models, and see also Eichenbaum (1995).

This brings us to the question of whether Bayesian econometrics has adequately distinguished itself into a methodological school. History has shown us that Bayesian methods are capable of results identical or comparable to those derived by classical statistic methods, regardless of whether they are from models built on the basis of the traditional paradigm or any other alternative modelling approach. The comparability stands at odds with the Bayesian belief that its advantage lies in its explicit subjectivism and its ability to be updated in accordance with emerging evidence. What has made such an ideological divide unnecessary in practice? Reflection on the design of various Bayesian estimators, especially from the sequential method by Tiao and Zellner (1964a) to Litterman's (1979) BVAR device, may provide us with an answer. There is a disparity between the Bayesians' ideological claims and Bayesian inference in practice. The inference has hardly ever worked with the inverse conditional distribution defined by Bayes' formula (cf. Qin, 1994). Indeed, the workhorse of Bayesian econometrics is the product, or joint distribution, of a prior distribution with the likelihood function. As data evidence accrues, the impact of the prior diminishes, yielding empirical results predominantly based on data information from likelihood, especially when the prior is modified by a subsample likelihood result, as in the case of recursive estimation. The 'Bayesian' approach is de facto a misnomer for the 'joint distribution' approach.

The relative independence of Bayesian methods from the Bayesian ideology and the absence of strikingly empirical distinctness may explain the phenomenon that the Bayesian approach has largely retained its popularity among theoretical econometricians as well as some empirically minded macroeconomists. Here, it is interesting to observe that the joint distribution of a prior with the likelihood function is regarded as equivalent to 'the *model*' (Poirier, 1988). Such a Bayesian perspective shows us how much Bayesian methods have extended as well as reoriented the field of vision of theoretical econometricians from a given structural model to its derived parameter space in terms of its sampling distributions. The extension has greatly enhanced and enriched the potential of econometric research into various issues relating to the distributions of the parameters within a given model, as rightly foreseen by Bayesian pioneers such as Drèze and Zellner. On the other hand, the parameter space as derived or transposed from the original model space also confines research attention within parametrically related issues, leaving aside those issues which are pertinent to a post-data model search or data-instigated model design before the set of parameters of interest become clearly specified (see Chapters 7 and 9 respectively for more discussion on the history of parameter specification and that of model selection and design). Consequently, the primary Bayesian contribution lies in the refinement of the probability foundation within a given modelling approach. Its capacity to extend or develop alternative approaches is often restricted by what it could observe as model search or design problems via the sampling distribution window of the derived parameter space. In that light, the contribution of Bayesian econometrics is more directly and substantially concerned with 'internal questions' than 'external questions' in the development of econometric methodologies (Carnap, 1950).

3

Rise of the VAR Approach

The VAR approach, which was built on the 'Vector AutoRegression' model form, emerged roughly at the same time as Leamer's Bayesian approach on model specification searches was taking shape.[1] Unlike Leamer's Bayesian approach, which targeted the general model estimation methodology within econometric practice, the VAR approach was developed to deal specifically with macro-econometric models using time-series data. The development thus evolved in close interaction with developments in macroeconomics, especially in the USA. The present chapter makes a historical assessment of the rise of the VAR approach by describing how the approach came about, what issues it endeavoured to resolve, and what methodological position it assumed, particularly with respect to the CC tradition.

It must be noted that the use of VAR models long precedes the rise of the VAR approach. The model was designated as the 'reduced form' of a structural model in its most general form in the CC tradition. Empirical studies of the dynamic properties of macro-econometric models by means of a VAR could be traced back to Orcutt's (1948) work on the famous Tinbergen (1939) model. The theoretical pursuit of efficient estimators specifically for the dynamic side of simultaneous-equation models (SEM) was explored by Sargan (1959). The early promotion of VARs as the dynamic representation of structural models was advocated by H. Wold under the banner of 'causal chain models' as opposed to the CC SEM approach (eg see Wold, 1954, 1960, 1964). However, it was not until the joint work of Sargent and Sims (1977), presented as a conference paper in 1975, that the VAR approach emerged as an alternative methodology to the CC-tradition-based mainstream econometrics (e.g. Pagan, 1987).

A major impetus for this chapter comes from the marked disparity between the 'atheoretical' attribute of the VAR approach popularly held within the econometrics profession and the devotion of many VAR modellers to

[1] This chapter is adapted from Qin (2011a).

identification, causality test, and structural modelling to support policy analysis—issues close to the heart of the CC tradition. Historical investigation reveals that the 'atheoretical' attribute is a partial and somewhat extremist label. The VAR approach has actually resulted from a fusion of the CC tradition and time-series statistical methods, catalysed by the rational expectations (RE) movement. It has arisen from and been nourished by extensive interplay between macroeconomics and econometrics.[2] It offers a systematic way of resolving the issue of model choice with respect to the dynamic side, an issue left out as by the CC group.

The rest of the chapter is organized in six sections as follows: Section 3.1 mainly describes the early econometric activities of C. Sims, the acknowledged leader of the VAR approach; Section 3.2 traces the RE movement with particular emphasis on the use of VARs. These first two sections provide a prelude to the VAR approach. Section 3.3 is focused on Sargent and Sims's (1977) joint work, which effectively launches the VAR approach. Section 3.4 describes Sims's (1980a) well-known manifesto of the VAR approach. Section 3.5 describes the emergence of the structural VAR (SVAR) approach. Section 3.6 gives a methodological assessment of the rise of the VAR approach.

3.1 Dynamic specification gap: from theory to data

The 1960s saw increasing efforts in the dynamic and stochastic modelling of agents' behaviour. Exemplary works include models for inventory control (e.g. the joint research by Holt et al., 1960), studies emulating the Koyck (1954) model in the area of production and investment (e.g. Jorgenson, 1963; Griliches, 1967), and studies on the dynamics of demand (e.g. Houthakker and Taylor, 1966). It was in this context that Sims undertook the mission to bridge 'the big gap between economic theory and econometric theory' (Hansen, 2004) when he started his econometrics career.

Sims's initial work was to study the effects of dynamic specification errors under the assumption that a theoretical model was known and given. Using a continuous-time distributed-lag model as the given theoretical model, he demonstrated two dynamic specification gaps in applying the model to data—one was the approximation of the continuous-time model by a discrete-time

[2] Much of the cross-fertilization occurred at the University of Minnesota, where Sargent, Sims, and Wallace, the key players of the VAR approach or RE models, were working in the economics faculty during the 1970s, e.g. Hansen (2004) and Evans and Honkapohja (2005). Apart from university duties, they were also involved part-time at the Federal Reserve Bank of Minneapolis. Most of the influential papers on RE models and the VAR approach appeared first as the FRBM Working Papers, see the Bank website archive <http://www.minneapolisfed.org/research/economic_research/>, last accessed 3 February 2013.

model (Sims, 1971) and the other the approximation of an infinite lag distribution by a finite lag distribution (Sims, 1972a). The two studies led him to the view that 'the usual approach … to proceeding to "efficient" estimation and inference as if the [empirical] model were exact, can lead to serious errors' (Sims, 1972a: 175). These studies are reminiscent of Haavelmo's (1943) discovery of the OLS 'bias' in the context of SEM.

However, Sims soon switched his research angle. He began to investigate the dynamic specification gap from data to theory, instead of starting from a given theoretical model. An important influence appears to be Granger's (1969) introduction of what is now known as the Granger causality test.[3] The test was designed to detect sequential causality between time-series variables and the design was inspired by mathematician N. Wiener. The test rules out one variable causing another if the past of the former exerts no impact on the latter in a dynamic bivariate relationship, which is effectively a VAR. More precisely, the test decides that y does not cause z if all the parameters $\pi_{21,i}$ are shown to satisfy $\pi_{21,i} = 0$ in the following bivariate VAR:

$$\begin{aligned} y_t &= \sum_{i=1}^{p} \pi_{11,i} y_{t-i} + \sum_{i=1}^{p} \pi_{12,i} z_{t-i} + u_{1,t} \\ z_t &= \sum_{i=1}^{p} \pi_{21,i} y_{t-i} + \sum_{i=1}^{p} \pi_{22,i} z_{t-i} + u_{2,t} \end{aligned} \tag{3.1}$$

where u_1 and u_2 are white-noise errors.

The significance of Granger's (1969) paper was recognized almost immediately by Sims for its close link to Wold's causal chain model, since Granger's notion of causality departed from the traditional static notion built upon SEMs. Sims (1972b) applied the Granger causality test to the money–income dynamic relationship in order to show how time-series information could help differentiate a 'unidirectional causality' concerning rival theories on the relationship between money and income. Sims's application was immediately followed by Sargent. He adopted the Granger causality test as the means to empirically assess the RE hypothesis (e.g. Sargent, 1973b; Sargent and Wallace, 1973; and Chapter 7 of this book). Subsequently, the Granger causality test became a core tool used by macroeconomists (see the next section and also Sent, 1998: ch. 3).

Sims's (1972b) paper elicited a great deal of interest 'because it came out at the peak of the monetarists–Keynesian controversy' (Hansen, 2004) and also because it fed into the methodologically controversial issue about the extent to which econometric models were capable of verifying behaviourally causal hypotheses. In response, Sims (1977a) put forward a substantive clarification on 'exogeneity

[3] Granger was first engaged in economic time-series research at Princeton University under the directorship of Morgenstern, see Phillips (1997).

and causal ordering in macroeconomic models'.[4] Here, Sims related Granger causality to the 'strictly exogenous' concept and presented Granger causality as an augmentation of Wold's causal ordering by making the ordering testable.

Interestingly, Sims (1977a) also discussed the position of Granger causality with respect to the CC approach on the general setting of an SEM. In his view, 'numerous maintained hypotheses of exogeneity' were needed for the implementation of the SEM, hypotheses which could easily lead to Granger causal ordering. Sims realized that the choice between SEM and Wold's causal chain model would inevitably invoke the issue of the identifiability of structural models. He tried to define 'structural' relations by two key attributes, following closely Hurwicz's (1962) definition. The first states that 'being "structural" is a property of the way we interpret the system as applying to the real world, not of the system's form'; and the second states that 'being behavioural is neither necessary nor sufficient to make a relation structural relative to an interesting class of possible interventions' (Sims, 1977a: 28–30). The latter was made in particular reference to the Lucas (1976) critique (see Section 3.2). Sims maintained that the essence of econometric modelling was to search for a data-coherent causal model 'to buttress a claim that the model is behavioural or structural relative to variations in the path of *x* [the forcing variable in the model] as identifying interventions' (Sims, 1977a: 31) and that the Granger causality test was a useful tool in that connection, even though there lacked sufficient ground to verify a model displaying Granger causality being indeed a 'structural' model.

Sims's (1977a) paper paved the way for the VAR approach. It demonstrated his research interest in methodological matters with particular attention on filling the model specification gap between data and economic theory. Unfortunately, the demonstration has not been recognized widely in econometric circles, probably because Sims's (1977a) paper has largely been overshadowed by his (1980a) critique.[5]

3.2 The rational expectations hypothesis and VAR model

The early 1970s witnessed major reform in macroeconomics. Widely known now as the RE movement,[6] this reform set out to configure macro theories into both dynamically testable and behaviourally optimizable models. In

[4] The draft of the paper was presented at a conference in 1975 sponsored by the Federal Reserve Bank of Minneapolis, see Section 3.2.

[5] According to Sims, 'virtually nobody has read and understood' his (1977a) paper (Hansen, 2004).

[6] The RM movement is commonly viewed as initiated by Lucas (1972a). There are numerous studies on the history and the methodology of the RE movement (e.g. Sheffrin; 1983; Maddock 1984; Sent 1998; and see also Sent 1997, 2002; Young and Darity 2001, for the early history of RE models).

the eyes of key RE proponents, macroeconomics was weak both technically and methodologically, especially in comparison with CC structural econometrics, and the weakness lay mainly in the lack of micro foundations in Keynesian methodology; empirically, the weakness had led to poor forecasts and misguided policy simulations by existing macro-econometric models (e.g. Lucas and Sargent, 1978; Sargent, 1980).

The RE movement stems from the RE hypothesis originally formalized by Muth (1961) in a microeconomic context. The RE hypothesis emphasizes the importance of considering agents' expectations of the modelled variables in a behavioural model; it postulates that the expectations be formulated from all the available information relevant to the model and that the useful content in the information lies with the past series of innovation shocks contained therein. Essentially, an RE model can arise from augmenting a simple static model, say a bivariate one:

$$y_t = \beta z_t + \varepsilon_t \tag{3.2}$$

by an expectation term, y^e, of the endogenous variables y_t:

$$y_t = \beta z_t + \lambda y^e_{t+k} + u_t. \tag{3.3}$$

When $k=0$, forward expectations are absent.[7] The RE hypothesis models the latent y^e_t on all of the available information set $\{I_{t-1}\}$, that is, the history of y_t and z_t in the present case:

$$\begin{aligned} &y^e_t = E\left(y_t \middle| \{I_{t-1}\}\right) = E\left(y_t \middle| \{y_{t-1}\},\ \ \{z_{t-1}\}\right) \\ &\Rightarrow \quad y_t - y^e_t = v_t \quad \Rightarrow \quad E\left(v_t \middle| \{I_{t-1}\}\right) = 0 \quad Var\left(v_t \middle| \{I_{t-1}\}\right) = Var\left(v_t\right) \end{aligned}. \tag{3.4}$$

As a result, the expectation error, v_t, should follow an innovative process, as shown in (3.4). Taking the conditional expectation of equation (3.3), we get:

$$E\left(y_t \middle| \{I_{t-1}\}\right) = (1-\lambda)^{-1} \beta E\left(z_{t-j} \middle| \{I_{t-1}\}\right) = (1-\lambda)^{-1} \beta z^e_t. \tag{3.5}$$

Equation (3.5) transforms the RE hypothesis into a model where the explanatory variable becomes the latent expectation of the original exogenous variable. The consequent need to model z^e_t is then referred to as the requirement of 'completing' the model (e.g. Wallis, 1980). For example:

$$z^e_t = E\left(z_t \middle| \{I_{t-1}\}\right) = \sum_{i=1}^{p} \pi_{21,i} y_{t-i} + \sum_{i=1}^{p} \pi_{22,i} z_{t-i}. \tag{3.6}$$

[7] The assertion of forward expectations brings a new technical issue into econometrics: the need for terminal conditions to assist unique solutions (e.g. see Pesaran, 1987: ch. 5). However, this issue does not add to nor heighten our methodological discussion here. For simplicity, we will only consider RE models with current expectations hereafter.

Combining (3.5) with (3.4) results in a closed VAR of the same form as (3.1).[8] This VAR is 'reduced' from the structural model (3.3) combined with the RE hypothesis in (3.5). It should be noted here that the derivation from the RE model to a VAR involved a modification of the implicit assumption of the RE model (3.4) that the dynamic model had an infinite lag length structure. The VAR could only have a finite lag structure, that is, denoted by p in (3.1), to be empirically operational. However, the practical implications of the assumption and its modification were left unheeded for a long time.[9]

The RE movement thus foreshadows the VAR approach in two key and interrelated respects. It highlights the need to formulate a closed model in the sense that all the variables considered in a model should be regarded as potentially endogenous; and it rationalizes a general dynamic specification by the RE hypothesis.

At the time, the first aspect gave rise to a powerful critique by Lucas (1976) on the use of macro-econometric models.[10] Essentially, Lucas' critique questioned the legitimacy of using constant parameter macro-econometric models for policy simulations because shifts in policy would affect the constancy of certain structural parameters. In terms of the simple RE model given above, policy shifts amount to value changes in π_{22} when z represents the policy instrument; the changes would obviously affect π_{12} of the first equation of VAR (3.1) (see footnote 8), thus destroying the constancy of that equation.[11] The Lucas critique has provoked a great deal of rethinking about the practice of a priori categorization of endogenous versus exogenous variables. Attempts to make the categorization testable led to further popularization of the Granger causality test (see Sargent, 1976a, 1978b).

Meanwhile, an important demonstration of the second aspect was Sargent's discovery of the 'observational equivalence' problem (1976b). In an effort to make it empirically possible to differentiate the impacts of Keynesian policy shocks versus shocks based on classical theories, Sargent stripped down what he considered to be a general structural model into a Wold representation:

$$\begin{aligned} y_t &= \sum_{i=0}^{\infty} \delta_{11,i} u_{1,t-i} + \sum_{i=0}^{\infty} \delta_{12,i} u_{2,t-i} \\ z_t &= \sum_{i=0}^{\infty} \delta_{22,i} u_{2,t-i} \end{aligned}, \tag{3.7}$$

[8] In the VAR, the parameters of the first set of equations are function of the parameters of (3.3) and (3.6), i.e.: $\pi_{11,i} = (1-\lambda)^{-1} \beta\pi_{21,i}$ and $\pi_{12,i} = (1-\lambda)^{-1} \beta\pi_{22,i}$.

[9] See Hendry and Mizon (2000) and Hendry (2002) for a critique of RE models concerning this assumption in particular in relation to model-based forecasting.

[10] The paper was first circulated in 1973 as a Working Paper at Carnegie-Mellon University.

[11] Notice that Lucas' original presentation did not use parameter shifts as such. He analysed the policy variable in terms of its permanent and transitory components. A shift in the permanent component is equivalent to a parameter shift in the autoregressive representation of the variable.

which was effectively an inversion of VAR (3.1) into a moving average (MA) model. Conceptually, Sargent assumed (3.7) as the 'structural' model by regarding the innovation errors of the RE hypothesis as policy shocks. In other words, if policy shocks are defined as unanticipated deviations of the policy variable from its expected path, equation (3.5) can become:

$$E\left(y_t \left|\{I_{t-1}\}\right.\right) = \sum_{i=0}^{\infty} \delta_{11,i}(z_{t-i} - z^e_{t-1}) + \sum_{i=1}^{p} \pi_{11,i} y_{t-i}, \tag{3.8}$$

which is equivalent to the y equation in (3.7). It is important to note that the structural status of (3.7) has departed substantially from the CC structural model framework and has echoed the Slutsky–Frisch impulse-propagation scheme (see Frisch, 1933; Slutsky, 1937). Nevertheless, it is obvious that (3.7) can be transformed into a VAR, so that the 'reduced-form' status of the VAR remains intact. Since the first equation in (3.7) embraces theories of both RE and non-RE types depending on how the shock variables are formulated, the corresponding reduced-form VARs set the empirical limit to the testability of the economic hypotheses concerned.[12]

For econometricians at the time, the RE movement essentially endorsed the VAR by assigning it a central position in bridging time-series econometrics with macroeconomics. Moreover, it turned their attention from estimation and identification of structural parameters of a priori formulated models towards testing and searching for data-coherent models. The VAR approach effectively emerged from the starting point of such searches.

3.3 Emergence of the VAR approach

The blueprint of the VAR approach was first presented at a conference on business cycle research in November 1975 sponsored by the Federal Reserve Bank of Minneapolis. It was published as a joint project between Sargent and Sims (1977),[13] entitled 'Business cycle modelling without pretending to have too much a priori economic theory'. The title emphasized their opposition to the common practice of assuming that structural models were a priori set and well-formulated. Sargent and Sims pointed out that many of those models lacked theoretical foundation and empirical support, thus reducing

[12] From the viewpoint of some macro theorists, however, this limit implies a lack of identification power of econometric models with respect to many theoretically interesting but parametrically sophisticated RE models. As a result, empirical macroeconomics branches into two directions (see, e.g., Summers, 1991), one still pursuing macro-econometrics, the other abandoning statistical methods to develop computable dynamic general equilibrium models (e.g. Kydland and Prescott, 1991, 1996).

[13] The conference proceedings were published in 1977 under the sponsorship of the Federal Reserve Bank of Minneapolis.

the practice to 'measurement without theory', a famous criticism with which Koopmans (1947) had charged the NBER thirty years earlier (see Qin, 1993: ch. 6). Having rejected the common practice, Sargent and Sims searched for a new, alternative modelling route in order to improve the credibility of the theoretical content of macro-econometric models.

Broadly, the Sargent–Sims alternative contained the following steps: First, they compared the NBER method of deriving a 'reference cycle' indicator of business cycles from a selected group of variables with the conventional macro-econometric model approach of explaining business cycles by modelling key macro variables, and decided to follow the latter.[14] Next, they showed VARs to be the general form of conventional macro models. They then proceeded to estimate the VAR and sought data-coherent ways of simplifying the VAR in order to identify and compare data-based model features with what had been postulated a priori in the form of a particular macro model nested in the VAR.[15] Finally, the simplified VAR was transformed into an MA model to enable the application of impulse response analysis for further assessment of the dynamic properties of the model and also for policy simulations.

To compare the Sargent–Sims route with the CC tradition, a closer examination is needed. Their choice in the first step appears to be in line with CC procedure, that is, to start econometric modelling from known theory. However, the structural model they chose is not SEM but a distributed-lag model, that is, a special case of the first equation in (3.1) or a dynamic augmentation of (3.2):

$$y_t = \sum_{i=0} \beta_i z_{t-i} + \varepsilon_t. \tag{3.9}$$

The model thus shifts the discussion from simultaneity to exogeneity and dynamics, a shift which virtually mirrors the RE movement in macroeconomics. Although its link to VARs follows naturally and apparently stays on par with the step of SEM $\rightarrow$ reduced-form VARs in the CC tradition, the motivation is distinctly different. Instead of utilizing reduced-form models to facilitate estimation, Sargent and Sims use the link to justify Liu's (1960) argument to start empirical modelling from unrestricted VARs. They find Liu's strategy very useful for abandoning the questionable assumption of a priori known structural models. But whereas Liu wants to produce better forecasts, Sargent and Sims's primary goal is to make the theory testable, which extends and enhances the CC tradition. Their goal renders crucial the third step of VAR simplification. Here, particular attention is given to testing exogenous

[14] They referred to the NBER method as the 'unobservable-index model' approach and the conventional macro-econometric model approach as the 'observable-index model' approach; they also reformulated the NBER method by factor analysis, see Chapter 6.

[15] This objective is plainly phrased by Sims (1977a: 5) as trying to 'use available data to increase our understanding of economic behavior'.

variables in order to identify data-coherent causal ordering of the equations in the VAR. It should be noted that the concept of identification is no longer equivalent to that in the CC procedure. It has blended in the Box–Jenkins (1970) time-series notion of identification. Another source of VAR simplification discussed is cross-equation linkages/restrictions postulated in various theories—restrictions which form a natural ground for testing the theories concerned. Finally, the utilization of the simplified VAR to conduct impulse response analysis via its MA representation shares the CC conviction that the ultimate aim of econometric modelling was for policy purposes.

Sargent and Sims (1977) demonstrate their new approach via empirical experiments with two simple index models—a Lucas model with one unobservable index and a five-variable Keynesian model. They note, among the findings, that model results can be highly sensitive to the choice of variables at the outset, especially those variables which are potentially active indices (forcing variables), and that some variables are associated with more indices than the others within the chosen variable set, suggesting a certain causal ordering of the system.

The Sargent–Sims joint venture attracted more criticism than approval (see the comments on their paper in the conference proceedings (Sargent and Sims, 1977)). In their defence, Sims emphasized that the new approach was developed more for hypothesis testing and evaluation than forecasting, see (Sims, 1977b), whereas Sargent (1979) acknowledged that the VAR approach was more for prediction than policy evaluation. Their apparently contradictory interpretations reflect the essential spirit of the Sargent–Sims joint venture. For a macro theorist, it is an opportunity to seek a systematic route to producing empirically operational theories, whereas for an econometrician, it is an opportunity to strengthen the theoretical underpinning of empirical models. The synergy adheres to the broad CC principle, but it was hidden almost completely by the combative veneer of their undertaking.

3.4 Manifesto of the VAR approach

A fuller methodological exposition of the Sargent–Sims joint venture is given by Sims (1980a) in a paper entitled ‘Macroeconomics and Reality’. This article has been widely regarded as the manifesto of the VAR approach as a departure from the CC structural modelling tradition.

The opening discussion of the manifesto was focused on the issue of identification, as implied by the title, ‘Macroeconomics and Reality’. The discussion echoed what had been argued in Sims’s joint paper with Sargent (1977). It drew its main inspirations from the then-recent RE movement; its specific attack was on those ad hoc dynamic restrictions frequently used for conventional identification purposes; it defined the research interest within the domain of business cycle modelling and stressed the importance of adopting a fully dynamic and closed model in order to capture the possible impact of the changing time paths of any policy variables; and more fundamentally, it anchored its methodology on building models for the purposes of theory testing and policy analysis, rather than merely forecasting, although it regarded its position as comparable to Liu's (1960) critique.

Compared to the Sargent and Sims (1977) paper, the identification discussion here substantially broadens the theoretical justification for the choice of the VAR route. The justification builds on the dynamic interaction among macro variables which constitutes the essence of RE models, with specific reference to the Lucas (1976) critique concerning the validity of the exogenous assumption of policy variables. In fact, the discussion points to a broad correspondence between a VAR and the fundamental theme of dynamically general equilibrium in macroeconomics. Such a correspondence was absent in Liu (1960), when the reduced-form VAR model was mainly valued for its data summary and forecasting capacity.

Having laid out a lengthy justification for his 'alternative strategy', Sims proceeded to describe the VAR approach in detail. The starting point was to set up a closed and unrestricted VAR to summarize data. There were two main technical issues to be considered: whether the VAR would adequately summarize data information and whether any patterns or regularities could be revealed in the summary. The former was achieved via the choice of lag length and the latter via a comparison of estimation results over different subsamples. Next, a possible simplification of the VAR was considered. A crucial simplification source was those 'hypotheses with economic content' which would normally limit 'the nature of cross-dependencies between variables' (Sims, 1980a: 15–16). While the initially unrestricted VAR facilitated hypothesis testing here, those hypotheses which survived the testing would also bring certain structural interpretations to the model. The simplified VAR was then transformed into an MA model to enable impulse response analyses. The MA representation was considered most useful in subjecting the model to 'reasonable economic interpretation', as well as in facilitating hypothesis testing via comparing the dynamic paths of shocks assumed of policy variables versus what was implied in the hypotheses concerned. Through a detailed illustration of the possible routes by which a monetary innovation shock could affect a real output variable, Sims endeavoured to show how the VAR-based MA model could empirically help define 'what battlefield positions must be' between rival theories.

Interestingly, Sims (1980a) attributes the hypothesis testing capacity of the VAR approach to the way in which the approach releases the hypothesis of interest from the 'burden of maintained hypotheses' under the conventional structural approach. The burden is essentially a consequence of the CC strategy of leaving aside the issue of 'model choice' when the group formalized econometric methods (e.g. Qin, 1993: ch. 6). In the 1950s, Theil attempted to address this 'maintained hypothesis' problem explicitly by proposing an experimental approach of using an array of mis-specification checks in a piece-meal manner (e.g. Theil, 1961). That approach becomes generalized in the dynamic specification aspect under the general → specific procedure of the VAR strategy. It should also be noticed how macroeconomics has evolved over the two decades. The RE movement, in particular, has turned theorists' attention from parameterizing behavioural propensities to tracing the sources of observed cyclical fluctuations and their dynamic characteristics. Sims's illustration reflects such a shift. It is none the less an identification attempt within the broad spirit of the CC enterprise, albeit one that has moved away from those tightly parameterized structural models underlying the CC identification conditions.

However, the very act of moving away has laid the VAR approach wide open to criticism. In the same way as his joint work with Sargent (Sargent and Sims, 1977), Sims's (1980a) paper elicited more suspicion than appreciation among econometricians. The VAR approach was branded 'atheoretical macroeconometrics' (see, e.g., Cooley and LeRoy, 1985), and described as dissenting 'vigorously from the Cowles Commission tradition' (Pagan, 1987), but was 'not an adequate substitute for the Cowles program' for drawing causal inferences (Leamer, 1985b).

3.5 Emergence of structural VARs

Once the ground-breaking work was done, research into the VAR approach turned to further improving the VAR techniques. The focal issues were, however, closely pertinent to the structural/causal interpretation of the VAR results. The improvement efforts led to the revision of the VAR approach into the 'structural' VAR (SVAR) approach. This revision can be seen as a natural defensive reaction to the criticisms it had encountered. It nevertheless reinforces the commitment of VAR modellers to build models that are useful for policy purposes, a commitment shared by modellers following the CC tradition.

A vital prerequisite for the interpretability of the VAR results is model simplification, as a general VAR is known to suffer from the curse of dimensionality—the number of parameters growing with the number of variables and the maximum lag in a multiplicative manner. A primary technical issue on the VAR research agenda was to find a systematic 'shrinkage' strategy. This led to the joint work of Doan et al. (1984), which implemented Sims's

(1980a) conjecture that the Bayesian approach might offer such a strategy.[16] Doan et al. (1984) imposed a number of Bayesian priors on each equation of an unrestricted VAR to help simplify mainly the lag lengths and possible time variation of the parameters. The equations were estimated one by one using the same set of priors to keep the symmetry of the VAR. A data-coherent and simplified VAR resulted from numerous experiments through the adjustments of the priors. Since parameter constancy was vitally important to allow the model to be applied to simulations of policy shocks, Kalman filter and recursive estimation were used to examine the time-varying feature of their parameter estimates. To their relief, most parameter estimates revealed few time-varying features in the experiments.[17]

However, the most controversial techniques were related to the impulse response analysis, which became labelled 'innovation accounting' by Sims (1978).[18] As the MA representation of (3.7) was designated the structural model status, it became crucial to justify the structural interpretability of innovation accounting. Obviously, the issue was intimately related to identification. But Sims found himself facing an apparently new identification problem—the impossibility of identifying an impulse shock with a particular structural shock if the error terms in the MA model were contemporaneously correlated across equations. Technically, he saw the solution to lie in orthogonalizing the error terms. Conceptually, he found that solution to be closely linked to Wold's causal ordering, which could therefore be tested by the Granger causality test. Sims thus proposed an experimental approach—use the Granger causality test to simplify the cross-equation interdependence between variables, examine the degree of cross-equation correlation between the residual terms of the simplified VAR, and then experiment with different Wold's causal orderings on the transformed MA; the experiment was to check either that a particular theory-based ordering would generate the kind of impulse responses expected by the theory or which data-based ordering would be the most robust and have the best interpretation in terms of theory.

This experimental approach elicited immediate criticisms, mainly of two aspects. Methodologically, it was excessively data driven. Technically, it further blurred what should be considered as the 'structural parameters' of the resulting model. In other words, the approach amounted to transforming those assumed 'structural shock' variables, u_{1t} and u_{2t}, in model (3.7) into another set of uncorrelated error terms; but it was impossible to transform

[16] An initial experiment with a Bayesian VAR approach was explored by Litterman (1979) around the time Sims was preparing his (1980a) critique.

[17] Several interesting results emerged in the Doan et al. (1984) experiment, including the assumption of a unit root in each variable, i.e. the prior on the own first lag taking the value of one.

[18] Sims (1980a) was first distributed as a University of Minnesota Economics Discussion Paper in 1977, and a part of his (1978) paper was later published in the book edited by Kmenta and Ramsey in 1982.

the structural status to the new set unless the orthogonalizing parameters used in the transformation were established as structurally interpretable.

Efforts to allay the criticisms led VAR researchers to a significant retreat from the data-driven position. After all, a primary agenda of the VAR research was to develop VAR-based policy evaluations (e.g. Sims, 1986). One obvious alternative was to seek orthogonalizing methods from conventional economic theories. For instance, a 'structural' identification method for the error terms was initiated by Blanchard and Watson (1984) and extended by Bernanke (1986). The method essentially exploited the relationship of the error terms between a conventional SEM and its reduced form. Interestingly, the method moved the SVAR approach closer to the CC tradition. It resorted not only to an SEM for support of the structural interpretation of the shock-driven models of type (3.7) but also to the traditional use of a priori static restrictions for orthogonalization. Just as it was often difficult for modellers of conventional SEMs to find enough and adequately solid a priori restrictions without using certain arbitrary ones, SVAR modellers found themselves not really cleared of the situation. They had to take arbitrary positions on whether certain cross-equation error correlations should totally reflect structural interdependence rather than 'passive responses' between equation disturbances (see Hansen, 2004; also Sims, 1986), because the available a priori restrictions from conventional structural models were often insufficient for orthogonalization.

The technical refinements were accompanied by VAR proponents' ardent defence of the usefulness of the VAR approach for theory-based policy analyses, notwithstanding the Lucas critique (e.g. see Sims, 1982a, 1986, 1987, 1989; Sargent, 1981). The defence was a natural reaction to the fast-rooted belief that VARs were good forecasting models but not structural models and thus incapable of structural inferences. Various VAR applications for testing rival macroeconomic hypotheses were presented by Sims (1980b; 1983), Blanchard and Watson (1984), Litterman and Weiss (1985), Bernanke (1986), and also Leeper and Sims (1994).[19] These applications demonstrated how flexible the VAR approach could be in relating empirical findings to particular dynamic theories of interest. In the late 1980s, this line of research found a new stimulus in the arrival of cointegration theory (see Granger, 1983; Engle and Granger, 1987; and also Chapters 4 and 5 of this book). Since a cointegrated system could be naturally represented by a VAR (see Sims et al., 1990), cointegration theory rendered VARs a powerful link to long-run equilibrium-based economic theories (see also Canova (1995) for a contemporary survey of the developments).

Meanwhile, research on the forecasting front came to an impasse. In preparation for a business-cycle forecasting conference organized by the NBER in May

[19] Note that Leeper and Sims (1994) examines empirically a completely specified dynamic general equilibrium model.

1991, Sims took over a small Bayesian VAR forecasting model from Litterman[20] and tried various experiments and technical extensions on it in order to improve its forecasts. In spite of his efforts, the nine-variable VAR failed to forecast the onset of the 1990–91 US recessionary downturn (Sims 1993). The relatively strong forecasting performance of VARs over conventional macro-econometric models was fading away as those models became strengthened in dynamic specification. On the other hand, the limited variable coverage of VARs, due either to the modellers' allegiance to the principle of general equilibrium theory or to the technical curse of dimensionality, severely restricted VAR's forecasting competitiveness with respect to other multivariate data methods such as dynamic factor analysis (see Chapter 6). Interestingly, the possibility of mis-specification due to omitted variables has been highlighted repeatedly in the VAR literature (e.g. Sims, 1980b, 1989), in relation to different causality test results due to changes in the number of variables included in the VARs.[21] But empirical VAR studies have remained on a very small scale in variable coverage nonetheless.

The VAR methodology gained firm acceptance from macroeconomists in the 1990s, led by influential works such as Bernanke and Blinder (1992) and Christiano et al. (1996). Meanwhile within econometric circles, the methodology entered a consolidatory period via more technical refinements (e.g. Watson, 1994). But the overall trend was clear: the research was following closely the CC motto of building better bridges between theory and data. Moreover, the criteria for better bridges became more explicit and specific, as best illustrated by Sims's description of an 'ideal model', that is, one which 'contains a fully explicit formal behavioural interpretation of all parameters', 'connects to the data in detail', 'takes account of the range of uncertainty about the behavioral hypotheses invoked', and 'includes a believable probability model that can be used to evaluate the plausibility, given the data, of various behavioural interpretations' (Sims, 1989: 489).

3.6 Methodological reflection

The historical investigation reveals that the VAR approach arises as a methodological revision and renovation of the CC tradition.[22] Stimulated by the RE movement in macroeconomics, the VAR approach offers a systematic

[20] Litterman was the key VAR modeller at the Federal Reserve Bank in Minneapolis, where a 46-equation monthly forecasting VAR model of the USA was built during the mid-1980s. The small model had six variables and Sims extended to nine variables.

[21] The issue was highlighted later in connection with the SVAR approach by Cooley and Dwyer (1998).

[22] A very similar view was actually put forward by Eichenbaum (1985: 306), who argued that the VAR approach 'should be viewed as a necessary complement to an important class of structural models'. But his view appears to have been totally ignored.

procedure to tackle the issue of 'model choice' bypassed by the CC group. The procedure embodies a synergy of various preceding methods explored by prominent econometricians and statisticians alike, such as Tinbergen's 'kitchen work' (1937), Liu's appreciation of a general VAR (1960), Leamer's 'sinning' 'in the basement' (1978a), as well as Granger's time-series approach to causality (1969) and the Box–Jenkins (1970) time-series notion of identification. In spite of its imperfections, the procedure has never abandoned the CC tradition of theory allegiance, policy-oriented research target, and technical rigour.

Why then has the VAR approach been regarded by many as 'atheoretical macro-econometrics'? On the face of it, the 'atheoretical' charge seems to be an over-generalization of using the unrestricted VAR model as the initial step of econometric investigation. In retrospect, this initial step indeed constitutes the most drastic proposition of the VAR approach, as it advocates a reversal of the structural → reduced-form sequence, a reversal with serious methodological implications of backtracking Frisch's (1937) 'structural' method in favour of Tinbergen's (1935) 'historical' method. Moreover, the polemics employed by VAR proponents against the macro-econometric practice of the time apparently encouraged the view that they had demarcated their methodological position from the CC methodology. But that is insufficient to brand them with the 'atheoretical' badge. Since the CC methodology had been consolidated into a paradigm by the late 1960s as described in Chapter 1, any attempts to challenge it would require substantive and sometimes radical justification.

Another likely reason for the 'atheoretical' charge is the confusion of the VAR approach with the statistical properties of VAR models. While a VAR model is known for its capacity of summarizing data regularities, the VAR approach is a strategy to develop VAR-based structural models for policy purposes on the basis of such capacity. In retrospect, the confusion appears to have stemmed from a rather confined evaluation of the VAR approach within econometrics. The evaluation effectively conceals the fact that the VAR approach has emerged in close correspondence to the RE movement in macroeconomics. As mentioned before, the RE movement has affected mainstream econometrics by changing the way macro theories are formulated. Its emphasis on dynamics and micro foundations substantially complicated the simple and static form of macroeconomic models upon which traditional econometric models and methods were built. RE models have given rise to shifting interpretations of 'structural' models commonly known to econometricians (Sims, 1991). The moving 'structural' post almost makes it no longer possible for econometricians to anchor their starting position on a priori given 'true' models 'about which nothing was unknown except parameter values' (Hansen, 2004: 276). The rise of the VAR approach could be seen as a systematic response to such a loss of position. The approach shifts the focus of econometric modelling away from measuring a priori given structural parameters and puts a greater emphasis on simulating shock-instigated dynamic movements of the modelled variables. The apparent abandonment of traditionally formulated 'structural' models and the emphasis on the coverage of the dynamic features of the modelled variables by the VAR approach renders it a more data-exploratory image than what has been widely accepted as the CC structural approach. It is only from this angle that the VAR approach may be seen as 'atheoretical'.

However, if it is viewed in the broad light of the reforming macroeconomics, the VAR approach is avant-garde and at the same time starkly faithful to the CC tradition. As shown in this chapter, VAR modellers have devoted a great deal of effort and thought to mending the link between structural and reduced-form models; they have placed the issue of structural identification at the top of their research agenda; they have attached far greater importance to theory testing and policy analysis than to forecasting; they have concentrated their applications within the domain of connecting macro theories with stylized facts. But more importantly, both VAR and CC researchers share the same fundamental conviction that macro-econometric modelling should always be formulated, specified, and estimated in a system of interdependent equations. While the CC group emphasize simultaneity at a static level and see that as corresponding to a Walrasian equilibrium model, VAR modellers accentuate dynamic interdependence and see that as corresponding to a dynamic general equilibrium model, following the lead of the RE movement. Moreover, the emphasis of VAR modelling on impulse response analyses entails an unequivocal 'structural' interpretation of the error terms, an interpretation which is absent in the CC modelling framework. In this respect, the VAR approach has actually taken a stronger structuralist position than the CC group. The structural shock interpretation assumes away the possible existence of any theoretical ignorance or omission within a specified VAR, virtually granting it the status of a maintained hypothesis.[23] Indeed, it has been widely taken for granted among applied modellers that the use of VARs should make their models immune from all mis-specification errors and that the choice of variables to be included in the VARs should largely remain the device of macroeconomists. The VAR approach is thus fundamentally theory-bound and enhances, rather than forsakes, the CC paradigm.

[23] The problem that the shocks might not be uniquely interpreted as structural was pointed out by Hansen and Sargent (1991); see also Chapter 8 for a more detailed discussion on the history of the error terms.

4

Rise of the LSE Approach

The LSE approach bears the hallmark of 'dynamic specification', deriving probably from the title of Hendry et al. (1984) and also Hendry's (1995) textbook *Dynamic Econometrics*. The hallmark embodies a collective and concentrated effort to improve the CC structural approach mainly with the help of time-series statistical methods. Unlike Sims's VAR approach or Leamer's Bayesian specification search, the LSE researchers have not overtly criticized the CC approach. Instead, they have chosen an eclectic position, drawing extensively not just from the works of the CC but also from its wider historical roots, particularly the earlier works of Frisch and other forefathers. They have also kept abreast of the major works of their contemporaries and drawn lessons from them. Their eclectic position has allowed the LSE group to put together a comprehensive strategy for dynamic model choices and designs. The strategy has resolved much of the model choice issue which was left to one side by the CC group and, methodologically, it arguably goes further than the other two approaches in this respect.

The origins of the LSE approach are described in Gilbert (1986, 1989, 1991a). Essential features of the approach are discussed in Pagan (1987, 1995), where they are compared with the Bayesian and the VAR approaches. Further historical material is available from interviews with leaders of the LSE approach, see, for example, Phillips and Sargan (1985) and Ericsson and Hendry (2004), as well as from the recollections of a few key players, for example Mizon (1995), Hendry (2003), Phillips (2003), and Sargan (2003).

The present chapter attempts to portray the development of the LSE approach with the help of those background sources, so as to place the development on a wider historical spectrum. It should be emphasized that the development discussed here is focused on the methodologically reformative aspect only, which is just one narrow part of what the LSE group has contributed to econometrics. The investigation covers mainly two decades—the 1970s and the 1980s. To set the scene, antecedents of the 1970s period are described briefly in the next section. Section 4.2 moves on to advances from empirical modelling during the 1970s. The subsequent conceptual formalization of the LSE approach forms the topic of Section 4.3. Section 4.4 returns to the empirical side and examines the rise to maturity of the LSE approach through the attempts at money demand modelling during the 1980s and up to the early 1990s. Section 4.5 concludes with a brief methodological assessment.

4.1 Preludes

Prior to the rise of the LSE group, the Department of Applied Economics (DAE) at Cambridge University was probably the most prominent econometric research institution in the UK. Under the directorship of Richard Stone, the DAE group made a substantive contribution to applied structural modelling with time-series data (Gilbert, 1991a; Qin, 1993, ch. 6.1; Gilbert and Qin, 2006). To a large extent, their contribution was prompted by G. Orcutt. Orcutt started from an empirical investigation into the time-series data features of Tinbergen's (1939) macro model of the USA and found extensive residual serial correlations in the estimated equations of the model (Orcutt, 1948). He then delved into the dynamics of single variables by regressing them individually in an autoregressive (AR) model. The experiment showed that many of the variables could be characterized by an AR of order two with a unit root, implying a loss of statistical optimality in the OLS method used by Tinbergen. The finding led to his collaboration with Cochrane at Cambridge and resulted in the development of a least-squares based estimation procedure, later known as the Cochrane–Orcutt (CORC) procedure (see Cochrane and Orcutt, 1949). The CORC procedure essentially followed the CC approach—in that it took a static structural model as a priori given and fixed, and the symptom of residual autocorrelations as data 'complication'—and appended the model with a residual autoregressive scheme in order to devise a statistically efficient estimation procedure for the structural parameters in the original model. In addition to the procedure, Orcutt and Cochrane also proposed an expedient prescription for applied modellers—to re-specify the original static model into a growth-rate model, when the serial correlations were found to be close to one (see also Chapters 7 and 8).

The CORC procedure and their prescription were adopted by Stone (1954) within the DAE. They were also disseminated by Tintner's (1952) textbook and the influential demand study by Wold and Juréen (1953). Moreover, their prescription of a growth-rate model re-specification was incorporated into Theil's (1961: ch. 6.2) mis-specification analysis. Another important DAE contribution which emerged during the same period as Orcutt and Cochrane's work was the Durbin–Watson (D-W) test for residual first-order autocorrelation (see Durbin and Watson, 1950, 1951). The D-W test soon became widely adopted in the econometrics community, probably also due to the dissemination of Tintner's textbook and the book by Wold and Juréen, in addition to its within-group adoption at the DAE.

The 1960s saw the UK centre of econometric research gradually shift from the DAE to the LSE. The shift resulted primarily from a notable expansion of econometrics provision at the Department of Economics at LSE during the mid 1960s. Furthermore, econometric research at the LSE benefited from the close link between the Department of Economics and the Department of Statistics as well as the Department of Philosophy, Logic and Scientific Method at LSE. The latter was particularly well-known for housing a strong positivist tradition in philosophy of science (Gilbert, 1989; see also Hendry, 2003; Sargan, 2003). Indeed, the strong influence of the positivist philosophies of science was discernible from the LSE's econometric research. The research started primarily from Sargan's endeavour to solve systematically the problem of residual autocorrelations, which would unify the DAE's research line with the CC approach.

Sargan actually set out his research prior to his LSE appointment.[1] In a similar way to Orcutt, Sargan (1959) approached the problem of residual autocorrelations by devising statistically optimal estimation methods, only his estimators were based on an autoregressive residual amendment to a general SEM and on the more elegant IV procedure rather than the least-squares used in the CORC procedure. Further pursuit of the issue led to his 1964 Colston paper, which is retrospectively ranked as the very first masterpiece breaking the ground for the LSE approach (Gilbert, 1989 and the special issue of *Econometric Theory* 2003, 3)). However, the paper has remained relatively unheeded and its historical importance long unrecognized, as is evident from the relatively small number of citations found from the economics literature (see Chapter 10).

The belated and inadequate appreciation of Sargan's Colston paper is probably due more to its 'unusual' style than to its publication medium.[2] Its coverage was too wide for the taste of many of his contemporary econometricians, whereas its empirical examination and policy discussion was too technical for many applied economists and policy makers of the time. It also lacks the kind of provocative polemics commonly expected of methodological path-breakers, especially when it is compared with writings by proponents of the Bayesian approach and the VAR approach. It reads like a laboratory record of an extensive experiment into how to produce a rigorously built applied model that is useful for policy analysis. There is nothing spectacular, as might be expected of a

[1] Sargan moved from the University of Leeds to the LSE in 1963 (Phillips and Sargan, 1985; Gilbert, 1989).

[2] The paper is reprinted in Hendry and Wallis (1984) but the reprint has not boosted its citations significantly.

declaration of a new methodological approach. Nevertheless, its experimental style has become the epitome for numerous subsequent research outcomes which shaped the LSE approach over the next two to three decades.

Sargan started his Colston paper from the CC paradigm. He carried on the theme of his 1959 paper to elaborate the IV estimation procedure for a static model appended with a residual autoregressive scheme:

$$Ax_t' = u_t \qquad u_t = ku_{t-1} + e_t \tag{4.1}$$

where A was a matrix of parameters, x_t a vector of variables, and the autoregression of u_t could be higher than order one. Combining the two equations in (4.1) revealed that the appended residual autoregression actually re-specified the original static model into a dynamic model, that is:

$$Ax_t' - kAx_{t-1}' = e_t, \tag{4.2}$$

which was a special form of a general dynamic model:

$$Ax_t' - Bx_{t-1}' = \varepsilon_t \tag{4.3}$$

with the parametric restriction $B = kA$. The revelation led to Sargan's device of a test procedure for the parametric restriction in equation (4.2). It should be noted that Sargan's derivation effectively refuted the commonly accepted view, through textbook popularization of the CORC procedure, that the addition of a residual autoregressive scheme to a static structural model was merely a statistical generalization of the error term and the generalization was innocuous to the specification of the given structural model. Sargan's test procedure implied a methodological reverse of the CC approach of starting econometrics from the a priori fixed structural model, since it suggested that it was statistically desirable to test the model within a more general, that is parametrically unrestricted, dynamic model framework.

In fact, Sargan explored just that in the empirical part of the Colston paper, where the Klein–Ball (1959) price–wage model was his chosen guinea-pig (see Chapter 5). Starting by estimating a simplified version of the Klein–Ball model using three methods—the OLS, the CORC procedure, and his IV procedure—Sargan found that the results of the first two were similar whereas those of the third had much larger standard errors. He thus abandoned the third procedure, which was his own device, since 'there seemed little point in trying to find a better set of instrumental variables' (Sargan, 1964: 39).[3] Sargan subsequently turned his attention to an extensive dynamic specification search to improve the Klein–Ball model formulation, in a style similar to that used

[3] This must have been extremely discouraging considering that the implementation of Sargan's IV procedure required a complicated computing program which took him several years to develop (Gilbert, 1989).

by Dicks-Mireaux and Dow (1959). The search process was to shape the LSE approach in several important respects. First, the search process was mostly assisted by simple OLS estimates as the task of searching for a better model specification overrode that of choosing the best estimators with respect to particular interim specifications under trial. Secondly, the linear form of the Klein–Ball model was compared with the log-linear form of the Dicks-Mireaux and Dow model (see Section 5.2), and the relative advantages of the latter form were expounded. Thirdly, white-noise residuals became a primary criterion for specification selection, and that led to the practice of adding lagged variables directly to structural models to circumvent the likely occurrence of residual autocorrelations when the initial models were of the static type. Fourthly, the resulting dynamic models were carefully reformulated to ensure that the parameter corresponding to each regressor was economically interpretable, a step later referred to as model re-parameterization. Here, Sargan's experimentation led to separately identified short-run and long-run effects, and more significantly, an error-correction (EC) term representing long-run wage and price homogeneity (see Chapter 5). The EC specification was to become the most prominent feature of his paper two decades later, after the theory of cointegration came into being (see Section 4.3 and also Chapter 7). But at the time, Sargan did not pay much attention to his EC specification per se. He was more interested in solving the implied long-run solution as well as the lag lengths of his end model in order to study its policy implications. In fact, it was the policy aspect, instead of the min/max axiom-based theory formulated in a scholarly way, which underlay what he viewed as achieving a satisfactorily derived econometric model. Such a pragmatic attitude towards structural model building was to prevail in the LSE approach.

Nevertheless, Sargan's empirical exploration in the Colston paper attracted little immediate attention. What was immediately heeded, mainly via the research topics adopted by a number of his PhD students, was his quest to integrate the DAE's time-series research into the CC approach. It turned out that progress from their subsequent research eventually elaborated and developed many of the ideas which resided in embryo in his Colston paper (e.g. Hendry, 2003). The development brought the LSE approach into being and with it, D. F. Hendry, its chief researcher, to the leadership of a long methodological expedition into dynamic specification (Phillips and Sargan, 1985; Ericsson and Hendry, 2004).

4.2 Dynamic specification from empirical modelling

Following Sargan, Hendry began his research from the CC tradition. One of his earliest empirical studies was an eight-equation SEM of final demand of the UK (Hendry, 1974).[4] Starting from the SEM and ensuring consistent estimation of it, Hendry devoted his attention to seeking a systematic way of making a choice from several different specifications of the model. A particular choice of concern was between residual autoregressions added to the SEM and a dynamically more general SEM, that is, the situation demonstrated in equation (4.2) versus equation (4.3). The search resulted in several methodological steps which went beyond the CC structural approach. First of all, estimation became used explicitly and extensively as an intermediate step to assist model specification choice, instead of a one-off step serving to obtain the best estimates for given parameters. To facilitate the choice, Hendry produced GIVE (Generalised Instrumental Variable Estimator), a computer program which soon came to be dubbed the 'model destruction programme' at LSE because of the high rate of specification rejection it generated through comparison of the estimated results of different specifications (Ericsson and Hendry, 2004). In particular, Hendry's comparisons revealed that the addition of residual autoregressions to an SEM was merely an expedient for its mis-specification in terms of omitted lag variables and signs of such mis-specification from residual tests were rather robust irrespective of whether simultaneity was incorporated in the choice of estimators. The finding thus led to at least two innovative steps. One was to make an explicit choice to extend the dynamic structure of the original SEM rather than mend it with residual autoregressions. The other was the extensive use of data-based criteria for model selection via estimation. White-noise and minimum standard deviation residuals became a fundamental criterion, which effectively rejected the residual autoregressive specification route. Another criterion was the forecasting performance of a fitted model. To implement the criterion, a few of the end-of-sample observations were retained from estimation during the model specification search step. Interestingly, Hendry compared his specification search procedure with Box–Jenkins's (1970) time-series identification approach.

Subsequent empirical investigations soon encouraged Hendry to move away from the dichotomy of simultaneity and dynamics and focus his exploration on the dynamic side. Choosing a simple three-equation SEM of the UK import demand as his guinea pig, Hendry (1975) studied closely the consequences of three common types of mis-specifications—omitted variables, omitted simultaneity, and omitted dynamics. Aiming to obtain 'a reasonable description of the underlying data generation process' (1975: 289), he conducted an exhaustive examination through estimating various re-specifications of the SEM. The omitted dynamics turned out to have the

[4] The initial version of the paper was presented at the 1971 European Meeting of the Econometric Society.

most pronounced consequences, which reinforced his view against the use of any estimation methods to 'camouflage' the dynamics of time-series data in model specification.

In his next empirical study, Hendry embarked on a mission to seek an eclectic route between the CC approach and the Box–Jenkins time-series approach, as the latter was becoming noticeably popular in econometrics. The study, modelling UK building society behaviour, was carried out jointly with Anderson (1975; revised version published in 1977). Hendry and Anderson regarded the Box–Jenkins approach as 'sensible' but somehow inadequate in representing interdependence between economic variables, a key characteristic of economic theory. In particular, they pointed out that the commonly used method of data differencing in the Box–Jenkins approach amounted only to retaining short-run data information while discarding the long-run information, which was what economic theory was mostly concerned about. To incorporate the Box–Jenkins approach with long-run equilibrium-based economic theory, they developed a sequential model specification procedure through iterations of estimation, testing, and model re-specification. The procedure essentially exploited the long-run and short-run information jointly from the two approaches to rule out any detectable signs of mis-specifications. The ultimate aim was to obtain the most data-permissible dynamic specification of the model.

The procedure was further improved in Hendry and Mizon (1978) in the context of evaluating UK money demand models. In that paper, a static structural model appended with residual autoregression like (4.1) was labelled the 'common factor' model or COMFAC, because the resultant parameter restriction as shown in (4.2) implicitly amounted to assuming an equal lag structure of the variables (see Section 7.4). The lack of a priori realistic reasons for the assumption strengthened Hendry and Mizon's position to embrace, as a generally viable route for applied modellers, the method of starting with a general dynamic specification of a model and sequentially simplifying it with respect to data permissibility.

A fuller development and exposition of this route was presented in a joint paper by Davidson et al. (1978) on modelling aggregate UK consumption. To avoid various fallacies from ad hoc model specification impositions, Davidson et al. embraced three principles from the basic philosophy of applied modelling: (i) a new model should be able to explain results from existing models, especially competing models, and outperform them; (ii) a theoretical framework is indispensible in any applied modelling; and (iii) an acceptable empirical model should be able to account for all of the salient data features.[5] The principles were designed to help discipline applied

[5] It is interesting to find that Leamer's (1974, 1975) papers are cited in their discussion on model specification search strategies.

modellers in the pursuit of a systematic and progressive model specification search, with the aim of designing the best possible model representation of the underlying 'data generation process', a concept later referred to simply as 'DGP' (e.g. Hendry et al., 1984), and assumed a key place in the LSE approach. To a large extent, this modelling route amplified and theorized Sargan's (1964) empirical venture, as described in the previous section. Single-equation time-series methods were used, primarily to ensure the identification of an adequately dynamic model with residuals satisfying the white-noise properties at the outset. The practice was equivalent to filtering and retaining the regular part of data information as much as possible in the modelled part. Economic theory was applied loosely and related essentially to choice of variables and the implied long-run relationship. The dynamic model was then re-parameterized into an EC form to circumvent the problem of collinearity (see Section 7.3) and to verify the economic interpretability of the model. In particular, the steady-state solution implied in the EC term was checked for its long-run consistency with theories such as the life cycle hypothesis and the permanent income hypothesis. The end model was produced after careful tests and simplification, as well as by comparisons with existing models to ensure that they were outperformed by the new model.

The joint study by Davidson et al. (1978) was seen as momentous and marked the rise of the LSE approach. It popularized the EC specification in econometric modelling using time-series data and, in the UK, set 'new standards of best econometric practice' for academic and government economists as well (Gilbert, 1989: 125). Moreover, it anticipated further conceptual developments to draw the LSE approach into a more self-contained methodology.

4.3 Conceptual formalization of dynamic specification

Hendry's inaugural professorial lecture delivered at LSE in 1979 probably served as the inauguration of the conceptual formalization process (Hendry, 1980). Based on a broad perspective of the philosophy of science in the LSE tradition, Hendry prescribed 'test, test and test' as 'the three golden rules' for combating inadequately specified applied models, which he regarded as the major cause of the widespread predictive failure of models during the turbulent 1970s. The prescription also signified a decisive shift in his theoretical research from estimation methods, as highlighted by his estimator generating equation (Hendry, 1976), to more general and methodological issues concerning empirical model design.

The first milestone came with the conceptual formalization of 'exogeneity', a joint work with R. F. Engle and J.-F. Richard (Engle et al., 1983).[6] The work evolved immediately from an earlier investigation by Richard (1980) into the changing status of endogenous versus exogenous variables in models of regime shifts.[7] Richard observed that applied modellers often assigned variables as exogenous in a careless and inconsistent way, especially in the context of policy variables in models which allowed for regime shifts to study the effect of policy changes. To tighten the definition of exogeneity, he brought the concept into the statistical framework of conditioning and marginalizing likelihood functions with respect to time-series data. The framework enabled him to define 'weak exogeneity' on individual parameters of explanatory variables in a static equation, which was seen as resulting from a conditional distribution within a joint distribution. In the case of a simple regression:

$$y_t = \alpha + \beta x_t + u_t, \tag{4.4}$$

the modeller essentially assumed a reduction of the joint data density $D(x, y)$ to its conditional component $D(y_t | x_t)$ only, since (4.4) was based on:

$$E(y_t | x_t) = \alpha + \beta x_t. \tag{4.5}$$

Hence, the exogenous status of x_t entailed the validity of the conditional reduction, which required the error term, u_t, to satisfy certain statistical conditions, for example u_t being identically and independently distributed, and in turn designated parameter β to a particular function of the second moments of the two variables, for example the ratio of their covariance to the variance of x_t. Consequently, the exogeneity of x_t should be associated to β and particularly to its estimator.

Richard then extended the discussion to dynamic models and the complications that the dynamics would involve for the definition of exogeneity. The dynamic aspect was further elaborated in the joint work of Engle et al. (1983), where the concepts of 'strong exogeneity' and 'super exogeneity' were introduced, to differentiate from a situation of weak exogeneity, that is, the situation discussed above in a static model. Extending from weak exogeneity, strong exogeneity was related to the notion of the Granger causality test in

[6] Their joint paper was first released in 1980 as a CORE Discussion Paper 80–83 because Richard was based at the Centre for Operations Research and Econometrics (CORE) in Belgium. The research in statistical theory at the CORE during the 1970s was acknowledged to have contributed vitally to their joint work, see the ET interviews of Engle (Diebold, 2003) and Hendry (Ericsson and Hendry, 2004).

[7] Richard's (1980) paper was first presented at the 1977 European Econometric Society Meeting and the audience was 'bewildered' by his new perspective according to Hendry's recollection (Ericsson and Hendry, 2004).

determining sequential dependence from a relevant set of all lagged parameters, and super exogeneity concerned the capacity of the parameter estimates in a conditional model to remain invariant under regime shocks in the form of parametric shifts in the marginal equation/model of the conditioning variable, eg x_t in (4.4) within a joint distribution framework. Theoretically, this formalization of 'exogeneity' helped to dispel much conceptual confusion over causality, invariance of structural parameters, the relationship between parameters of interest, variables and model specification, as well as the desired zero correlations between explanatory variables and the error term.[8] Empirically, it encouraged tests of exogenous assumptions in addition to the appropriate choice of parameter estimators.

Around the time the 'exogeneity' project was under way, Hendry and Richard also delved into a formal conceptualization of the empirical modelling procedure using time-series data. The conceptualization was based on the same statistical approach underlying the formalization of exogeneity (Hendry and Richard, 1982).[9] Starting from the observation that available theory never gave 'a complete quantitative characterisation of the world', they designated model design as an essential task of applied econometric modellers and set about putting the design process on a 'progressive research strategy'. The strategy was built on the three principles proposed by Davidson et al. (1978), as described in Section 4.2. Here, the principles were condensed to build models which were consistent in theory, coherent in data, and parsimoniously 'encompassing' with respect to rival models. Moreover, the principles were used to organize and categorize existing model selection criteria from previously disparate studies. The categorization was done through a particular partition or taxonomy of the available information to modellers. The partition separated data information into the past, present, and future, and the non-data information into theoretical knowledge and competing models, and models were again represented by means of the statistical approach of likelihood-function based conditioning and marginalization. For example, the white-noise criterion widely imposed on model residuals was paired with the past data information and regarded as part of the data-coherent principle, while weak exogeneity was paired with the present data information and as part of the theory-consistent principle. Overall, the categorization made it clear that models designed by any single principle, for example the theory-consistent principle or the data-coherent principle, were generally inadequate. A model came to be regarded as a 'tentatively adequate conditional data characterisation' of the underlying DGP

[8] For a historical study of the concept of exogeneity, see also Aldrich (1993).

[9] The initial results were presented at a conference on model selection in 1980 and the abstract was published in *Journal of Econometrics*, 1981.

only when its design was shown to fulfil all the three principles.[10] It should be noted in particular that the notion of being 'tentatively adequate' indicated that there was possibly room for models to develop progressively as both data and theoretical information accrued (Phillips, 1988).

The formal conceptualization of model design and the taxonomy of information and criteria were applied to a systematic study of conditional dynamic models by Hendry and Richard (1983). In that joint paper, the behavioural notion of 'contingent plans' was related to the concept of weak exogeneity and used to rationalize the likelihood-function based conditioning reduction from a DGP in terms of a joint distribution of all the variables concerned. Various commonly used single-equation linear models from the applied modelling literature were summarized and classified into nine basic types, and eight out of the nine were shown to be special cases of an EC model, dynamically the most general conditional model type.[11] The typology helped to explain the respective properties of many empirical models in previously disparate applied studies and relate the properties particularly to model restrictions which were usually implicitly assumed without tests. The explanation therefore highlighted the advantages of following a 'general-to-specific' model search strategy as opposed to the traditional strategy of the 'simple-to-general' route.[12] Moreover, the discussion helped to place the use of various model diagnostic tests and the choice of parameter estimators into the appropriate steps in the model design framework.

The discussion on conditional dynamic model design and the advantages of the EC model type was extended in 'Dynamic Specification', a chapter in *Handbook of Econometrics* jointly written by Hendry, Pagan, and Sargan (1984).[13] Economic theories, the authors argued, were not effective in providing precise information about the lag structure of dynamic models, while the information that they offered most effectively was of static and conditional relations with a long-run equilibrating implication. It was thus recommended that the search of empirical models should start from dynamically most general conditional models corresponding to the theoretical relations of interest and end in the form of the EC model type to facilitate economic interpretation. The nine model types

[10] A summary of the model design principle with sorted criteria was subsequently provided in 'Editors' Introduction' of Hendry and Wallis (1984).

[11] The typology was initially discussed in an unpublished paper by the two authors in 1980.

[12] The 'general-to-specific' strategy evolved from the earlier applied modelling experience, e.g. Davidson et al. (1978) and Hendry and Mizon (1978). This summary term as opposed to the 'simple-to-general' route was explicitly discussed in Hendry (1979), where Leamer's (1974) criticism of 'excessive presimplification with inadequate diagnostic testing' was used to describe the key problem of the latter route. The term 'general-to-specific' was later used interchangeably with 'general-to-simple' (e.g. Gilbert, 1986).

[13] This paper was virtually completed in 1981, as shown from the reference list of Hendry and Richard (1982).

and their properties were further discussed in that connection. Other routes of dynamic specification were also discussed and evaluated, including the VAR approach initially proposed by Sargent and Sims (1977) and the disequilibrium model approach based on the Samuelson–Tobin model.[14] Interestingly, the former was regarded as being from a data-instigated modelling perspective but the latter from a theory-led perspective.

The conceptual framework of the LSE approach was further strengthened by various technical developments during the 1980s, for example, a unified procedure for conducting encompassing tests by Mizon (1984) and Mizon and Richard (1986). But the most influential development was probably 'cointegration' theory. This theory was initially formalized by Granger (1981, 1983) and subsequently supplemented with a single-equation estimation procedure by Engle and Granger (1987) and a simultaneous-equation estimation procedure by Johansen (1988).[15] The notion of a co-integrated time series was first discussed in Granger (1981), closely connected to the EC model type (see Section 7.4). Methodologically, the power of the notion essentially lay in its being the first provision of a statistically operational representation of the age-old idea of the equilibrating trend of an economic system or unit within which individual factors might nevertheless exhibit disequilibrating tendencies. The representation therefore furnished the EC model type with arguably the most convincing justification in terms of economics.

Meanwhile, empirical modelling utilizing the LSE approach was made widely and easily available by the public release of PcGive, a personal computer-based program evolved from GIVE (Hendry, 1986b). The software was designed on the basis of the LSE methodology and served as the workhorse of the LSE approach in practice (see Doornik and Hendry, 2007). In particular, the software facilitated general-to-specific model searches via a batch of 'testimation', a term coined by Trivedi (1984) to describe a combined use of estimators and tests in order to reach an end model which would adequately capture conditional data characterization tentatively, that is, only with respect to the available sample information.[16]

The development of the LSE approach reached its peak when Hendry's textbook, *Dynamic Econometrics*, came out in 1995. The book evolved, with years of delay, from a manuscript, which was in circulation as early as the late 1980s, of a lecture series that Hendry delivered on econometric methodology

[14] Samuelson's (1947) book *Foundations of Economic Analysis* was cited here.

[15] Granger acknowledged (see Phillips, 1997) that the cointegration idea came from conversations with Hendry during the 1975 conference where Sargent and Sims (1977) presented their joint paper on the VAR approach (see footnote 4 of Chapter 3). A more detailed account of their exchange of ideas was given in Ericsson and Hendry (2004).

[16] The testimation process was eventually automated in the software, PcGets, which was first released in 2001 (Hendry and Krolzig, 2003).

at Oxford University during the 1980s (see the preface of the book and also Ericsson and Hendry, 2004). Methodologically, the peak was largely embodied in chapter 9 of the book,[17] where the LSE approach was wrapped up in 'the theory of reduction'. Essentially an extension of the conceptual framework drawn up by Hendry and Richard (1982), the theory systematized the dynamic specification search and the model design process into twelve stages from the perspective of information reduction and allocated various commonly used tests into appropriate stages.

4.4 Dynamic specification in action: money demand studies

The formalization work provided an eloquent rationalization for the applied modelling route pioneered by LSE modellers during the 1970s, as described in Section 4.2. But the formalization alone cannot adequately account for the rise to prominence of the LSE approach from the 1980s onwards. The group also distinguished themselves by producing insightful findings through applied model building. One of the most influential cases of their applied work was a series of studies led by Hendry on building aggregate money demand models of the UK and the USA, closely linked to key policy concerns of the Bank of England at the time (Ericsson and Hendry, 2004). This section examines the evolution of this case in order to see how applied model research interacted with the formalization and what key features there were which distinguished Hendry's money demand models from previously existing models.

Almost all the previously existing models suffered extensive predictive failure in the wake of the 1973 oil crisis and the shift of major currencies from fixed to floating exchange rate regimes. While 'structural breaks' or 'regime shifts' in the economy under study were ascribed to the failure by many modellers, Hendry took economic turbulence as reflected in the data as information that was valuable for weeding out poorly designed and specified models (Hendry, 1985). In a study of the UK money demand models, Hendry (1979) effectively made predictive success a key criterion of model evaluation. The criterion imposed parameter constancy on adequately built models, an imposition which anticipated the notion of 'super exogeneity' and, at the same time, rehabilitated Frisch's description of 'structural' models (see Chapter 7). Using quarterly data, mainly from 1963 to mid-1976, Hendry designed, following the general-to-specific route, a single-equation error-correction model of UK money demand with a long-run unit elasticity of real money with respect to real income. He demonstrated that his model

[17] The chapter was based on the joint paper by Cook and Hendry (1993).

produced satisfactory predictions for the then newly released data observations up to the end of 1977, while the predictions by existing money demand models using the same variable coverage but built by the simple-to-general route were significantly off the mark. Moreover, he put those poor predictions down to either a lack of adequate dynamic specification or missing long-run adjustment factors in the models.[18] Comparison of the models established the tentative encompassing power of Hendry's model.

Hendry's involvement in money demand studies was substantially lengthened by the publication of Friedman and Schwartz's 1982 book, *Monetary Trends in the United States and the United Kingdom: Their Relation to Income, Prices, and Interest Rates, 1867–1975*. At a request from the Bank of England, Hendry conducted, jointly with Ericsson, an extensive evaluation of the econometric modelling work of the UK long-run money demand relation in Friedman and Schwartz's book.[19] Following virtually the same approach used in Hendry (1979), the evaluation exposed serious pitfalls in Friedman and Schwartz's econometric work. The pitfalls were ascribed mainly to the simple-to-general modelling methodology. Using the general-to-specific route, Hendry and Ericsson (1991a) produced a much more robust UK money demand model using annual data for 1878–1970. Among other things, their evaluation significantly undermined Friedman's monetarist position as being empirically unfounded. Their methodological contest also revealed the general weakness of many economic theorists for holding biased positions in searching for a priori preferred evidence from data without properly studying the historical information of the data at hand.

Indeed, careful examination of data and the relevant historical context formed an essential part of Hendry's money demand studies and sometimes played a vital role in helping his search for constant-parameter models which could ride out turbulent periods of volatile data. That was probably best shown by his study, jointly with Baba and Starr, of the US narrow money demand using quarterly data from 1960 to the 1980s (see Baba et al., 1992).[20] In view of the fact that the data period under investigation covered episodes known as 'missing money, great velocity decline and narrow money explosion' (Baba et al., 1992: 25), Baba, Hendry, and Starr devoted a great deal of attention to choosing financial variables as proxies of monetary innovations, such as various types of interest rates as well as a bond yield-based risk index,

[18] Methodological evaluation of existing UK money demand models was initially carried out in Hendry and Mizon (1978).

[19] The evaluation was first reported at the Bank in 1983 as a consultant paper (see the reference list in Hendry, 1985); the revised version was eventually published as Hendry and Ericsson (1991a), see Ericsson and Hendry (2004) for more detailed historical background of the event.

[20] The initial study first came out in 1985 as a working paper at the University of California at San Diego. The data coverage ended in 1984 in that working paper. The sample was extended to 1988 in the published paper.

and to embellishing the choice with financial theories. The primary aim of including these variables was to filter out volatile data shifts of those historical episodes such that it would become possible to find an economically interpretable, constant-parameter error-correction model of money demand over the entire sample period (see also Hendry and Ericsson, 1991b).

The Baba, Hendry, and Starr model demonstrates the possibility of obtaining empirical verification of conventional macro theories through exploiting knowledge of specific historical events in the form of 'vernacular' variables missing from those theories.[21] The verification was effectively secured by reformulating the structural model, in addition to a dynamic specification search, through mixing a priori given variables with data-instigated vernacular variables. Interestingly, such an active use of vernacular knowledge in variable choice and model design has not been explicitly formalized in the LSE dynamic specification procedure. The usefulness of vernacular knowledge is implied mainly in the emphasis, during empirical model design, on the importance of data information, particularly its complementary role to the limited role of economic theory. The implication is discernible from the following observation by Hendry:

> If econometrics could develop good models of economic reality, economic policy decisions could be significantly improved. Since policy requires causal links, economic theory must play a central role in model formulation, but economic theory is not the sole basis of model formulation. Economic theory is too abstract and simplified, so data and their analysis are also crucial. (Ericsson and Hendry, 2004: 759)

The possibility of designing relatively robust, constant-parameter empirical models to survive turbulent data periods, as shown by those money demand models built by Hendry and his associates, gave a timely boost to applied modellers. In particular, the models served as powerful refutation to the Lucas (1976) critique of econometric models being incapable of having parameter constancy over policy or regime shifts. However, the models also attracted scepticism and criticism. In the case of the Baba, Hendry, and Starr model, the criticism mainly came from two directions. One was the possibility of finding more than one economically interpretable, constant-parameter EC model using the same data set, and the models could be mutually none-encompassing, but with distinctly different economic implications (e.g. Boughton, 1993). The other, a more destructive criticism, was the relatively high probability that model predictive success was merely short-lived after the model was released, as demonstrated by Hess et al. (1998). Although the eventual performance failure of empirical models was already foreseen and

[21] The term 'vernacular' is adopted from Swann (2006) to denote local and specific factors not formalized in the academic economics.

covered by the formal designation of such models being no more than tentatively adequate conditional data characterizations of available data samples, demonstrations of such failures would likely deter more applied economists from adopting or appreciating the LSE approach. The eventual breakdown of the Baba, Hendry, and Starr model probably played a key role in turning Hendry's attention to a systematic study of model forecasting (e.g. Clements and Hendry, 1994).

Meanwhile, the criticisms highlighted the inherent uncertainty in empirical model building and the limitations of any formalized procedure in helping to reduce such uncertainty. The formalized dynamic specification procedure was indeed helpful in assisting Hendry to weed out inferior models, but it offered little guidance when it came to creatively extending empirical models with respect to particularities in data information. Hendry was not unaware of the situation, as shown from his reiteration: 'No sufficient conditions for validating models can exist in an empirical science; and failure to reject one of the necessary conditions does not establish that the model is valid, only that it is not demonstrably invalid' (Hendry, 1985: 75). In fact, he regarded the situation as one of endowing applied modellers with an opportunity of discovery, and designated the success of empirical model design to 'creativity' and 'luck' in his 'four golden prescriptions' of empirical modelling (Hendry, 1987). Remarkably, the designation echoes Tinbergen's (1937) analogy between empirical modelling and 'kitchen work'.[22]

4.5 Methodological reflection

The previous sections show how the LSE approach was methodologically established around the mid 1980s, although its prototype emerged during the mid to late 1970s. The development span of over a decade enabled the LSE approach to benefit from the development of other approaches of the time. It is particularly interesting to note that, among the three methodological approaches discussed respectively in Chapters 2–4, the LSE camp was the only one which paid close attention to developments of the other two, as is evident from the citations of their works. Consequently, the LSE approach shares certain noticeable similarities with the other two camps. The LSE modellers agree with Leamer's general view that model specification forms the bottleneck of econometric research; they also shared his particular

[22] In fact, Hendry has long been an admirer of 'what Jan Tinbergen called "kitchen-sink econometrics," being explicit about every step of the process' (Ericsson and Hendry, 2004: 760). Note also that the issue of empirical model discovery has been repeatedly discussed in Hendry's later studies on econometric methodology (e.g. Hendry, 2009, 2011).

view that collinearity is essentially an issue of parameter interpretation and thus its possible solutions lie in reparameterization. Their advocacy of the general-to-specific modelling route keeps clear of the pitfalls of the conventional simple-to-general route exposed by Leamer. The advocacy also agreed with that of the VAR approach on the need to specify dynamically general models such that no autocorrelation would occur in the residuals in the first place.

Nevertheless, the LSE approach has kept substantial methodological differences from the other two approaches. Apart from the obvious one of adhering to classical as opposed to Bayesian statistics, the LSE approach is less faithful to the task of merely estimating a priori given structural parameters than Leamer's Bayesian approach. On the other hand, the LSE approach is more insistent on having time-invariant and economically interpretable parameter estimates as part of adequate empirical model designs than either of the other approaches. In particular, the LSE modellers are not convinced by the VAR approach which reorients attention away from individual parameter estimation and interpretation and towards the residual-based shock simulations, even though the reorientation was largely in response to developments in macroeconomics led by the RE movement, as described in Chapter 3. For the LSE camp, residuals remain non-structural, because they are not autonomous but are derived from models and represent what modellers cannot yet explain (see Chapter 8).

From the standpoint of the CC tradition, the LSE approach is probably the most deviant of the three approaches, in spite of its eclectic veneer. Instead of focusing on the established task of finding statistically best estimates for given structural parameters, LSE modellers assume their central task to be the search for both theory-consistent and data-coherent empirical models, in which structural parameters are no longer a priori fixed but are partially data-instigated or identified in the sense of the Box–Jenkins time-series approach. Unsurprisingly, empirical models built under the LSE approach have been criticized as merely reduced-form models that are not up to the standard of structural models following the CC tradition (e.g. Sims, 1991; also Faust and Whiteman, 1997). To a certain extent, the criticisms largely reflect a difficulty with the LSE approach in accommodating its empirical model discovery strategy to advances of theoretical models, mostly in macroeconomics. The limited use of a priori theoretical information, which is confined mainly to variable choices and static, long-run equilibrium conditions between the chosen variables, encourages applied modellers to detach themselves increasingly from those theoretical models which take more elaborate forms than simple static, long-run conditions. The detachment poses a potential challenge to the CC tradition of confining econometric research to better statistical measurement of available economic theories.

On the other hand, the formalized rationale of the LSE approach in terms of distribution-theory based model reduction has caught the attention of philosophers of science as an innovative research strategy to tackle the problems of causal inferences in the non-experimental sciences. In particular, the LSE approach is regarded as a viable alternative to the traditional 'hypothetico-deductive' route in that the approach seeks to effectively exploit data and the associated background information to imply hypothetic causal relationships (Cartwright, 1988, 1995). For many economists trained in the CC tradition, however, it is not at all easy to forsake the traditional route. They are especially distrustful of the frequent use of vernacular knowledge in the LSE approach because it is difficult to verify (i) whether the properties of that part of an empirical model which is derived from vernacular knowledge would hold beyond the data sample information, and (ii) since the vernacular knowledge part often plays a decisive role in safeguarding the properties of the more conventional theory-based part of the model, it is difficult to verify in general if the theories concerned are indeed empirically confirmable. The difficulties may explain why the LSE approach has mostly been successful in building empirical models in which the conventional theories involved are relatively established, and also why it is often hard to reduce the risk of post-sample failure in spite of the confirmed theories.

The frequent use of vernacular knowledge also poses an educational difficulty for the LSE approach. The use entails a non-negligible element of tacit skills in the empirical model design process, making the approach relatively hard to impart, imitate, or master. One of the best illustrations of the difficulty is an experiment reported by Magnus and Morgan (1999). In the experiment, one postgraduate student independently carried out an empirical modelling project with a given data set by imitating Leamer's Bayesian approach, the VAR approach, and the LSE approach respectively, and the imitation turned out poorest for the last approach. The result is not surprising in view of the strong scientific image that mainstream econometrics teaching has been tirelessly advocating, an image which goes against the use of tacit skills because it lacks rigour and undermines the hope of unambiguously assessing the results of different empirical models built by different approaches. Although subsequent attempts have tried to improve the degree of objectivity in the model design process, for example, either through computer automation of the model selection procedure (Hendry and Krolzig, 2003) or through increased clarification of model evaluation criteria (Allen and Fildes, 2005), the need for ad hoc use of vernacular knowledge in good model designs has withstood replacement.

From a historical perspective, however, the indispensible use of tacit skills to tap modellers' creativity in the LSE approach may point to a compromising but practically feasible route in applied modelling between the orthodox stand of the CC paradigm and the heterodox position which maintains the importance of data discovery, for example as argued by Vining (1949), and more broadly of a historical approach, as shown in the case of modelling business cycles (see Chapter 6).

5

Case Study One—Modelling the Phillips Curve

This chapter and the next examine the history of econometrics through two case studies. The cases are chosen to illustrate the two major tasks of the application of econometrics—economic theory-based policy analysis and forecasting. The present case, that is, the econometric modelling of the unemployment–inflation/wage trade-off, has remained a lively research topic for half a century and established relatively high empirical credibility in macroeconomics with highly policy-related implications.[1] Furthermore, the subject has been intriguingly linked to several milestone works in the history of econometrics, such as Sargan's (1964) Colston paper described in the last chapter and R. Lucas's (1976) critique, and has involved econometric applications from numerous directions.[2] Skimming through the literature, one is soon lost in a labyrinth of economic and econometric issues and debates. Although the topic has been reviewed and surveyed periodically, little is available on the econometric side.[3]

This case study is particularly motivated by a number of questions. What econometric tools and routes were chosen by modellers to model the Phillips curve? How did their choices help shape the ways in which they obtained, interpreted, and theorized the empirical evidence? How did the different concerns and problems that they encountered feed back into the development of econometrics? We seek answers from several clusters of econometrically significant works during the three decades after Phillips's 1958 seminal paper. Our journey starts with the original Phillips and its early extensions (Section 5.1); we then look at wage and price models developed almost in parallel to the Phillip curve (Section 5.2) and the rise of the inverse Phillips curve led by Lucas nearly a decade later (Section 5.3); subsequent research

[1] This chapter has been adapted from Qin (2011b).

[2] For example, see Blinder (1997) and Mankiw (2001).

[3] The following is a list of reviews and surveys: Goldstein (1972), Lipsey (1978), Santomero and Seater (1978), Desai (1984), R. J. Gordon (1990; 2011), Berndt (1991: ch. 10), Cross (1995), Leeson (2000), Mankiw (2001), Sims (2008); of these, Desai (1984) is the closest to the present discussion.

trends up to the late 1980s are outlined in Section 5.4; Section 5.5 assesses the bibliographic impact of the major works examined in the first three sections and concludes with some retrospective methodological observations.

5.1 The Phillips curve

The Phillips curve is named after a single-equation empirical model built by A. W. H. Phillips (1958).[4] Based on a scatter diagram of UK annual time-series data of wages and unemployment for 1861–1957, net of the interwar period, Phillips conjectured a hyperbolic function between the growth rate of wages, w, and unemployment rate, U:

$$\left(\frac{\Delta w}{w} - a\right) = bU^{z} \tag{5.1}$$

where Δ denoted a difference, $(\Delta w / w - a)$ denoted the mean-adjusted wage inflation, and parameters a, b, and z were expected to satisfy $a > 0, b > 0, z < 0$. Equation (5.1) was transformed into a log-linear form for estimation:

$$ln\left(\frac{\Delta w}{w} - a\right) = ln(b) + z\,ln(U). \tag{5.1$'$}$$

Phillips estimated (5.1′) by a novel procedure: he reduced the first fifty-three observations of the sample into six averages to estimate b and z while choosing the value of a by graphical inspection through trial and error. Crucially, z was found to be significantly negative,[5] implying a trade-off between wage inflation and unemployment. The fitted equation was shown to give good forecasts of the subsequent subsample.

Phillips's econometric work was unorthodox if judged by the CC econometrics developed not long before (see Chapter 1). But that did not prevent P. Samuelson and R. Solow (1960, 1965) from recognizing its macroeconomic significance . Their naming of 'the Phillips curve' played an important role in popularizing Phillips's (1958) work in macroeconomics.

Meanwhile, R. Lipsey (1960) made a major effort to elaborate Phillips's econometric work. Apart from providing a theoretical explanation of the wage–unemployment trade-off, Lipsey carried out extensive statistical analysis to bring Phillips's model closer to 'standard statistical methods', especially in terms of the functional form. Based on the linear-in-parameter model

[4] For a more detailed historical account of the Phillips curve, see Humphrey (1985), Wulwick (1987), and also the contributions by Klein, Laidler, Lipsey, Yamey in Leeson (2000).

[5] Note that Phillips did not report t-values or standard errors of his parameter estimates. These were supplied by Gilbert (1976), where a detailed discussion on Phillips' estimation procedure is also given.

form, Lipsey experimented with alternative specifications to represent the nonlinear data phenomenon by various variable formations, for example taking reciprocals of the unemployment variable:

$$\frac{\Delta w}{w} = a + b\frac{1}{U} + c\frac{1}{U^2} \tag{5.2a}$$

$$\frac{\Delta w}{w} = a + b\frac{1}{U} + c\frac{1}{U^2} + d\frac{\Delta U}{U}. \tag{5.2b}$$

These specifications were fitted to data with different samples/subsamples, and the results were compared mainly by R^2. The changing rate of unemployment, $d\frac{\Delta U}{U}$, in (5.2b) was added on the grounds that the rate was normally uncorrelated with the level and thus deserved separate consideration. To verify its significance, Lipsey performed an auxiliary regression of the residuals from (5.2a) on the changing rate of unemployment (parameters with a circumflex indicate estimates):

$$\frac{\Delta w}{w} - \left(\hat{a} + \hat{b}\frac{1}{U} + \hat{c}\frac{1}{U^2}\right) = d\frac{\Delta U}{U}.$$

Note that the above treatment was in tune with the specification bias analysis by Griliches (1957) and Theil (1957), although neither work was referred to in Lipsey (1960).

Lipsey also examined the possible effect of the cost of living on wages. This was initially tested via a scatter diagram between the residuals of (5.2b) and the real wage rate,that is, the money wage rate net of inflation, $\Delta p / p$, where p stood for the consumer price index. The examination led to an augmentation of (5.2b) into a three-variable model and to further experiments with the following alternative specifications over various sample periods:

$$\frac{\Delta w}{w} = a + b\frac{1}{U} + c\frac{1}{U^2} + d\frac{\Delta U}{U} + e\frac{\Delta p}{p} \tag{5.3a}$$

$$\frac{\Delta w}{w} = a + b\frac{1}{U} + c\frac{1}{U^2} + e\frac{\Delta p}{p} \tag{5.3b}$$

$$\frac{\Delta w}{w} = a + b\frac{1}{U} + d\frac{\Delta U}{U} + e\frac{\Delta p}{p} \tag{5.3c}$$

$$\frac{\Delta w}{w} = a + b\frac{1}{U} + c\frac{1}{U^4} + d\frac{\Delta U}{U} + e\frac{\Delta p}{p}. \tag{5.3d}$$

Lipsey stated in footnotes that no evidence of residual autocorrelation was found during the experiments, but no specific tests were presented. In short, the experiments showed that inflation was significant but estimates of its parameter, e, were found to be far smaller than one, too small to warrant the postulate of relating unemployment to real wage directly, and that the parameter estimates would vary with changing samples, casting doubt on the over-time constancy of the wage and unemployment trade-off.

Formal statistical tests of the constancy through Chow tests were carried out by G. Perry (1964, 1966) when he modelled the Phillips curve using US data. Perry also applied the Durbin–Watson test for residual autocorrelation diagnosis. Perry followed Lipsey's model specification approach closely rather than that of Klein and Ball (1959) (see Section 5.2), although he cited the latter work. Similar to Lipsey, Perry experimented with various specifications of the three-variable model, and also with adding other variables, such as rates of productivity and profit rates. Following Dicks-Mireaux and Dow (1959) (see Section 5.2), Perry explored fitting the model with disaggregate data, for example for the durable-goods industry and the nondurable-goods industry separately. Perry's main finding was in favour of modelling the Phillips curve at disaggregate levels using multiple explanatory variables.

In short, the econometric side of the Phillips curve has been significantly formalized through the works of Lipsey and Perry. In particular, Lipsey's work has stimulated research toward more explicit dynamic specification (e.g. Desai, 1975), whereas Perry's work has encouraged more disaggregate and micro data studies.

5.2 Price and wage modelling

Around the time Phillips was working on his 1958 paper at LSE, Klein was heading a project to build a quarterly UK econometric model at Oxford University (see Klein et al., 1961). One by-product of the project was a paper by Klein and Ball (1959) on modelling the price and wage relationship.

The Klein–Ball price and wage model was exemplary of the CC paradigm—a four-equation SEM for wage, price, earning to wage differential, and work hours. The wage equation, key to the model, was actually a quarterly extension of the adjustment equation for the labour market in the Klein–Goldberger model (1955) explaining annual wage change mainly by the annual average unemployment, the annual average inflation, and a policy dummy F:[6]

$$\begin{aligned}\Delta w_t = \alpha_0 &+ \frac{\alpha_1}{4}\left(U_t + U_{t-1} + U_{t-2} + U_{t-3}\right)\\ &+ \frac{\alpha_2}{4}\left(\Delta p_t + \Delta p_{t-1} + \Delta p_{t-2} + \Delta p_{t-3}\right) + \alpha_3 F_t.\end{aligned} \tag{5.4}$$

[6] The original equation also includes quarterly dummies; these are omitted here for simplicity.

Note that (5.4) was defined by quarterly data, where Δ denoted annual difference, for example $\Delta w_t = w_t - w_{t-4}$. LIML was used in estimation, since p_t was endogenous (sample coverage 1948–56). The OLS estimates were also calculated, and the results were 'hardly distinguishable' from the LIML estimates (see Chapter 4). Residual autocorrelation was checked by the von Neumann ratio and the Durbin–Watson test.

Among other things, a significantly negative parameter was estimated for the unemployment variable in (5.4). The finding corroborated the Phillips curve, despite the difference between (5.1) and (5.4) in terms of variable definition, choice of explanatory variables, functional forms, sample periods, data frequency, and estimation methods. Klein and Ball actually compared their results with Phillips's (1958) paper briefly and disapproved of his nonlinear functional form. However, Klein (1967) later adopted the log-linear form in modelling wage and price.

An influential study that probably helped the wide adoption of the log-linear form was carried out by Dicks-Mireaux and Dow (1959). With UK quarterly data at hand, they postulated the following basic model between annual wage inflation and price inflation:

$$ln(w_t) - ln(w_{t\text{-}4}) = \alpha_0 + \alpha_1\left[ln(p_t) - ln(p_{t\text{-}4})\right] + \alpha_2 ln(d_t) \tag{5.5}$$

where $d>0$ denotes an index of the excess labour demand using primarily unemployment and vacancy data (see Dow and Dicks-Mireaux, 1958). The model was estimated by two methods: the OLS and the Cochrane–Orcutt estimator. The two sets of estimates were found not to differ significantly. Again, the Durbin–Watson test was used for checking residual autocorrelation.

In fact, a considerable part of Dicks-Mireaux and Dow's study was devoted to verifying the 'precise form' of model (5.5) and its robustness. They experimented with different specifications, including altering dynamic formulations via the time lags of the variables, for example using biannual differences instead of annual ones, and adding new variables such as the trade union effect. Moreover, they estimated the model with disaggregate data, for example data of sub-industry groups, in order to check the validity of the coefficient estimates of the aggregate model. They also discussed, under the issue of identification, the validity of assuming the causal direction of price $\rightarrow$ wage. Their defence for the assumption was mainly built on the observed time lag in the data formation between price and wage changes. Meanwhile, they recognized the possibility of wage having a feedback effect on price, but argued that the possibility implied a recursive system and that the second estimation method (i.e. the Cochrane–Orcutt estimator) should suffice in such a system. Dicks-Mireaux and Dow acknowledged that price could depend on import costs and other factors, and related the issue to Klein–Ball's (1959) model.

Notably, Dicks-Mireaux and Dow's discussion of identification covers the two most important epistemic aspects of the issue—simultaneity and endogeneity, their discussion on the latter including both the dynamic feedback formation and the variable coverage of a structural model. But the discussion stays away from the identification conditions formalized by the CC, since Dicks-Mireaux and Dow have not adopted the simultaneous-equations model form.

A synthesis of the two 1959 works, Klein and Ball (1959) and Dicks-Mireaux and Dow (1959), was made by Sargan (1964). As described in Chapter 4, the paper consists of two parts—a theoretical part on generalizing estimators for SEM with autocorrelated residuals and an applied part initially planned for trying his newly designed IV estimators. The Klein–Ball model (5.4) was chosen as his guinea pig. However, Sargan's attention quickly shifted to model specification search and he proposed simplifying (5.4) to

$$w_t - w_{t-1} = \alpha_0 + \alpha_1 U_t + \alpha_2 \Delta p_t + \alpha_3 F_t. \tag{5.4′}$$

He then modified and extended (5.4′) to

$$w_t - w_{t-1} = \alpha_0 + \alpha_1 U_{t-1} + \alpha_2 \left(p_{t-1} - p_{t-4}\right) + \alpha_3 \left(w - p\right)_{t-1} + \alpha_4 F_t + \alpha_5 t \tag{5.6}$$

so as to take into consideration the real wage effect ($w - p$) and a possible time trend effect, t, as well as to circumvent simultaneity by lagging the unemployment and inflation variables. Note that the real wage effect was added by reference to Dicks-Mireaux and Dow (1959). Remarkably, the way this effect was specified in (5.6) introduced an EC mechanism around an imposed wage–price long-run homogeneity (see Chapter 4). Further model specification experiments led Sargan to the final choice of a log-linear model:

$$ln\left(\frac{w_t}{w_{t-1}}\right) = \alpha_1\, ln(U)_{t-4} + \alpha_3\, ln\left(\frac{w}{p}\right)_{t-1} + \alpha_5 t + \alpha_6\, ln\left(\frac{w_{t-1}}{w_{t-2}}\right) + \alpha_7\, ln\left(\frac{w_{t-2}}{w_{t-3}}\right). \tag{5.6′}$$

Sargan then examined the dynamic properties of the wage rate via transformation of (5.6′) into a weighted moving average of past unemployment and prices. The economic implication was discussed via the long-run static solution,

$$ln\left(\frac{w}{p}\right) = -\frac{\alpha_1}{\alpha_3}\ln(U) - \frac{\alpha_5}{\alpha_3} t \tag{5.7}$$

embedded in (5.6′). As already mentioned in Chapter 4, however, Sargan's (1964) work remained relatively unheeded for well over a decade.

5.3 The inverse Phillips curve

A new wave of interest in modelling the Phillips curve emerged around 1970, anticipating the RE movement. Two aspects of the Phillips curve, at least, sustained the interest—the dynamic nature of the inflation–unemployment trade-off and the interpretability of the unemployment variable as representing the real sector demand–supply gap. A dominant figure leading the new wave is Robert Lucas (see Chapter 3).

Lucas first engaged himself in empirical studies of aggregate labour supply and demand because the topic formed 'a cornerstone of both neoclassical growth theory and short-run Keynesian-type employment theory' (Lucas and Rapping, 1969a). In this joint work with L. Rapping, a conventional simultaneous-equations model of labour demand and supply was set up and augmented by Phelps's (1968) expectations hypotheses. More precisely, adaptive expectations for price, p, and wage, w, were assumed which resulted in the labour supply equation taking a partial adjustment form (defined by employment, L, per household, H). The same form was assumed of the demand equation (defined by quality weighted employment per output, Y, where an index Q was used to represent labour quality) on the simple justification that lagged employment and output had been empirically shown to be significant in demand equations. The labour demand–supply gap defined unemployment rate, U, resulting in an inverse Phillips curve—the unemployment rate being explained by wage rate and inflation with cross-equation parameter restrictions,[7] as shown in the last equation of the following three-equation structural model:

$$
\begin{aligned}
&\underbrace{ln\left(\frac{LQ}{Y}\right)_t}_{\text{labour demand}} = \beta_{10} - \beta_{11}\, ln\left(\frac{w}{Q}\right)_t + \beta_{12}\, ln\left(\frac{LQ}{Y}\right)_{t-1} + \beta_{13}\Delta\, ln(Y)_t + u_{1t} \\
&\underbrace{ln\left(\frac{L}{H}\right)_t}_{\text{labour supply}} = \beta_{20} + \beta_{21}\, ln(w)_t - \beta_{22}\, ln(w)_{t-1} + \beta_{23}\Delta\, ln(p)_t + \beta_{24}\, ln\left(\frac{L}{H}\right)_{t-1} + u_{2t} \\
&U_t = \beta_{30} - \beta_{31}\Delta\, ln(w)_t - \beta_{31}\frac{\beta_{23}}{\beta_{21}}\Delta\, ln(p)_t + \beta_{24}U_{t-1} + u_{3t}
\end{aligned}
\tag{5.8}
$$

where u_i were error terms and where most of the coefficients had expected signs or magnitude range conditions, for example, $\beta_{31} > 0$ and $0 < \beta_{24} < 1$. Assuming that the wage was endogenous, Lucas and Rapping (1969a) estimated (5.8) by 2SLS using annual US data for 1930–65. They interpreted as corroboration of their theoretical model the relatively good fit of (5.8) and

[7] Actually, Klein (1967) makes unemployment endogenous by adding an autoregressive unemployment equation, though without expectations theory to interpret the equation. However, a much earlier precedent to the inverse Phillips curve is Fisher's 1926 work (see Fisher, 1973).

the basic confirmation of those significant coefficient estimates within their expected restrictions. In particular, the inflation variable in the unemployment equation was found significant, as normally expected. Interestingly, Lucas and Rapping mentioned in footnotes and the appendix that (5.8) was actually selected from estimations of several variants of their basic theoretical model, variants such as adding an interest rate variable, a wartime dummy, and a time trend to one of the three equations at a time.

Subsequently, Lucas and Rapping (1969b) extended their inverse Phillips curve by introducing alternative forms of the price expectations. In addition to the simple adaptive expectation scheme underlying the partial adjustment model form of the inverse Phillips curve in (5.8),[8] the RE hypothesis was postulated, which led to a general autoregressive distributed lag model of unemployment:

$$\begin{aligned} \underset{\text{rational expectations}}{ln\left(p_t^*\right)} &= \sum_{i=0} b_i \ln\left(p_{t-i}\right) + \sum_{j=1} a_j\, ln\left(p_{t-j}^*\right); \qquad \sum b_i + \sum a_j = 1 \\ \Rightarrow \quad U_t &= \beta_0 + \sum_{i=0} \beta_{1i} \Delta ln\left(w_{t-i}\right) + \sum_{i=0} \beta_{2i} \Delta ln\left(p_{t-i}\right) + \sum_{j=1} \beta_{3j} U_{t-j} + u_t. \end{aligned} \tag{5.9}$$

Annual US data for 1900–65 were used, and subsample estimates of the two alternative unemployment equations were obtained. The results rendered more support to the one in (5.9) than that in (5.8) and were interpreted in favour of the RE hypothesis. The long-run static solutions and accompanying significance test statistics (e.g. the hypothesis of $\sum \beta_{2i} = 0$) were then derived from the various subsample estimates of (5.9). The solutions suggested absence of a significant long-run inflation–unemployment trade-off. That was interpreted as endorsing the theories of a vertical long-run Phillips curve derived from the Phelps–Friedman expectations hypothesis.[9] Another major finding by Lucas and Rapping (1969b) was the lack of constancy in parameter estimates. This led to the view that empirical Phillips curves did not have much value in terms of assisting policy decisions.

Lucas's research forked, after his joint works with Rapping, in two directions that were to impinge enormously on both macroeconomics and macro-econometrics. The first direction was modelling the output–inflation trade-off, which bore close similarity to the inverse Phillips curve as unemployment was considered economically comparable to an output gap. Again, Lucas's main interest was to test the long-run implications of the RE theory, especially Friedman's natural rate hypothesis (see Lucas, 1972b, 1973). In terms of model (5.9), the natural rate was the rate at which the long-run unemployment–inflation trade-off was absent, a rate also known as the non-accelerating inflation rate of

[8] A simple adaptive expectation of price amounts to assuming: $ln\left(p_t^*\right) = \lambda ln\left(p_t\right) + \left(1 - \lambda\right) ln\left(p_t^*\right)$, where p^* denotes permanent price.

[9] The hypothesis is commonly seen as originated from Phelps (1968) and Friedman (1968).

unemployment (NAIRU).[10] The other direction was embodied by Lucas's (1976) critique of the validity of using structural econometric models for policy purposes. Notably, the Phillips curve was the theme of the Carnegie-Rochester conference volume in which the critique was published. In the critique, Lucas used, as an example, an unemployment–inflation model similar to (5.9) to show that the coefficients of inflation (β_{2i}) in the unemployment equation would not remain constant if policy shocks occurred in the form of changing parameter values in b_i or a_j of the price equation. The example became the keystone to his general argument that few econometric structural models had invariant coefficients owing to agents' RE behaviour under frequent policy shocks.

Interestingly, the econometrics employed by Lucas basically follows the 1960s textbook approach, that is, starting from a rigorously formulated theoretical model and using econometrics for the best estimates of those a priori defined structural parameters. After all, Lucas's primary motive for doing econometrics was to find empirical support for his a priori formulated theoretical models. The route of experimenting with alternative specifications with respect to sample data information, as used by Lipsey, Perry, Dicks-Mireaux and Dow, and Sargan, was formally forsaken, despite his admission that 'many coefficient estimates vary rather widely depending on which other variables are included' (Lucas and Rapping, 1969a: 747). Deprived of the data-instigated outlet, Lucas's attachment to the textbook econometrics became loosened, as he experienced more mismatches between what the textbook econometrics delivered and what he had expected to achieve out of his theoretical interest. Most of his subsequent studies simply used the OLS estimator.

The RE-instigated theories and the related empirical studies explored by Lucas gave rise to new econometric issues and controversies. The job of providing better estimation methods for RE models was tackled relatively quickly and successfully (e.g. Wallis, 1980), but the task of resolving other modelling issues turned out to be far more challenging and baffling (e.g. Pesaran, 1987). As a result, econometric practice became greatly diversified from the mid–late 1970s onwards.[11]

5.4 Diversified practice

One macroeconomist who played a pivotal role in extending Lucas's work on modelling the output–inflation trade-off was Thomas Sargent. Augmenting Fisher's theory of the real interest rate with the RE hypothesis, Sargent

[10] The literature on the natural rate hypothesis is vast; for general surveys, see Cross (1995), Ball and Mankiw (2002).

[11] In his account of the history, Gordon (2011) chooses 1975 as a demarcation year and describes the post-1975 period as a 'less well understood' period when macro theories forked in the road.

(1973b) deduced that a convenient way to test the augmented theory was via the use of the natural rate of unemployment as a proxy for the output gap. Two tests were proposed. One used Clive Granger's (1969) causality test, that is, testing whether unemployment could be significantly explained by, other than its own lags, the lagged variables that the RE hypothesis was conditioned upon. The other regressed unemployment on two separate parts of inflation—the expected and the unexpected inflation. The regression aimed at checking whether the first part had any explanatory power. The latter test was more sophisticated as it involved formulating unobserved expectation variables and circumventing possible simultaneous-equations bias. Using quarterly US data for 1952–70, Sargent obtained mixed results from the two tests. He played down the results of the second mainly because the results of the first showed more constancy over different subsample periods. In a subsequent five-equation RE model that Sargent (1976a) postulated, the test of the natural rate hypothesis became solely reliant on the Granger causality test.

Further contemplation of the connection between RE-based structural models and the time-series vector autoregression (VAR) model underlying the Granger causality test led Sargent to a new revelation: observational equivalence between the natural rate model based on Keynesian theories and the model based on classical theories (Sargent, 1976b; and Chapter 3). Here, Sargent chose to represent the theoretical/structural models in a moving average form:

$$\begin{pmatrix} y \\ z \end{pmatrix}_t = \sum_{i=0} \begin{pmatrix} A_i & B_i \\ 0 & D_i \end{pmatrix} \begin{pmatrix} \varepsilon_y \\ \varepsilon_z \end{pmatrix}_{t-i}, \tag{5.10}$$

where y could denote output and z, a policy instrument; ε_y and ε_z were assumed to be uncorrelated white-noise 'structural' shocks. Mathematical equivalence between a moving average model like (5.10) and a VAR such as

$$\begin{pmatrix} y \\ z \end{pmatrix}_t = \sum_{i=1} \Pi_i \begin{pmatrix} y \\ z \end{pmatrix}_{t-i} + \begin{pmatrix} \varepsilon_y \\ \varepsilon_z \end{pmatrix}_t; \quad \Pi = f(A, B, D) \tag{5.11}$$

led Sargent to interpret (5.11) as the 'reduced form' of (5.10). Further, Sargent showed that both Keynesian models and the classical models shared (5.11) as their reduced forms and hence might not be empirically differentiable.

A well-cited case of using (5.10) as a structural model in macroeconomics was the four-equation model of money growth and unemployment built by R. Barro (1977, 1978) (see also Barro and Rush, 1980). In Barro's model, output, price, and unemployment dynamics were assumed to be mainly driven by unanticipated money growth, which was defined as the residuals of the money growth equation:

$$\varepsilon_{mt} = \left(\frac{\Delta m}{m}\right)_t - \left[\sum_{i=1} \hat{a}_i \left(\frac{\Delta m}{m}\right)_{t-i} + \hat{b}\left(\frac{\Delta U}{U}\right)_{t-1} + \hat{c}Z_t\right], \tag{5.12}$$

where m, U, and Z denoted money, unemployment, and exogenous fiscal variables respectively; the estimated coefficients were denoted by a circumflex. Both the current and the lagged ε_m were found significant in explaining unemployment and output:

$$\begin{aligned} U_t &= \alpha_1 + \sum_{i=0} \beta_{1i}\varepsilon_{m_{t-i}} + \gamma_1 z_{1t} + \upsilon_{1t}; \quad \Rightarrow \quad \bar{U} = \alpha_1 \\ ln(y_t) &= \alpha_2 t + \sum_{i=0} \beta_{2i}\varepsilon_{m_{t-i}} + \gamma_2 z_{2t} + \upsilon_{2t}; \quad \Rightarrow \quad \Delta ln(\bar{y}) = \alpha_2 \end{aligned} \tag{5.13}$$

where t was a deterministic time trend and a constant long-run equilibrium rate, or 'natural rate', α_1, was assumed. Among other things, Barro's model stimulated much interest in testing the relationship between unanticipated monetary shocks and the natural rate hypothesis, that is, whether it was the unanticipated shocks alone that would drive output to deviate from its 'natural rate'.

With respect to econometrics, models such as (5.13) raised two representation issues, albeit little heeded by macroeconomists, namely, (i) justifying that the theoretical entities of unanticipated shocks, such as monetary shocks and real supply shocks, were equivalent to model-derived residuals, and (ii) justifying the representation of the anticipated long-run movement by a constant rate. Econometric efforts to resolve the issues led to a renewed interest in latent-variable models (e.g. Geweke and Singleton, 1981), and in the NBER business cycle research tradition of decomposing the permanent and transitory components of variables by their time-series properties (see Chapter 6).[12] Research along these lines helped foster, during the 1990s, the revival of factor models and the use of time-series filters to define latent theoretical entities, such as time-varying NAIRU.

Apart from those measurement issues, econometricians were also confronted with the demand for better or sharper tests to differentiate competing theoretical models. Various attempts emerged. For example, Pesaran (1982) used Cox's non-nested testing procedure to evaluate Barro's model results against the Keynesian alternative, and Ilmakunnas and Tsurumi (1985) and Leamer (1986) applied Bayesian methods to evaluate the output–inflation

[12] Later, a similar time-series approach was extended to multiple series and applied to the study of the long-run output–inflation trade-off (e.g. Geweke, 1986; and also King and Watson, 1994).

trade-off and the unemployment–inflation trade-off. Unfortunately, statistical uncertainty of the empirical results was repeatedly found to be too large to sustain a clear verdict between rival theories despite the use of refined tools. To a large extent, the evidence reinforced Sargent's 'observation equivalence'.

An alternative modelling route to circumvent 'observation equivalence' was to drop the theorists' stance of 'pretending to have too much a priori economic theory', a route explored by Sargent and Sims (1977) that evolved into the VAR approach (see Chapter 3). Applied macroeconomists were particularly attracted to the VAR approach because it facilitated impulse response analysis through model simulation, as it made shock-based models such as (5.10) empirically operational, and also by its continued allegiance to the general equilibrium tradition. But one fundamental problem cropped up: how should modellers causally sequence the contemporaneous shocks when these terms were correlated with each other? In his (1980a) paper Sims simply followed the inverse Phillips curve in ordering the triangle shock matrix of his six-variable VAR model, that is, letting the contemporaneous wage and price shocks precede that of unemployment. However, it was soon shown by Gordon et al. (1982) that the reverse ordering in accordance with the Keynesian school could work equally well. They also highlighted another problem of the VARs—the results would often vary considerably when the VARs were altered in terms of which variables were included.

The inclusiveness of macroeconomic evidence motivated some empirical researchers to go for micro evidence from disaggregate data, leading to a boom in labour economics (e.g., Oswald, 1985; Pencavel, 1985). Meanwhile, there came a rising interest in time-series methods. Apart from the VAR approach, there emerged numerous studies on the compatibility between the properties of observed single time series and the corresponding time-series process a priori postulated in models based on the RE hypothesis, such as the autoregressive scheme of the monetary instrument implied in (5.10) and the autoregressive distributed lag structure for inflation in (5.9). These studies revealed the wide existence of non-stationary features in economic variables. For example, Altonji and Ashenfelter (1980) showed, through various tests including the then newly developed Dickey–Fuller unit-root test, that aggregate wage rates exhibited significant random walk properties. Such findings severely undermined those RE models that disregarded nonstationarity and pointed to transitory shocks as the only source of dynamics.

The rise of time-series econometrics culminated in the mid-1980s with the birth of cointegration theory (see Sections 4.3 and 7.4). A vital spur to cointegration theory was the empirical success of ECM, since the EC term in such models was frequently made up of cointegrated non-stationary variables (see Phillips, 1997). Interest in Sargan's (1964) work was finally revived (e.g. Dawson, 1981; Hendry and Wallis, 1984; Hendry, 2003). In fact, Granger and Weiss (1983) used the Sargan-type wage equation as the first empirical example of an error-correction model, and Sargan's (1964) paper was cited again in Engle and Granger (1987). But cointegration and the associated error-correction model were not found immediately helpful in resolving issues that perplexed macroeconomists, such as whether the impact of monetary policy was transitory or permanent on the inflation–unemployment trade-off and how to measure accurately the NAIRU if the impact was more than short-run (e.g. Mankiw, 2001; Ball and Mankiw, 2002).

For many macroeconomists, the time-series econometric approach was too data-driven. A more theory-driven approach was explored by Kydland and Prescott (e.g. see their 1982 paper) and grew into a methodological enterprise known as the 'real business cycle' (RBC) model. RBC models were built on economic theories such that the policy issues of interest were fully specified by model construction. The models were then calibrated through simulations, aiming mainly at generating macroeconomic data that would possess the key time-series features observed from actual data. However, it was beyond the RBC approach to ascertain whether such key features would or should remain invariant, especially when the models were used to study the effect of various policy shocks.

5.5 Impact assessment through citation analysis

To improve the assessment of the impact of the historically important works discussed in sections 5.1–5.3, a citation database was constructed. The database was based on twenty-six relevant major works during the three decades starting from Phillips's (1958) paper.[13] Over 4000 citations were collected from JSTOR (for the pre-1970 period) and the Web of Science (for 1970–2005). Every entry in the database was classified in line with the Journal of Economic Literature (JEL) system. It should be noted that the JEL system often classifies one work into several categories. For example, Lucas and Rapping (1969b) has the categories of 'C' (econometrics), 'E' (macroeconomics), and 'J' (labour economics). Entries under 'C' were further classified into three types: 'applied', 'theoretical', and 'educational'.

To study the citation patterns of the key papers discussed in Sections 5.1–5.3, these papers were assigned to three base groups: group I (Phillips, 1958; Lipsey, 1960; Perry, 1964, 1966), group II (Dicks-Mireaux and Dow, 1959;

[13] The twenty-six root works are mostly from the reference list of this paper. A few citations by papers in books and conference collections are added, but the database is primarily made of journal papers. Citations of a non-research nature such as book reviews are filtered out.

Klein and Ball, 1959; Sargan, 1964), and group III (Lucas and Rapping, 1969a, 1969b; Lucas, 1972b, 1973). The entropy-based *s*-index of Silagadze (2009) was used here and adapted to measure the citation impact of each base group. Denoting the citation count of base group *i* for year *t* by $\Gamma_{i,t}$, the document number of the base group by n_i, the initial year of the citation data by y_0, and the final year by y_T, the adapted *s*-index was defined as[14]

$$s_i = \frac{1}{2}\sqrt{\frac{\sum_{t=y_0}^{y_T}\Gamma_{i,t}}{n_i}\frac{\varepsilon_i}{ln(y_T - y_0)}};\varepsilon_i = -\sum_t p_{i,t}\,ln(p_{i,t})\text{and } p_{i,t} = \frac{\Gamma_{i,t}}{\sum_{t=y_0}^{y_T}\Gamma_{i,t}}. \qquad (5.14)$$

Since all the citations were classified, disaggregate *s*-indices of particular JEL categories could also be calculated. A simpler measure, which follows the index of topic transfer (ITT) by Mann et al. (2006) to reflect the degree of topic diffusion, was also calculated:

$$ITT_{i,t}(L) = \frac{\text{Number of citations of group } i \text{ in JEL category } L \text{ by time } t}{\text{Number of citations of group } i \text{ by time } t}. \qquad (5.15)$$

The index measures the degree of impact of base group *i* in research field *L* relative to its overall impact. Here, we are particularly interested in $ITT_{i,t}(C)$, $ITT_{i,t}(E)$, and $ITT_{i,t}(J)$.

The total citation counts of the three base groups are reported in Table 5.1 and their time series are plotted in the upper left panel of Figure 5.1. Table 5.1 also gives the *s*-indices calculated by various data arrangements. Since the three groups emerge from different years, a set of subsample *s*-indices was also calculated for the same period 1986–2005. It is easily seen that group III enjoys the largest level of citations and the highest impact measures in spite of its having the shortest span of citation years. Note also that the JEL category 'E' is where group III has its highest impact and 'J' its lowest impact, whereas group II enjoys its highest impact within 'C'. In comparison, group I has its impact more evenly distributed across the three categories (i.e. its three disaggregate *s*-indices vary least). The pattern of these *s*-indices is further reflected in the ITT series plotted in Figure 5.1. Interestingly, group III experienced a transitory period of rising attention in econometrics during the 1970s, but macroeconomists remained its main audience. Group II retained its audience mainly among econometricians, an audience that was gradually increasing owing crucially to the cointegration implication

[14] Notice, n_i is not in the original index, which is designed for measuring the impact of one individual or one work. Note also that we take the natural logarithm in the entropy calculation.

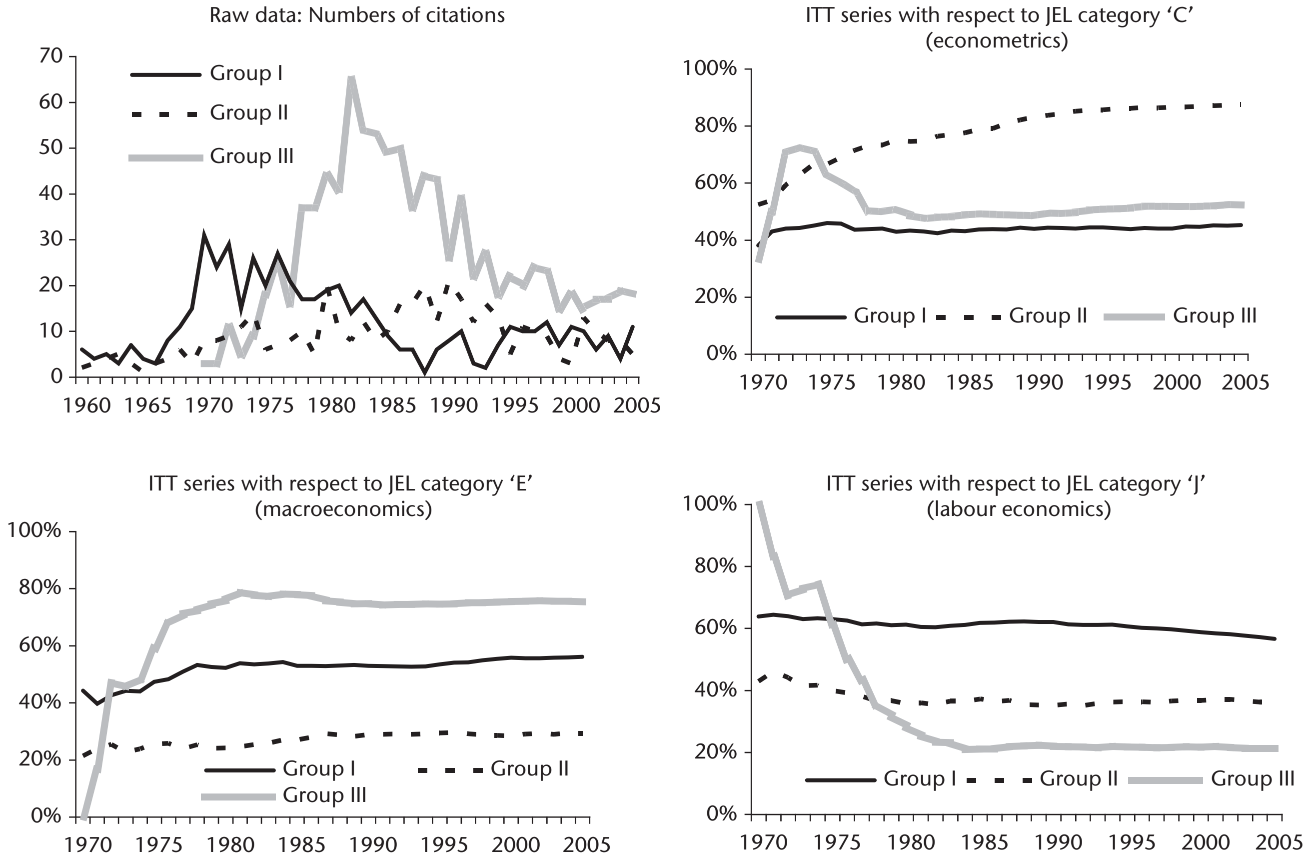

Figure 5.1. Citation series and ITT series

Note: See the note in Table 5.1 for the group information.

Table 5.1 Citation counts and *s*-indices of citation impact

	Group I	Group II	Group III
Total citation counts (1960–2005)	537	417	984
Aggregate *s*-index (initial sample year)	5.40 (1967)	5.67 (1965)	7.63 (1974)
s-index for JEL category 'C'	3.63	5.34	5.51
s-index for JEL category 'E'	4.04	3.05	6.66
s-index for JEL category 'J'	4.03	3.35	3.41
Aggregate *s*-index for subsample 1986–2005	3.01	4.29	5.60
s-index for JEL category 'C'	2.09	4.20	4.17
s-index for JEL category 'E'	2.38	2.34	4.80
s-index for JEL category 'J'	1.96	2.51	2.56
Citation counts by Journals (1960–2005)			
Econometrica	7	20	19
Journal of the American Statistical Association	2	7	1
Econometric Theory	0	6	1
Journal of Econometrics	0	5	3
Journal of the Royal Statistical Society	2	5	0
Share of the total citations	1.7%	10.3%	2.4%

Note: Group I: (Phillips, 1958; Lipsey, 1960; Perry, 1964, 1966); Group II: (Dicks-Mireaux and Dow, 1959; Klein and Ball, 1959; Sargan, 1964); Group III: (Lucas and Rapping, 1969a, 1969b; Lucas, 1972b, 1973).

of Sargan's (1964) paper. In comparison, the impact of group I was the most diffused with respect to the three categories.

Figure 5.2 plots the citations under the 'theoretical' category within 'C'. Citation counts of five econometrics/statistics journals and their shares in the total counts are given in Table 5.1. These statistics show that group I hardly appealed to theoretical econometricians directly, that group III's success with them was transitory in the early 1980s, and that only group II had maintained certain visibility over the thirty-five-year span, again owing solely to Sargan's (1964) paper.[15] Theoretical econometric research appears not to have cared much about the Phillips curve modelling.

What about the other side of the interaction, that is, how much were modellers of the Phillips curve attracted to theoretical econometric works? To find answers, we selected a sample of 125 papers from the 'applied' category under C. We checked the reference list of each of the 125 papers and picked out those items that fall under the 'theoretical' category within 'C', for example, the Granger (1969) causality test paper. The selection was made primarily from the reference lists of the literature surveys cited in the present chapter. The sample was then enlarged by including, from the database, applied

[15] It is, however, difficult to assess the secondary impact since the present database does not present citation trees.

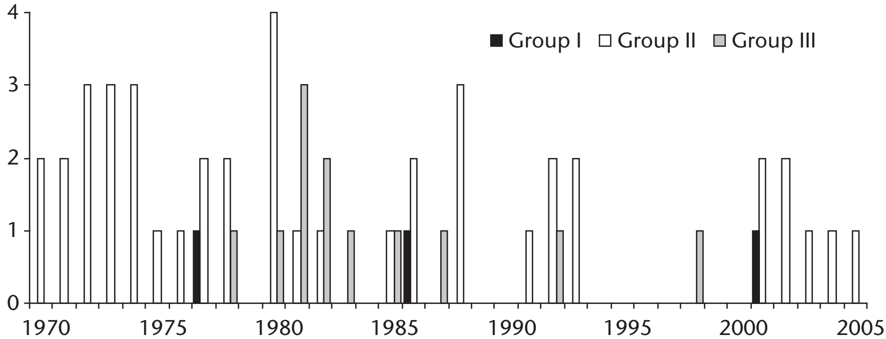

Figure 5.2. Citations by theoretical econometrics papers
Note: Self-citations are removed from the counts.

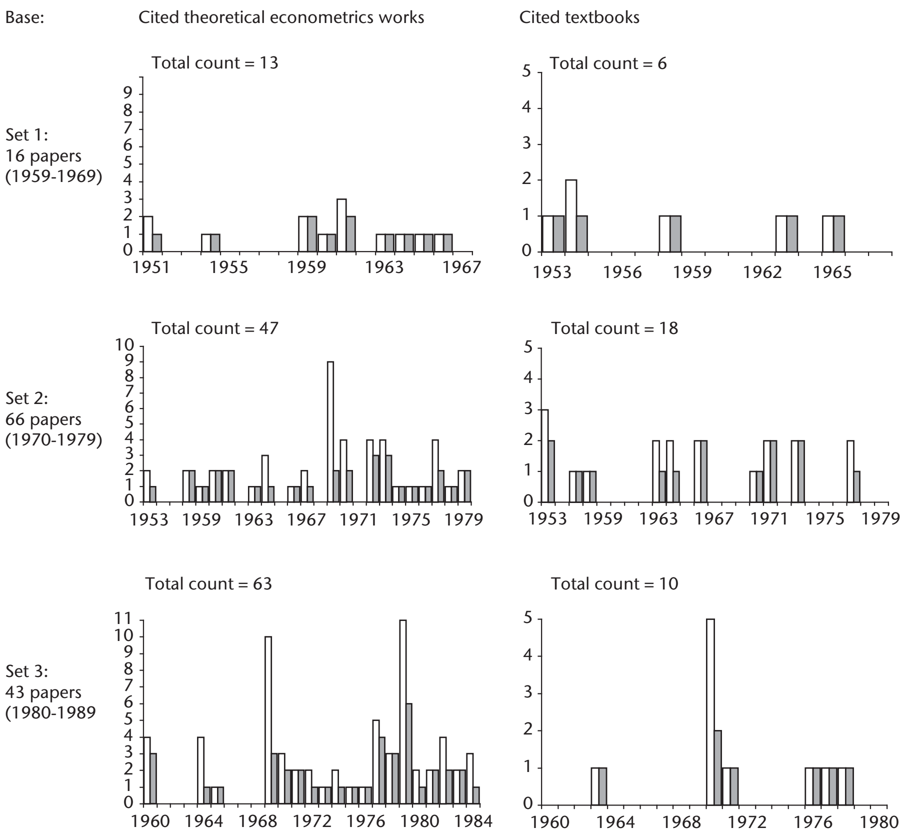

Figure 5.3. References to econometrics works of a sample of 125 applied papers
Note: Self-citations are excluded. Clear bars indicate the raw counts and the grey bars are the counts net of the works which are cited by more than one paper. For example, the two counts of 1951 (grey bar) in the top left panel are citations of a single paper (clear bar) by Durbin on serial correlation tests published in *Biometrika*.

papers published in *Economica* and the *Journal of Political Economy*, two key journals on the subject.[16]

Figure 5.3 graphs the summary statistics of the citation counts, which were further divided into three sequential sets. The graphs show that the reference counts increased over time. An inspection of the reference lists reveals that the references are relatively up-to-date and are mainly on tests, for example, the Chow test, the Ramsey test, and autocorrelation tests as well as exogeneity tests; the last is most noticeable from the middle and bottom left panels, where Granger's (1969) paper tops the counts; in contrast, references on estimators are few and far between. One might infer that many modellers would refer to textbooks on estimation matters. Indeed we see a steady reliance on textbooks from the right-side panels, although there are signs of weakening and more lagged reliance in the 1980s set. Interestingly, the top count in that set is the Box and Jenkins (1976) time-series book. On the whole, the sample evidence suggests that applied economists have been fairly knowledgeable and receptive of econometrics and have become increasingly so since the 1970s.

5.6 Retrospective assessment

In view of the history, the relatively strong citation impact of group III, especially its transitory impact rise in econometrics during the 1970s, and the belatedly increasing citation impact of group II reflect the consolidation process of the CC paradigm and the subsequent reforms that it has initiated (see Chapter 1). As described in Section 5.2, Klein was avant-garde in applying the simultaneous-equations model approach to the price–wage relationship in the late 1950s. It took roughly a decade for that model approach to be widely adopted among econometric practitioners, as shown from Goldstein's (1972) survey and also the joint study by Lucas and Rapping (1969a). Emulations of Lucas's works in group III represent an extensive acceptance of the CC paradigm in that ad hoc practice of data-instigated model specification was formally banished to 'sins' in the basement (Leamer, 1978a). Mainstream economists have been trained to use econometrics merely as a measurement toolbox to help them postulate more sophisticated theoretical models.

Methodologically, the CC paradigm meets the need of theory corroboration from mainstream economists. This explains why group II has relatively lower and slower dissemination rates. For most economists, the inflation–unemployment trade-off bears far more economic significance than the

[16] The detailed sample list is too long to attach here, but it is available from the author on request, along with the list of cited econometric works.

wage–price relationship (e.g. Gordon, 1990); the Phillips curve is particularly attractive because of its simple and heuristic model form, its close policy relevance, and its rich macroeconomic interpretability. Technical aspects associated with its empirical verification are merely secondary. Once the inflation–unemployment relationship is brought into alignment with the inflation–output trade-off, the Phillips curve becomes well grounded in the macroeconomic tradition of having a simple but closed model representation within the general equilibrium paradigm. The RE movement led by Lucas was aimed essentially at making the dynamic aspect of that model a priori complete. Econometrics became useful only for providing measured proofs of that model. The CC stance of focusing econometric tasks on best measuring a priori well-specified structural models was thus further consolidated.

Empirical RE models of the Phillips curve have, however, failed to yield more conclusive results than before, especially concerning the dynamic characteristics of the curve. Here it is interesting to see how soon Lucas abandoned more elegant estimators than the OLS advocated by the CC econometrics, replicating Sargan's (1964) abandonment of his own IV method in favour of the OLS. But neither Lucas nor Sargan was the first to experience the disappointing performance of supposedly more consistent estimators than the OLS. The first collective rehabilitation of the OLS actually occurred around 1960 (see Waugh, 1961; Gilbert and Qin, 2006). What is remarkable is the lapse of time and reoccurrence it takes for such disappointments to accrue to a major criticism of a methodology, and, in the meantime, it nonetheless becomes more widely established. It is also interesting to see how Sargan and Lucas reacted differently to their disappointments. Sargan turned to a painstaking data-instigated model specification search, which laid the foundation for the LSE approach; Lucas lashed out at macroeconomics for a lack of rigorously testable theories and at the use of macro-econometric models because of that lack, which led to the RE movement. Their different choices probably reflected the different econometric background that each came from. Regardless of that, neither Sargan nor Lucas intended to forsake the CC paradigm.

Nevertheless, the CC paradigm is losing its dominance, as seen from the diversified econometric practice since the late 1970s. Notice that much of the diversification stems from the different angles that modellers take in interpreting their empirical results. While those strongly theory-minded largely abandon econometrics for the simulation-based RBC approach, economists who still practise econometrics are also divided according to how much they are willing to relinquish the structural model approach. Some let go of the constancy of structural parameters for time-varying parameter models; others let go of structurally parametric models for random shock models or for dynamic factor models; and still others let go of the general equilibrium tradition for data-instigated single-equation models with loose theoretical guidance. Applied economists on the whole have become increasingly willing to forsake textbook teachings and let data speak more, although it is not yet a prevailing position to forgo the general equilibrium tradition and embrace empirical models explicitly with partial and incomplete structural interpretation. Active searching for empirically robust specifications is no longer considered a sin, although the results of theorization are still widely regarded as superior to and more interesting than those of an empirical nature. The CC paradigm still remains influential on a broad scale.

On the other hand, the diversification also embodies an 'externalization' process in that econometricians' attention has been increasingly shifted from devising measurement instruments for parameters given in an a priori formulated model to devising other tools for testing, evaluating, and revising the model (Gilbert and Qin, 2007). The externalization breaks loose from the essential measurement role of the CC tradition. After all, econometrics as an applied toolbox can find its lasting usefulness only in serving applied economic research.

6

Case Study Two—Modelling Business Cycles

Business cycle studies occupy a prominent position in the history of econometrics. To a large extent, modern macroeconomics and econometrics arose from business cycle studies of the 1930s in the wake of the Great Depression, see Morgan (1990: part I). Econometric business cycle research has evolved a great deal during the past seven decades. Nevertheless, macro-econometric models still fell considerably short of predicting the onset of the 2008 recession in the West. That failure has triggered the present study. This chapter examines how econometric methods for business cycle analysis evolved during the three decades from 1960 to 1990, approximately, especially the post-1973 oil-crisis period, since that was a period during which most of the methods used currently for macroeconomic forecasting were developed. We are especially interested in finding out how business cycle research was affected and led to by the reformative ideas and approaches previously discussed in Chapters 2–4, and what methodological lessons we could draw from the impressive technical advances of the modelling toolbox on the one hand and the shaky practical achievements on the other.

There have been numerous surveys of business cycle research since the end of World War II, for example Gordon (1949), Koopmans (1949b), Roose (1952), Hickman (1972), Zarnowitz (1985, 1992), Laidler (1992), and Jacobs (1998). But none of these are exclusively from the angle of the history of econometrics.

The present study will start from a brief description of the historical background in Section 6.1. Section 6.2 describes how the RE movement in macroeconomics in the wake of the global economic recession triggered by the 1973 oil crisis revitalized econometric business cycle research. Section 6.3 turns to the waves of formalization of the NBER business cycle measures as a result of the rise of time-series econometrics in the 1980s. The consequent re-orientation of macro-econometric modelling research towards business cycle forecasting is discussed in Section 6.4. The final section offers a brief historical assessment.

6.1 Background and preludes

Tinbergen's macrodynamic models, especially his model of the US economy (1939), are widely acknowledged as the first major econometric endeavour to model business cycles. Subsequent methodological debates over Tinbergen's models have played a vital role in catalysing the formalization of econometrics, as mentioned in Chapter 1 (see also Qin, 1993). A methodological summary of econometric modelling of business cycles of the time was provided by Koopmans (1949b) and the methodology basically followed Frisch's structural approach (1937). The backbone of the methodology was the Slutsky–Frisch impulse-propagation scheme, see Frisch (1933), Slutsky (1937), also Bjerkholt (2007b) and Louçã (2007: ch. 2), which assumed that business cycles were embedded in the dynamics of certain macro aggregate variables, such as GDP, and that the dynamics was driven by a few such variables according to available economic theories plus some random shocks. Under the methodology, the task of econometricians was to obtain statistically best estimates for the structural parameters of those dynamic models postulated by theorists. Explanation of business cycles was achieved once the best fit was found.

Noticeably, the above approach sidestepped certain statistically fundamental issues concerning the identification of business cycles and the measurement of the extent of their impact on various economic activities/sectors. These issues actually formed the very agenda of business cycle studies at the NBER. Starting from the early 1920s under the leadership of Wesley C. Mitchell, the NBER business cycle programme had, by the mid 1940s, developed a relatively mature procedure for establishing an empirical chronology of business cycles (see Burns and Mitchell, 1946).[1] Based on the definition of business cycles being cyclical movements in aggregate economic activities with the key features that the movements were recurrent but non-periodic in terms of timing, duration, and amplitude,[2] the chronology comprised mainly of measures of: (i) aggregate cycles; (ii) the turning points, lengths, troughs, and peaks of the cycles; (iii) the extent of cyclical effect. The aggregate time series, GDP or GNP, was a most commonly used indicator from which an aggregate measure of cycles was built, but it was also common

[1] In this classical work, specific cyclical analysis was carried out on 1277 individual time series of monthly, quarterly, or annual frequencies with various sample lengths for four countries, France, Germany, the UK, and the USA. The main method of composing leading indicators for business cycles followed their earlier joint work (Mitchell and Burns, 1938).

[2] The highly quoted NBER definition is: 'Business cycles are a type of fluctuation found in the aggregate activity of nations that organize their work mainly in business enterprises: a cycle consists of expansions occurring at about the same time in many economic activities, followed by similarly general recessions, contractions, and revivals which merge into the expansion phase of the next cycles; this sequence of changes is recurrent but not periodic; in duration business cycles vary from more than one year to ten or twelve years; they are not divisible into shorter cycles of similar character with amplitudes approximating their own' (Burns and Mitchell, 1946: 3).

practice to use the measure of 'reference cycles', that is certain averaging of a group of 'specific cycles', each derived from the seasonally-adjusted time series of a particular economic activity, such as coke production (Burns and Mitchell, 1946: ch. 2). Possible erratic movements were also filtered out from the series. The cycles were characterized via dating of their turning points, troughs, and peaks. Since a large number of series were analysed, diffusion indices were constructed as an indicator for the extensiveness of the cycles. The index was based on the proportions of upturn/expanding or downturn/contracting points at each observation of all the series. The phase difference of specific cycles was also analysed to identify series of 'leads and lags' for forecasting purposes (Burns and Mitchell, 1946: ch. 4).

The NBER research method was criticized as 'measurement without theory' by Koopmans (1947). The criticism was responded to by Vining (1949), who attacked the CC approach as being too narrow to allow for any discovery or hypothesis seeking. Their debate set a methodological divide between the CC approach and the NBER approach. The former became subsequently accepted as the paradigm of econometric research, as described in Chapter 1. In business cycle modelling, the paradigm consolidation could be felt clearly at least from two aspects. One was the adoption of the Tinbergen and the CC type of structural models as the means for modelling business cycles, even at the NBER, for example see Chow and Moore (1972), thanks possibly to the growing influence of the macro econometric modelling activities led by Klein during the 1960s. The other was a shift of focal attention of macro-econometric modelling research away from cyclic movements to dynamic convergence towards equilibriums defined essentially in terms of comparative static economic theories. A good illustration of this was probably the continuous-time econometric modelling approach pioneered by A. B. Bergstrom (e.g. Bergstrom and Wymer 1976).

Meanwhile, however, the Burns–Mitchell route of business cycle studies was by no means abandoned and chronometric studies of cycles were still carried on at the NBER. But a historically more significant event was a major research project set up by O. Morgenstern under the Econometric Research Programme of Princeton University in the late 1950s. The research comprised mainly of an empirical study of the international propagation of business cycles through financial markets following the NBER route. The results came out in the NBER Book Series Studies in Business Cycles (Morgenstern 1959). In the book, Morgenstern analysed a large number of financial time series of mainly monthly frequency from France, Germany, the UK, and the USA for the periods of the gold standard era (1870–1914) and the interwar period (1925–1938). Particular attention was paid to the co-movement (covariation) of cross-border financial series as well as to the cyclical movements between the financial series and the reference business cycles of each of the four countries. Data evidence was also used to verify theories such as interest rate parity. The findings revealed a considerable gap between data and available theories. Morgenstern thus concluded that methodological advance was needed for both theoretical and econometric research. In particular, he argued that theories should shift away from notions of 'equilibrium' and 'stability' to games and strategies between market players while 'more penetrating mathematico-statistical analysis of data may produce surprises' (Morgenstern 1959: ch. 11).

The more penetrating approach on Morgenstern's mind was probably spectral analysis.[3] Based on his empirical business cycle research experience, Morgenstern (1961) was convinced that the future of research lay with Wald's (1936) decomposition of economic time series into trend, cycles, seasonal, and irregular fluctuations rather than the Frisch–Slutsky impulse-propagation scheme. A key figure he brought into his research team was C. W. J. Granger. Their initial study on weekly New York stock price series by means of cross-spectral analysis revealed that the 'business cycle' component was insignificant in the price series and that their role in indicating/leading macro business cycles was weak. The result cast doubt on the existence of stock market 'specific cycles' derived by the NBER method (Granger and Morgenstern, 1963). However, when cross-spectral analysis was applied to a number of NBER business-cycle indicators, the identified cyclical components were found to confirm broadly those derived by the NBER method, although the duration of the average lead or lag was significantly longer than that estimated by the NBER method (see Granger and Hatanaka, 1964: ch. 12).

Interestingly, the exploratory time-series research of the Princeton Programme was criticized by Wold (1967) as 'empiricism without theory' for the main reason that the nonparametric approach of spectral techniques was ill-suited to the parametric tradition of structural econometric modelling. In the case of business cycles, it was obviously difficult to equate cycles identified by spectral techniques with what economists reckoned and described as business cycles in reality. But the criticism was soon shown to be unwarranted by Granger's (1969) introduction of a causality test, via cross-spectral methods, on the basis of the feedback mechanism of a bivariate VAR model. Ironically, Granger's approach was noted to be essentially identical to Wold's causal chain modelling approach (Sims, 1972b and Chapter 3). The test has generated enormous interest in the econometric circle and marked a new era in business cycle research—a rapid fusion of time-series methods into the structural econometric modelling approach (e.g. Granger and Newbold, 1977b).

[3] It is recorded in a number of historical studies that von Neumann suggested the spectral method to Morgenstern (see Cargill, 1974; Phillips, 1997).

6.2 Theory led time-series reforms

As mentioned in the previous section, mainstream econometrics based on the CC tradition assumed that business cycles were adequately embedded in the dynamics of a few aggregate macro variables. Accordingly, examination of the applicability of macro-econometric models to business cycle analysis was mainly conducted through dynamic model simulations. A seminal work on such simulations was Adelman and Adelman (1959). In 1969, a large scale examination was organized at a conference at Harvard University, sponsored partly by the NBER. The dynamic properties of several macro-econometric models were tested including the Wharton model and the Brookings model (Hickman 1972).[4] Most of these models were built in the spirit of the Slutsky–Frisch scheme. Interestingly, the source of cycles emerged as a contending issue through various simulation results. Purely random/erratic shocks were found unable to produce cycles; they would only arise from either shocks of assumed autocorrelated error terms or perturbations of exogenous variables. But the inevitable risk of model mis-specification, especially when models were assumed to have autocorrelated error terms (see Chapter 8), made it difficult to rule out the possibility that the source should have been structurally internal. In other words, correctly postulated theoretical models should be dynamically cyclical.

Indeed, more theoretical models containing the property of self-sustaining cycles have been postulated since the mid 1960s. One type of model, which gained rapid prominence, postulated that cycles arose from the expectation-augmented disequilibrium in short-run wage–price dynamics (see Chapter 5). The research was led by M. Friedman and E. S. Phelps and extended by R. E. Lucas. The subsequent rise of the RE movement in the early 1970s effectively moved the focal point of macroeconomic modelling from comparative static equilibrium to dynamics, especially short-run dynamics and its transitory properties as compared to long-run equilibrium solutions. In that respect, the lack of dynamically adequate structural models was blamed for poor econometric model performance in forecasting the oil-shock-induced business cycles of the 1970s (e.g. Lucas and Sargent, 1978). In response, econometric business cycle research evolved along two diverging methodological strands—one with reduced reliance on available a priori structural models and the other with attention turning away from econometric estimation towards computer simulations of shock-based models.

The first strand evolved around the VAR modelling approach. As described in Chapter 3, the approach was initiated by Sargent and Sims (1977) under the proposition to do 'business cycle modelling without pretending to have

[4] For a more detailed description of the history of these models, see Bodkin et al. (1991: part II).

too much a priori theory'. With the intention to reform the mainstream econometric approach by adapting the 'NBER style quantitative business cycle analysis', they first examined the NBER method of identifying the 'reference cycle' by reformulating the method into what they referred to as 'unobservable-index models'. They chose fourteen time-series variables, all detrended quarterly aggregates over the 1949–71 period,[5] and extracted, using factor analysis, one common factor from the set as well as from different subsets of the variables. The factor was regarded as the 'reference cycle' indicator of the chosen variable set. They then showed that the extracted factor was generally inadequate in representing the co-movement of a chosen variable set, and concluded from the finding that the NBER 'reference cycle' measure for business cycles was generally inadequate. The conclusion led them to the 'observable-index model' approach, that is the mainstream econometric approach in modelling key macro variables. There, their innovation was to start from a general dynamic model, known as a VAR, instead of an a priori tightly parameterized structural model. In particular, they built a five-variable VAR to capture US business cycles.[6] To locate the sources of cyclical movements, they resorted to a Granger causality test for identifying cross-variable sequential (lead and lag) dependence. To evaluate the magnitude of random shock impact, they performed impulse analysis to simulate short-run dynamics caused by 'structural' shocks (see Chapter 8). The two techniques were soon to become the pillar of the VAR approach, with the latter being particularly popular among macroeconomists.[7]

The second strand grew from the real business cycle (RBC) approach, which was initiated by Kydland and Prescott (1982). Kydland and Prescott opposed the monetary school's attribution of monetary disturbances as the source of business cycles and postulated a model in which the source came from technological shocks (i.e. a 'real' factor rather than a nominal factor). In their model, business cycle features were assumed to be embodied in the autocorrelation of real output in terms of GDP and its covariance with other aggregates such as total consumption and fixed investment. Simulation of cyclical features formed the primary goal of their modelling activities. Methodologically, they chose to build their model within the general equilibrium system and calibrate the structural parameters following the 'computable general equilibrium' (CGE) modelling approach.[8] Different from extant CGE models, their

[5] They also examined some monthly series when such observations were available.

[6] The variables are money, unemployment rate, price and wage indices, and a demand–pressure proxy by unfilled orders for durable goods/total shipments.

[7] Further examples include Sims' exploratory studies in monetary business cycles (1980b, 1983).

[8] The general argument for calibration was the unidentifiability of structural parameters, especially when structural models become more disaggregated. See Mitra-Kahn (2008) for more on the history of CGE models.

model was focused on finding a feasible dynamic and stochastic propagation channel to explain business cycles, thus extending the CGE approach to a new branch—the dynamic stochastic general equilibrium (DSGE) models. The use of econometrics was minimized to simple time-series statistics of the aggregates concerned and they served largely as references for adjusting and evaluating model simulation results. For example, the sample standard deviation of the real output was used in the Kydland–Prescott model to anchor the magnitude of their simulated real output.

Subsequently, econometrics became relegated to producing time-series properties of aggregate variables, properties which were used by DSGE modellers as references to target the simulated results of their conjectured RBC models as similar to what was generally observed from data. For example, Long and Plosser (1983) postulated a multi-sector RBC model which enabled the economic norm plus stochastic behaviour of producers and consumers to generate business cycles by sector-specific shocks. The time-series features of significantly autocorrelated output and strong co-movement between outputs of various sectors formed their target of model simulation—to mimic simple time-series properties of the outputs of six sectors including agriculture, manufacturing, and service. Their model was extended to include money and banking by King and Plosser (1984) to account for the phenomenon of significant co-movement between money, inflation, and real economic activities. The phenomenon was presented by both static and dynamic regressions between the aggregate output growth and growth rates of monetary and nominal variables.

The DSGE approach has carried forward and formalized the NBER tradition of emphasizing the role of sector-specific shocks in business cycle research, but at the expense of replacing econometric estimation by calibration and consequently nullifying the relevant econometric criteria for model testing. But the approach has not really repudiated econometrics in spite of the contentious position that Kydland and Prescott (1991) took in denouncing the CC structural approach as ill-suited for DSGE modelling of business cycles, and of the consequent acrimony between the DSGE camp and business cycle econometrician modellers.[9] Econometrics has proved useful at least in two respects. One involves using parameter estimates from extant econometric studies, especially micro and sector studies, as references for parameter calibration. It is in that sense that the calibration step can be seen as a kind of estimation (see Gregory and Smith, 1990). The other is to use the time-series features of aggregate economic variables from econometric studies to help assess the empirical performance of DSGE models, for example how well

[9] For more discussion about the acrimony, see Quah (1995) and the papers following Quah's summary in the 'Controversies Section'.

the model-simulated variables could match these features. The assessment can also be formalized into a statistical test procedure (e.g. Watson, 1993). The latter aspect has prompted positive feedback for the rising popularity of time-series econometric research.

6.3 Time-series formalization of business cycle measures

The 1980s saw a rapid formalization of NBER business cycle measures by time-series econometrics. One of the leading topics was related to the nonstationary feature of economic variables, especially those exhibiting significant trends. It was an old and well-established view that trend and cycle were two separable components in economic time series. Although a trend component was not filtered out in the original Burns–Mitchell procedure of dating specific cycles, they were not unaware of the desirability of filtering out secular trends before identifying the cyclical component and attributed the reason for not doing so to resource constraints at the time (Burns and Mitchell 1946). Moreover, in his earlier works, Mitchell had actually already used detrended business activity indices in dating US business cycles for the pre-1927 era, as shown by Romer (1994).

An explicit trend filter was introduced at the NBER by Mintz (1969) when she was engaged in a project of dating German business cycles by the Burns–Mitchell procedure. Mintz found many of the German time-series indices were highly trended and experimented mainly with two ways of detrending those indices. The first was to define a long-run trend by 75-month (6–7 years) centred moving averages of the indices. The cycles produced from such detrended indices were defined as ‘deviation cycles’. The second way was simply using (monthly) growth-rate indices as the base of extracting cyclic measures. The resulting cycles were defined as ‘step cycles’. Mintz found that it was harder and required more complicated criteria to extract step cycles because of ‘highly jagged’ growth-rate series and the unfeasibility of delimiting ‘cycle phases’ directly by the peaks and troughs in the data. The German business cycle index that Mintz chose eventually was based on deviation cycles alone. Subsequently, cyclical measures built on undetrended level data series came to be called ‘classical cycles’, as against ‘growth cycles’ which were derived from detrended data series.[10]

Mintz’s work demonstrated the intimate dependence of business-cycle dating measures on trend decomposition methods. But the latter remained ad hoc until the notion of nonstationarity was introduced as the statistical base

[10] Mintz (1969) quoted a remark by R. A. Gordon at a London conference in 1967, which argued for examining business cycles around the growth rate of output and employment and called such cycles ‘growth cycles’.

for filtering trends by Beveridge and Nelson (1981). Essentially, the Beveridge–Nelson trend filter assumed nonstationarity for all of the economic variables to be used for dating business cycles. Since (weak) nonstationary (technically known as 'integrated') stochastic processes could be decomposed into a stochastic nonstationary trend and a stationary component, Beveridge and Nelson proposed using the former as the trend filter and dating business cycles from the latter part alone. To justify their proposal, Beveridge and Nelson related their decomposition to Friedman's (1957) classic work in dissecting income into permanent and transitory parts, albeit their decomposition did not involve any economic principles. Technically, the Beveridge–Nelson filter was defined upon a particular univariate $I(1)$ (integrated of order one) time-series model known as ARIMA (autoregressive integrated moving average) model. For instance, a simple random walk with drift $I(1)$ series, y_t, has an ARIMA representation of its first difference, $\Delta y_t = y_t - y_{t-1}$:

$$\begin{aligned} y_t &= \mu + y_{t-1} + \varepsilon_t \\ \Delta y_t &= \mu + \varepsilon_t. \end{aligned} \tag{6.1}$$

However, different assumptions for the time-series properties would result in different filters.[11] For instance, Hodrick and Prescott (1981) chose to filter the trend by the Whittaker–Henderson method used in actuarial science, which effectively extended the assumed nonstationarity of y_t to an $I(2)$ series.[12] Even under the same assumed degree of integration, filters could vary with different assumptions on the source of the random drift in the trend. For example, starting from the conventional decomposition of y_t into a trend, a cycle, and an irregular component:

$$y_t = \tilde{y}_t + \psi_t + \varepsilon_t. \tag{6.2}$$

Harvey (1985) assumed $I(1)$ for the trend component, $\tilde{y}_t$:

$$\tilde{y}_t = \mu + \tilde{y}_{t-1} + \eta_t. \tag{6.3}$$

Substituting (6.3) into (6.2) and taking the first difference would result in:

$$\Delta y_t = \Delta\psi_t + \mu + \eta_t + \Delta\varepsilon_t. \tag{6.4}$$

This model differs clearly from the lower equation in (6.1) unless $\Delta\psi_t + \eta_t = \varepsilon_{t-1}$, that is, when there is only one single stochastic factor as the source of shocks for both the trend and cyclical components. Harvey referred to (6.2) as the structural model and the ARIMA specification as its reduced form, although

[11] More general ARIMA models result from more complicated formulations of the upper equation of (6.1). For example, model $y_t = \mu + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \lambda_0 \varepsilon_t + \lambda_1 \varepsilon_{t-1}$ becomes an ARIMA(1,1,1) when the characteristic function of the autoregressive part of y_t contains a unit root.

[12] Note, however, that Leser (1961) had actually proposed the same filter following Whittaker two decades before Hodrick and Prescott's work. I thank Stephen Pollock for this historical fact.

(6.2) bore little resemblance to the kind of structural models commonly referred to by the CC paradigm (see Chapter 7 for further description of the notion of 'structure'). Nevertheless, Harvey's discussion highlighted the need for additional information concerning the random source of the trend component once the component was assumed to be stochastic rather than deterministic (e.g. Stock and Watson, 1988). Within the univariate context, such information had to be assumed, as none of the components of a single time series were directly observable. Various assumptions led to various trend filters. The lack of consensus laid bare the information inadequacy of identifying a unique stochastic trend from univariate series. The impasse was brought to light by the fast rise to fashionable status of 'cointegration' analysis in the late 1980s, as described in Section 4.3. Among other things, the notion of cointegration offered an endogenous explanation for a nonstationary trend in the sense that such a trend in individual variables could and should be explained by their co-trending with other trended nonstationary variables. If the stochastic trend component of a variable were endogenously generated from a group of variables, the time-series approach to detrending single variables would be meaningless.

Although the issue of how best to detrend variables exhibiting nonstationary features remained unsettled, the discussion encouraged many applied modellers to work with growth-rate data as a convenient way to avert nonstationarity. The practice was commonly seen in VAR models and was adopted in those formalized techniques of identifying turning points of business cycles (see below). Mintz's differentiation between 'deviation cycle' and 'step cycle' was buried under 'growth cycle' and lost. There was somehow a confusion that 'step cycle' was shorthand for 'deviation cycle'.[13]

In the NBER dating method, location of specific cycles was the prerequisite of identifying turning points as these were selected from the peaks and troughs of specific cycles. The selection involved certain 'censoring rules', such as mid-run duration and large enough amplitudes, factors which were essentially underpinned by economic judgement.[14] The aggregate turning points could be derived from the mode of the specific turning points (e.g. Mintz, 1969) or from the aggregate reference cycle (e.g. Bry and Boschan, 1971). Comparison between disaggregate turning points and the aggregate ones formed an important step in the NBER dating method. It not only enabled the classification of specific series into leading, coincident, and lagging indicators to facilitate ex-ante forecasting, but also helped determine, via the ex-post forecasting performance of the indicators, whether the aggregate

[13] Klein and Moore (1985: Introduction) credit Mintz's (1969) work as the major methodological turning point from classical cycles to growth cycles.

[14] See also Harding and Pagan (2002) for a summary of the NBER's selection method.

turning points thus produced could be verified as the appropriate measures. Failure of the latter could evoke revisions of the aggregate turning points, which actually made the dating procedure an iterative one (e.g. see Klein and Moore, 1985: 7–8).

The NBER procedure was, however, regarded as lacking statistical rigour in terms of probability specification, and the lack was held responsible for impeding the design of statistical models for forecasting turning points (e.g. Wecker, 1979). Attempts to formalize the NBER procedure by time-series statistics were centred on automating the selection process from binary series of peaks and troughs. For example, Neftci (1982) proposed using discrete-state Markov processes in single time-series models as the specification to characterize stochastic turning points and illustrated how such models could be used to forecast macroeconomic variables, such as unemployment.[15] A significant extension of Neftci's route was carried out by Hamilton (1989). Taking as a correct fact the Beveridge–Nelson finding of the widespread nonstationary feature in level variables, Hamilton chose to apply the Markov-process specification to the first difference of an $I(1)$ series, such as y_t in (6.1), that is, treating its growth rate as a nonlinear stationary process.[16] A simple two-state extension of the lower equation in (6.1) would be:

$$\Delta y_t = \mu_{S_t} + \varepsilon_t \qquad S_t = \begin{cases} 1 & t = 1, \cdots, t_0 \\ 2 & t = t_{0+1}, \cdots \end{cases} \qquad Pr\left(S_t = j \middle| S_{t-1} = i\right) = p_{ij}. \tag{6.5}$$

The model effectively identified business cycles within the short-run growth movement of y_t by defining the cyclical turning points as distinct shifts associated with very small probabilities in the time-varying parameter, μ_{S_t}. Hamilton applied a version of (6.5), in which an autoregressive ε_t was assumed, to model the quarterly series of the postwar US real GNP growth rate and found recurrent shifting points, which were shown to conform largely to the NBER-dated turning points of recessions. Hamilton's device gained great popularity among applied modellers as its application to single macro variables yielded numerous shifts, which were handily interpreted as evidence of regime shifts or structural breaks.

It should be noted that the interpretation has strengthened the gap between the time-series notion of 'structure', such as the time-series decomposition in equation (6.3), and the traditional econometric concept of a structural model, which is crucially dependent on multivariate interdependence. It should be also noted that applied studies of dating business-cycle turning points using

[15] Neftci (1984) also tried the same specification for identifying asymmetry in single macro series, as asymmetry was believed to be an important feature of business cycles.

[16] The Markov-switching regression model was first introduced into econometrics by Goldfeld and Quandt (1973).

the Hamilton method have effectively retraced the 'step cycle' route abandoned by Mintz two decades earlier. However, the model method has now glossed over the fundamental problem associated with that route, namely that growth rate data, especially those derived from higher than annual frequency time series, have essentially lost most of the mid-range frequency information upon which business cycle measures were originally defined.

One problem arising with those newly invented time-series methods was how they should be assessed as adequate substitutes of the traditional NBER measures. Inventors of those methods would usually promote their new inventions by illustrating how their methods could generate measures in rough conformance with the NBER business cycle chronology. But by doing so, the NBER chronology was taken as the only and correct business cycle measure, and thus the NBER methodology was assumed to be the only correct one. A tougher method of assessment required proof of those inventions being capable of producing measures which would outperform the NBER chronology in forecasting the dynamic movements of key macro variables.

6.4 Forecasting business cycles with time-series modelling

One common criterion in using macro variables to define the start of a recession is a decline in real GNP/GDP for two consecutive quarters. The Neftci–Hamilton method was devised to provide estimates which would forecast such events with a relatively high success rate. Empirical evidence was, however, inconclusive on whether the method could indeed significantly outperform simple autoregressive time-series models in forecasting significant and prolonged GNP downturns (e.g. Goodwin, 1995). Various other routes to elaborate simple time-series models were explored, for example, models with an autoregressive scheme augmented by leading indicators and explicitly specified Bayesian loss functions for forecasting values (e.g. Zellner et al., 1990). Despite the efforts, ex-ante forecasts of recessionary downturns remained disappointing.

To some modellers, single time-series models were clearly incapable of capturing the information of interdependence between economic variables. Once it came to forecasting based on multivariate time-series models, the VAR approach presented an obvious route for experimenters. One of the pioneering experiments was carried out at the Federal Reserve Bank in Minneapolis, where a 46-equation forecasting model of the US was built using monthly data and the time-varying parameter and Bayesian VAR (BVAR) technique developed by Doan et al. (1984) (see also Chapter 2). While maintaining that model, Litterman, the key modeller, set up a six-variable quarterly BVAR model mainly for research purposes (see Litterman, 1986b).

He later expanded the model to nine variables in an attempt to improve its inflation forecasts.[17] The model was subsequently taken over by Sims. In an attempt to rectify its forecasting deterioration, Sims (1993) extended the BVAR technique by adding various probabilistic assumptions to make the model as unrestrictive as possible.[18] In particular, he introduced nonstationary mean priors to trended time series and drastically relaxed the classical conditions usually imposed on the error terms to allow for them being conditionally heteroscedastic and non-normally distributed. Unfortunately, his attempt failed to pay off when it came to forecasting the onset of the 1990–91 recessionary downturn in the GNP growth rates. The forecasts tracked closely behind the data series, virtually the same outcome as Chow and Moore (1972) had obtained using the conventional modelling approach almost two decades before. Unfortunately, ex-post model forecasting tests showed no warning signs.[19]

Meanwhile, a more exploratory route of multivariate forecasting was explored by Stock and Watson. They resumed the experiment, abandoned by Sargent and Sims (1977), of using factor analysis to reformulate the NBER 'reference cycle' measure with the key purpose of obtaining probability-model based forecasts of recessions. Stock and Watson (1989) started from filtering, by a dynamic factor model (DFM), a single coincident index from the variable lists used by the NBER for its coincident indicator.[20] To circumvent the problem of possible nonstationary trends, they took the first difference of those trended series just as Hamilton did. For example, a simple DFM would be:

$$\begin{aligned} \Delta X_t &= \beta_0 + \beta_1 \xi_t + U_t \\ \xi_t &= \alpha_0 + \alpha_1 \xi_{t-1} + e_t \end{aligned} \tag{6.6}$$

where ΔX_t denoted a set of detrended or stationary series and U_t an idiosyncratic component. The common factor, ξ_t, estimated by means of the Kalman filter, was regarded as representing the co-movement of ΔX_t and hence called 'the coincident index'. Next, a small set of leading indicators/variables was selected to form a VAR, which also included ξ_t in order to predict its future values as well as the associated probabilities. The predicted ξ_{t+j} was referred to as the 'leading index' and used to forecast the GNP cycles. Stock and Watson tried their model on US monthly data. A six-month ahead VAR forecast series of ξ_{t+j} was shown to track well the

[17] The original six variables are real GNP, the GNP price deflator, real business fixed investment, the 3-month Treasury bill rate, the unemployment rate, and the money supply; the added three variables are exchange rate, SP 500 stock price index, and commodity price index.

[18] The paper was presented at an NBER conference in May of 1991.

[19] Note, the lack of predictive power of this kind could not be identified by those commonly used forecasting tests based on the averaging of modelling errors, e.g. Fair (1984).

[20] Four variables are used in this case: the growth rates of industrial production, personal income, employment, and sales from manufacturing and trade sectors.

real GNP at its business cycle frequencies. However, the model missed the downturn when it was used to forecast the US 1990–91 recession. A thorough re-examination of the model led Stock and Watson (1993) to the conclusion that it was mainly the inadequate choice of specific leading indicators, rather than their modelling method, which caused the misprediction. Their finding highlighted the importance of identifying in a timely manner from the economy of concern specific data with background information which would be mainly responsible in triggering nonperiodic business cycles.

In fact, Watson was already aware of the importance of this element. The statistical nature of shocks formed the subject of one of his earlier empirical studies, which was undertaken jointly with Blanchard (1986). Their study traced the source of American business cycles to a mixture of large and small shocks, rather than purely small shocks as portrayed by the Slutsky—Frisch impulse-propagation scheme. Moreover, the shocks were found to have stemmed in equal likelihood from the fiscal and monetary side as from real-sector demand and supply side. The finding probably played a key role in motivating Watson to explore the DFM route in his subsequent collaboration with Stock. But the lack of expected forecasting success of their 1989 experiment kept many modellers in doubt of the adequacy of the DFM device mainly because of its lack of economic theory, especially when judged by the CC paradigm. While time-series modellers continued to elaborate various statistical devices, for example, to combine DFMs with switching-regime models and to experiment with probit models to turn the measures of forecasting turning points into probability ones, more conventionally trained modellers endeavoured to build dynamically robust structural models which would survive regime shifts. The most prominent models there were the error-correction type, often with embedded long-run cointegrating relations. These models were meant to accommodate the postulate that recessionary turning points also indicated shifts in the long-run trend of co-trending variables in addition to transitory swings in individual short-run variables. Still, more theoretically minded modellers pursued the DSGE approach in the belief that more accurate forecasts should result from larger scale DSGE models because they offered clearly defined causal rationales of how shocks from various sectors would propagate through systems built on established economic theories. The 1990s became an era of diverse research pursuits in business cycle modelling. A pronounced common feature of these pursuits was a significant improvement in model building in terms of internal consistency, technical complexity, and a reduction in the ad hoc judgements involved. Nevertheless, timely forecasts of the onset of recessions appeared to remain tenaciously difficult to obtain.

6.5 Retrospective assessment

There have been over four decades of econometric research on business cycles, a fact which exhibits a significant shift away from the macro-econometric tradition of Tinbergen's models and the Slutsky–Frisch impulse-propagation scheme, the tradition upon which the CC structural approach was elaborated. Modellers' attention has shifted not only from SEMs to dynamic models, but also from estimating given structural parameters to simulating shock effects and devising statistical measures to characterise cyclic phenomena. Moreover, the choice of modelling route has become much wider, with the tightly theory-based DSGE route at one end and the heavily data-based DFM route at the other. The traditional macro-econometric modelling route has long lost its dominance.

A closer reflection on the history, however, reveals an opposite facet—a methodological assimilation of the NBER tradition by the CC paradigm, with the assimilation being catalysed by time-series statistical methods. As shown in the previous sections, statistical formalization of NBER's ad hoc measures and procedures has, to a great extent, led the econometric business-cycle research of the past decades. The formalization essentially aims at a scientification of those measures and procedures in the sense that they should be built on the probability foundation, following the Haavelmo–CC enterprise, with maximum internal rigour and minimum use of outside-model ad hoc human judgements. Furthermore, the formalization process is reminiscent of the consolidation process of the CC paradigm, as described in Chapter 1, in that econometric business-cycle studies have become focused on more detailed, segmented, and narrowed-down issues, such as whether cyclic measures should be based on trended or detrended data, whether cycles should be asymmetric with respect to their upturns and downturns, and whether the shocks supposedly triggering business cycles should be small or large, purely erratic or autocorrelated, or be originated from the real sector or the monetary sector. The extensive and synthetic style of the Burns–Mitchell tradition has long been forsaken.

The formalization has undeniably improved the scientific strength of business cycle measures when compared to those produced during the Burns–Mitchell era. The rift between the CC and the NBER groups has also long been forgotten. Econometric research has extended its field from adding empirical content to a priori postulated business-cycle models to exploring data features and devising new representative measures for business cycles. Interestingly, the extension has been so significant that a new rift has occurred between econometric modellers of business cycles and the DSGE camp, a rift echoing the 'measurement without theory' controversy, only with econometricians on the accused side this time.

But the significance of the formalization becomes more difficult to identify when it is assessed from the applied perspective, especially when the success rate in ex-ante forecasts of recessions is used as a key criterion. The fact that the onset of the 2008 financial-crisis-triggered recession was predicted by only a few 'Wise Owls'[21] (Bean, 2010) while missed by regular forecasters armed with various models serves us as the latest warning that the efficiency of the formalization might be far from optimal. Remarkably, not only has the performance of time-series data-driven econometric models been off the track this time, so has that of the whole bunch of theory-rich macro dynamic models developed in the wake of the RE movement, which derived its fame mainly from exploiting the forecast failures of the macro-econometric models of the mid-1970s recession. Whereas it remains a deeply held conviction among business-cycle modellers that rigorously designed methods and models are bound to produce better results than those arrived at by ad hoc means, it has been repeatedly shown that human judgement plays an indispensible role in the making of business forecasts even with the aid of models, for example see Turner (1990) and Clement (1995), and that a simple pool of forecasts often outperforms individual forecasts based on particular modelling methods, for example see Stock and Watson (1999). These observations indicate that inaccuracy or inadequacy is practically inevitable in individual modelling work, that econometric methods are limited by their statistical approach in analysing and forecasting business cycles, and moreover, that the explanatory power of generalized and established theoretical relationships is highly limited when applied to particular economies during particular periods alone, that is, if none of the local and institution-specific factors are taken into serious consideration.

Actually, the limits have been critically pointed out periodically but somehow ignored by the core research community. One early major critique by Morgenstern (1928) even predates the Slutsky–Frisch scheme.[22] During the postwar period, severe doubt about both the CC approach and the NBER approach was expressed by Wright in a sweeping statement, 'I simply do not believe that any set of econometric models, or any set of mathematical formulae, will ever suffice for reliable economic forecasting over any great length of time. The element of novel social conception is always breaking in' (Wright 1951: 147). Shortly before that, Gordon (1949) grouped the two approaches under the name heading of the 'statistical approach', as opposed to the 'historical approach', which placed its focus on explaining particular cycles using all kinds of relevant information, and argued for a blend of the

[21] They include N. Roubini and W. R. White, who independently foresaw the coming of the crisis based on no-model-assisted human judgement.

[22] For a critical summary of the book, see Marget (1929).

two, a 'quantitative-historical' approach, as the promising direction of future research (see also Roose, 1952). After years of statistical research in business cycles, Burns, a founding father of the NBER approach, acknowledged 'that it is vital, both theoretically and practically, to recognize the changes in economic organization and the episodic and random factors that make each business cycle a unique configuration of events. Subtle understanding of economic change comes from a knowledge of history and large affairs, not from statistics or their processing alone' (Burns 1969: 85). These messages were reiterated twenty years later by Zarnowitz: 'because business cycles are not all alike and are subject to historical changes along with the structure and institutions of the economy, it is not surprising that the numerous efforts to model them as a uniform by-product of one type of random shock or another have failed' (Zarnowitz 1992: 17).

It is precisely this aspect of the unique social-historical conditions of different business cycles that has been neglected in the formalization. Reformation of the NBER methods by the CC methodology has certainly led the 'statistical approach' further away from the 'historical approach'. The wide conviction of the superiority of the methods of the science has converted the econometric community largely to a group of fundamentalist guards of mathematical rigour. It is often the case that mathematical rigour is held as the dominant goal and the criterion for research topic choice as well as research evaluation, so much so that the relevance of the research to business cycles is reduced to empirical illustrations. To that extent, probabilistic formalization has trapped econometric business cycle research in the pursuit of means at the expense of ends. It is thus not surprising that the resulting studies would fail to generate any significant breakthrough in predicting and explaining business cycles.

On the other hand, the apparently insurmountable goal of forecasting business cycles and the growing research interest in designing better forecasting models and methods calls into question one fundamental tenet of the CC approach—forecasting is an easier and lower-ranking task than policy analysis since the latter entails elaborately designed structural models while the former requires only simple reduced-form models. Evidence from our two case studies shows us the opposite. Progress has been relatively easier when the econometricians' main task has been largely to furnish a priori postulated models with numbers derived from data, than when the task becomes tracking the data well enough to predict forthcoming events of significant change in the real world. The latter compels econometricians to the job of model discovery which has been assumed to be the realm of theoretical economists under the CC approach. Such a division of labour is simply impossible when econometricians themselves directly face up to the task of exploring analysis of data.

Nevertheless, the history of science tells us that major paradigm shifts will not occur until all possible routes within the existing paradigm have been trodden. Once the formalization attempts have gone significantly astray from what is needed for analysing and forecasting the multi-faceted characteristics of business cycles, the research community should hopefully make appropriate ‘error corrections’ of its overestimation of the power of a priori postulated models as well as its underestimation of the importance of the historical approach, or the ‘art’ dimension of business cycle research.

7

Evolving Polysemy of Structural Parameters

The estimation of structural parameters forms the core of econometrics, as taught in textbooks and embodied in the practice of mainstream econometricians.[1] As already described in Chapter 1, devising appropriate estimators for the parameters of structural models a priori postulated by economists has become the prime task of econometrics resulting from the legacy of the CC paradigm; Haavelmo's discovery of the 'OLS bias' and the CC invention of the LIML estimator have given rise to an extremely vibrant production line of estimators for various models, even long after the majority of applied modellers have abandoned the LIML and related methods in favour of the OLS method. In Chapter 1, the sustained supply of estimators is attributed to the compartmentalization of econometric measurement issues by the CC programme. Estimation, in particular, has become a well-defined technical issue, which falls comfortably into the domain of mathematical statistics, provided that the structural model of interest is given and known to be fixed. In practice, however, a priori postulated structural models are seldom known to be fully fixed after they have entered the empirical laboratory. Applied modellers who are keen to obtain statistically robust and economically sensible parameter estimates often find themselves in a far less compartmentalized situation than what is required for those statistically optimal estimators. They cannot avoid the thorny issues of deciding which 'structural' parameters are estimable from available data samples and which economic attributes the coefficients should embody.

This chapter looks into the history of how such decisions were made and how they interacted with the device of estimators during the post-CC and reformative era. The investigation is focused on tracing those decisions which have led to diverging research routes and which challenged the CC

[1] 'Parameters' are often used interchangeably with 'coefficients' in econometrics. The term 'parameter' is used here for its wider connotation than 'coefficient'. However, the topic of this chapter is on 'structural parameters' or parameters of (economic) interest, which exclude 'nuisance' parameters such as the variance of a residual error term.

paradigm. The history is portrayed via three distinct but not unrelated areas: remedies and methods developed for non-constant parameter estimates (Section 7.2), remedies and methods developed for estimating individual parameters when they contained joint effects of a set of correlated explanatory variables (Section 7.3), and methods developed for estimating apparently static structural relationships with time-series data (Section 7.4). The history reveals that diverse decisions, often made inadvertently, have led to diverse remedies and methods, some serving conflicting purposes (see Section 7.5 for a more detailed discussion). To a great extent, the diverse decisions reflect equivocal and often arbitrary use of the notion of 'structural' models or parameters. Hence, it is necessary to first prepare a background for our investigation, that is to describe briefly how the concept of 'structural' parameters emerged and evolved in econometrics. That is the topic of Section 7.1.

7.1 Prelude: the conceptualization of structural parameters

The notion of 'structural parameter' is undoubtedly derived from that of 'structural model' or 'structural relation'. The concept was first introduced and specified by Frisch. In his 1930 Yale lecture notes (see Bjerkholt and Qin, 2010), Frisch classified three types of relations in economic theory: 'structural, confluent and artificial' (Bjerkholt and Qin, 2010: 73), and defined a '*structural relation*' by two specific properties: (i) 'it holds good identically in the variables involved' and (ii) 'it holds good even though we do not assume any particular side relation to be fulfilled' (Bjerkholt and Qin, 2010: 75). In the present terminology, property (i) means that the relation holds constant and valid irrespective of what values the variables involved will take, and property (ii) states that the relation holds constant and valid even when the way (i.e. 'the side relation') in which the explanatory variable in the relation is generated has been altered. A relation which satisfied (i) but not (ii) was defined as a '*confluent relation*' while one in which neither property holds was a '*artificial relation*' (Bjerkholt and Qin, 2010: 75). To Frisch, structural relations formed the core of economic theory. In a subsequent lecture on business cycles given at the Nordic Economic Meeting in 1931, Frisch used 'structure' for the 'economic theoretical structure' characterizing business cycles and 'prime relations' as an alternative to what he had defined earlier: 'A prime relation is a relation between a certain number of variables, the relation being such as to hold good identically in all the variables involved, and furthermore such as to hold good quite independently of whether certain other relationships are fulfilled or not. Thus economic prime relations are relations of an autonomous sort' (Qin, 2013, vol. I: 21). The term 'autonomous' later became a summary adjective of the essential properties for those economic relations, e.g. in Frisch (1938) and also in Haavelmo (1944).[2]

However, the notion of 'prime relations' was soon submerged under 'structural' relations, and the relations were essentially labelled as those a priori postulated theoretical ones due probably to the influential impulse-propagation business cycle model by Frisch and Waugh (1933). It was in that paper that 'structural coefficients' were used to denote the parameters of the structural relations. Frisch's classification of a structural relation versus a confluent one was retained in Haavelmo's (1938) paper, where the '*law*'-like attribute of the structural relation was embodied by 'absolutely fixed coefficients'. Haavelmo modified his position to some extent later when he described the attribute of 'autonomy' as 'a highly relative concept' with respect to the state of whether a relation was affected by 'structural changes' (Haavelmo, 1944: 29). However, he did not specify precisely what was meant by 'structural changes'.

The term 'economic structure' was described as the 'very mechanism' generating economic data in CC Monograph 10 (Marschak, 1950). Marschak also designated economic theories as 'hypotheses' about the structure, which generally took the form of a system of simultaneous relations. On the basis of a structural SEM, Marschak referred to all the parameters contained in the model as 'the structural parameters', and designated their sample-based estimates as embodying the 'observational structure'. He acknowledged that there could be 'structural changes' in a different time period or geographical area and distinguished two kinds of structural changes: 'controllable' policy changes and 'uncontrollable' changes set off by uncontrollable exogenous variables. In his view, it was the primary task of economists to analyse the effects of controllable policy changes and hence 'structural determination' was essential in econometric research.

In CC Monograph 14, Marschak (1953) further grouped structural parameters and exogenous variables together under the label of 'conditions' as opposed to 'results', that is the set of endogenous variables. Specifically, 'conditions' which underwent changes during the period of observation were assigned to exogenous variables while conditions which remained constant during the observation period were assigned to the 'structure', represented by the structural parameters, which might or might not change in future. Marschak summarized,

> In economics, the conditions that constitute a structure are (1) a set of relations describing human behavior and institutions as well as technological laws and involving, in general, nonobservable random disturbances and nonobservable random errors of measurement; (2) the joint probability distribution of these random quantities. (Marschak, 1953: 26)

[2] For a historical study of the term, see Aldrich (1989).

Another related notion that Marschak defined was 'dynamic structure'—'a structure that would admit variations of observed endogenous variables, even if exogenous variables did remain constant and if there existed no random disturbances' (Marschak, 1953: 18). Examples of such a structure included Tinbergen's 'final equations', that is, deterministic difference equations characterizing the time paths of endogenous variables.

It is particularly interesting to observe here that the essential property of parameter invariance used originally by Frisch to define structural relations was significantly compromised in the CC programme—parameter constancy becomes limited only to the available sample period, within which the structural model of interest is fitted. Meanwhile, a priori postulated relations depicting certain economic behaviours are granted the essential condition to define a structural model and an SEM is the most general form of such a model.[3] That position has been reinforced through the formal discussion of identification and the related causal aspects (e.g. Hood and Koopmans, 1953: chs 2 and 3). The discussion gives 'structural parameters' effectively the status of 'causal parameters'. The position has been largely maintained in textbook econometrics. In the first set of the *Handbook of Econometrics* (Hsiao, 1983), for example, structural parameters are simply given to the parameters of any a priori assumed probability distributions of stochastic relations without explicit reference to any economic contents.

7.2 Diagnosis and treatment of non-constant parameter estimates

As described in Chapter 1, Klein was the only practitioner who carried out an immediate and major application of the CC's theoretical work to modelling the US economy. His experiments soon illustrated that indeed not all the structural parameter estimates would remain virtually constant when data sample periods extended, despite the statistical optimality of the estimators used. Consider the famous Klein–Goldberger model (1955) for example. The model was estimated twice, first with subsample data of 1929–1950 and then the full sample 1929–1952. A number of parameter estimates differed significantly between the two sets. Klein and Goldberger referred to the difference as 'structural change' and observed:

> The problem of detection of bias from the pattern of recent residuals is intimately connected with the problem of structural change. A pattern of several

[3] Note that the initials 'SEM' can also stand for 'structural econometric model' or 'structural economic model', as has been used popularly in the literature much later, e.g. Reiss and Wolak (2007).

> residuals of the same sign and similar magnitude may actually reflect a changed parameter. In this case the bias is more permanent and, in place of temporary corrections for nonzero values of disturbances, we must search for more permanent alterations of parameters. (Klein and Goldberger, 1955: 78)[4]

From the practitioner's viewpoint, Klein and Goldberger recommended that 'the best practice in forecasting from econometric models is to recalculate all parameter estimates from the latest revised set of data' (Klein and Goldberger, 1855: 89). As for their model-based forecasts, an explicit condition was attached—'no structural change between sample and forecast period' (Klein, 1953: 252).

Evidence of cross-sample shifts in parameter estimates was also observed from micro-econometric studies. For instance, Wold and Juréen (1953) estimated the market demand elasticity for butter and margarine using Swedish data samples of 1921–39 and 1926–39 respectively. They obtained relatively constant elasticity estimates for butter but significantly shifting elasticity estimates for margarine. The results were interpreted as evidence of a changing demand pattern for margarine. Since forecasting was not as often the primary purpose here as in macro-econometric modelling, cross-sample shifts in parameter estimates tended to be interpreted as a manifestation of economic behavioural shifts and accepted as such.

However, empirical evidence of cross-sample shifts in parameter estimates and referencing of these shifts as 'structural changes' led to the perception that the occurrence of such shifts was normal and could be tackled by better devised statistical methods, such as estimators and tests devised specifically for regression models 'obeying two separate regimes' (Quandt, 1958). One of the most popular devices developed at the time was the Chow test, a test of equality between parameter estimates from two sample sets. In the opening statement to substantiate the importance of his test, Chow (1960) maintained that 'often there is no economic rationale' in assuming the same parameter value for different sample groups or periods. Note that this view has virtually revoked the CC position in that parameter constancy is no longer required even for the available sample period, since, in principle, the period could be arbitrarily divided into subsamples. Parameter constancy has thus ceased to be an inherent property of a structural model, but a testable hypothesis at best, if not a stroke of luck at the mercy of recalcitrant data.

Subsequent development of the time-varying random parameter models showed just how accommodating the device of statistical methods could be. Caught in between the classical assumptions of the residual error terms from

[4] An easy type of permanent alteration is to revise the intercept values during model updating. The revision became known as 'intercept correction' and was commonly used in maintaining macro-econometric models in action (e.g. Evans and Klein, 1968).

the statistical side and the given fixed structural models from the economic side, explicit randomization of parameter estimates naturally became the only option. A formal treatment of random-parameter models was presented by Hildreth and Houck (1968), who related their research to Ruben (1950) for justification, a very short paper in CC Monograph 10.[5] Specifically, Hildreth and Houck proposed to relax a fixed-parameter static regression model, such as:

$$y_t = \alpha + \beta x_t + u_t \tag{7.1}$$

by a random-parameter specification:

$$y_t = \alpha + \beta_t x_t + u_t; \qquad \beta_t = \beta_0 + v_t \quad \text{or} \quad \beta_t = f(z_t) + v_t. \tag{7.2}$$

Interestingly, Hildreth and Houck gave the option of conditioning the randomized β_t on a new set of variables, z_t, in (7.2). Their explanation was that certain case-specific factors, z_t, were often omitted in structural models like (7.1) and thus caused 'instability' in estimated β over different samples. But they did not delve into the implication of adding z_t to the given structural model specification and, instead, dedicated themselves to pursuing the derivation of statistically consistent estimators for β_t.

The 1970s saw thriving research into time-varying random parameter models. A significant contribution was made by T. F. Cooley jointly with Prescott. In their initial paper, Cooley and Prescott (1973) regarded a typical equation in econometric models, such as (7.1), as a linear approximation of a more complex relation. They diagnosed the frequently observed phenomenon of residual autocorrelation when model (7.1) was fitted using time-series data as an omitted-variable problem. More specifically, the autocorrelation was believed to be caused by non-transitory shocks from certain latent omitted variables. Drawing inspiration from Friedman's (1957) classification of income and consumption into permanent and transitory components and Muth's (1960) adaptive forecasting model, Cooley and Prescott proposed decomposing the error term into a transitory part and a permanent part, and to explain the latter by a time-varying intercept. The resulting model was referred to as an 'adaptive regression' model:

$$\begin{aligned} &y_t = \alpha_t + \beta x_t + u_t; \quad \alpha_{t+1} = \alpha_t + v_t, \quad \operatorname{cov}(u_t \quad v_t) = 0; \\ &\left.\begin{aligned} &\operatorname{var}(u_t) = (1-\theta)\sigma^2, \\ &\operatorname{var}(v_t) = \theta\sigma^2 \end{aligned}\right. \qquad 0 \le \theta \le 1 \qquad . \end{aligned} \tag{7.3}$$

When $\theta = 0$, (7.3) collapsed into (7.1). Cooley and Prescott argued that model (7.3) was superior to the commonly used autoregressive error procedure,

[5] Further justifications were later provided by Raj and Ullah (1981), which went back as far as the works of J. Neyman and Frisch during the 1930s.

known as the Cochrane–Orcutt procedure in textbooks (see Section 7.3), for not imposing any implicit common dynamic patterns to the omitted variables, and also for providing an explicit justification as well as a formal framework for the ad hoc practice of intercept corrections by empirical modellers. Similar to Hildreth and Houck (1968), Cooley and Prescott focused their attention on the derivation of statistically consistent parameter estimators for (7.3).

In a later paper, Cooley and Prescott (1976) further generalized model (7.3) and revised their position on the model justification. On the basis that 'in many instances economic theory suggests that relationships will change over time', they argued:

> The structure of an econometric model represents the optimal decision rules of economic agents. From dynamic economic theory we know that optimal decision rules vary systematically with changes in the structure of series relevant to the decision makers. It follows that changes in policy will systematically alter the structure of the series being forecasted by decision makers, and therefore, the behavioural relationships as well. (Cooley and Prescott, 1976: 167)

It is interesting to note that the argument no longer associates time-varying parameter estimates with the omitted-variable problem. Instead, time-varying parameters are regarded as the structural representation of the changing behaviour of agents as they adapt to changing economic reality, a position which bears close similarity to Lucas's (1976) critique.[6] Time-varying parameter models in this vein assume the image of being more general than the conventional models such as (7.1). Otherwise, research along this route follows the pattern of the CC paradigm, that is, the focus of attention is on devising optimal estimators and identification conditions for given time-varying parameter models (e.g. Raj and Ullah, 1981; Machak et al., 1985).

Time-varying parameter models apparently provided a solution to the Lucas critique, but the models were ill-suited to the practical purposes of model-based forecasting and policy simulations. A more conventional route was simply to rectify what was perceived to be a problem directly according to the diagnosed cause. In the context of time-series models, for example, the problem of residual autocorrelation was diagnosed as the omission of dynamic variables and therefore a general dynamic specification strategy was offered as a systematic treatment, for example as shown in the VAR approach and the LSE approach described earlier in Chapters 3 and 4 respectively.

The VAR approach was particularly interrelated with the RE movement and the Lucas critique. However, Sims was doubtful of the empirical

[6] Note that Prescott was working at Carnegie-Mellon University where the leading figures of the RE movement, such as Lucas and Sargent, were working at the time.

relevance of the Lucas critique. In his view, allowance for parameter value shifts was not necessarily the only way to represent policy shifts in structural models (Sims, 1982a, 1986) and many problems in the existing econometric models were caused by oversimplified dynamic specification in a priori theoretical models. Hence, a general VAR specification was prescribed as an effective and systematic treatment (Sims, 1980a).

To study the parameter constancy of VAR models, Sims and his co-workers employed the Kalman filter algorithm to estimate VAR models recursively, and viewed the process as mimicking the data updating process in model forecasting (e.g. Doan et al, 1984).[7] Plentiful estimation results yielded relatively constant parameter estimates over various historical episodes of known significant policy shifts. However, this route was not without cost. A vital weakness of VAR models was over-parameterization. Consider a bivariate VAR corresponding to (7.1) for example:

$$\begin{pmatrix} y \\ x \end{pmatrix}_t = \begin{pmatrix} a_{10} \\ a_{20} \end{pmatrix} + \sum_{i=1} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}_i \begin{pmatrix} y \\ x \end{pmatrix}_{t-i} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}_t . \tag{7.4}$$

It contains many more parameters (with the number increasing with i) than (7.3) and the parameters are also susceptible to the multicollinearity problem (see Section 7.3).

Nevertheless, the recursive estimation results by Doan et al. (1984) encouraged applied modellers to build models which would survive the Lucas critique. The LSE approach (Chapter 4), in particular, advocated the use of parameter constancy as a modelling criterion for defining exogenous variables and for checking the validity of the causal assumption in a conditional model (Engle et al., 1983). In that joint paper, exogeneity was defined on individual structural parameters and classified into three levels: 'weak', 'strong', and 'super'. Weak exogeneity was concerned with the validity of conditioning y on x via parameter β in a static relationship such as (7.1), whereas strong exogeneity was related to a Granger causality test within a VAR framework. In the case of (7.4), x_t would be strongly exogenous for y_t if the current and lagged y were found to play no significant role in explaining x_t, and the model could be transformed into:

$$y_t = a_{10} + \sum_{i=1} a_{1i} y_{t-i} + \sum_{j=0} b_{1j} x_{t-j} + u_{1t} \tag{7.5a}$$

$$x_t = a_{20} + \sum_{i=1} b_{2i} x_{t-i} + u_{2t} . \tag{7.5b}$$

Super exogeneity depicted a case when x_t was found to have time-varying parameters in its marginal distribution, for example via a_{20} or b_{2i} in (7.5b),

[7] Recursive methods were applied in econometrics as early as 1940, see Flood (1940), but got lost somehow for nearly three decades until the late 1960s (e.g. Dufour, 1982).

but the conditional model of y_t on x_t, that is (7.5a), remained parametrically invariant. Exogeneity defined as such exploited parameter shifts in single time series, for example $\{x_t\}$, in the search for parametrically invariant conditional models. The latter was seen as having close correspondence to what economists viewed as a theoretical relation in a dynamic context (see Chapter 4 and also Section 7.4). Interestingly, the definition virtually restated Frisch's two specific properties proposed over half a century before.

On the other hand, attention to marginal models like (7.5b) became revived with the rise of time-series econometrics. It should be noted that such models are in fact a stochastic version of Tinbergen's 'final equations'. Since parameter non-constancy was frequently observed from the estimation of such models, formal representation of the phenomenon found its way into stochastic switching regression models. The most popular of such models was probably Hamilton's regime-switching model built on a first-order differenced equation, because numerous 'regime shifts' or 'structural breaks' were estimated and confirmed when the model was applied to single economic time series, as described in Section 6.3. Unfortunately, the issue of how many of these stochastically located structural breaks indeed corresponded to policy-induced changes in reality was seldom addressed or even raised.

7.3 Diagnosis and treatment of collinear parameter estimates

Another area where confusion abounds is concerned with the situation where some explanatory variables of a multiple regression are found correlated with each other. The problem is referred to as 'collinearity' or 'multicollinearity' in many textbooks. The concept is rooted in Frisch's (1934) confluence analysis (see Hendry and Morgan, 1989), but its present connotation emerged after the CC formalization work in the 1950s. One of the earliest textbook definitions can be found in Tintner's (1952) textbook, where 'multicollinearity' was described as lack of 'sufficient accuracy' for individual parameter estimates (Tintner, 1952: 33), say $\hat{\beta}_1$ and $\hat{\beta}_2$ of the following regression, when there was significant correlation between x_1 and x_2:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + u, \tag{7.6}$$

and the inaccuracy was reflected in relatively large standard errors for $\hat{\beta}_1$ and $\hat{\beta}_2$.

Collinearity in the context of an SEM was discussed in Klein and Nakamura (1962). Taking the SEM as a priori fixed, Klein and Nakamura were mainly concerned with the effect of collinearity on different estimation methods, such as OLS, 2SLS, and LIML, and treated the problem largely as a computational complication. Since variable inter-correlation formed the essence of an SEM, collinearity was seen as fundamentally a data problem; its ultimate cure was believed to lie in having more data and additional information as well, for example see Johnston (1963: 207) and also Goldberger (1964: 193).

In the late 1950s, however, the problem was approached from another angle in association with the diagnosis of model 'specification errors', see Griliches (1957) and Theil (1957). It was demonstrated that parameter estimates of a regression model could be biased from the 'true' parameters of interest if the model omitted certain relevant variables which were correlated with the existing explanatory variables of interest. For example, Griliches (1957) showed how such bias could affect the parameter estimates of a production function when certain factor inputs were neglected because of the correlative nature between various factor inputs. He reckoned that correction of the bias entails due recognition and incorporation of potential collinearity. In a simple regression context, the parameter estimate, $\hat{b}$, of the following model:

$$y = a + bx_1 + e \tag{7.7}$$

would differ from $\hat{\beta}_1$ of (7.6) when x_1 was correlated with x_2 and (7.6) was believed to be the correct model. As such, this bias became known as the 'omitted-variable bias'.

Notably, both 'collinearity' and 'omitted-variable bias' stem from a set of inter-correlated explanatory variables, but the correlation interpretations implied in the two are so different that Farrar and Glauber (1967) describe them as 'schizophrenic'. From the perspective of 'collinearity', the correlation is seen as a threat to the possibility of isolating the effect of one explanatory variable from a group, but, from the perspective of 'omitted-variable bias', the correlation becomes useful in correcting the estimates of the parameters of interest. Ill-conditioned and non-experimental data are blamed for causing the collinearity problem whereas the culprit of the omitted-variable bias problem is claimed to be the inadequate model specification for excluding relevant and related variables. Generally, collinearity is undesirable in the case of multiple regression models, as implied in the usual textbook remedy to rid the models of those highly correlated explanatory variables, but in the omitted-variable bias case it becomes undesirable to neglect those collinear variables. Collinearity is discussed in a more generic way in textbooks; omitted-variable bias is a problem more often left to applied modellers to deal with on an ad hoc basis.

In fact, applied modellers had worked out how to circumvent collinearity before theorists agreed on a general way of treating the issue. For example, Wold and Juréen (1953) imposed, in their demand analysis, the sum of income elasticity and price elasticity to unity when income and price were found significantly correlated, and reformulated their models using a relative price variable to circumvent the high correlation between two price variables; Koyck (1954) proposed, in modelling investment dynamics, a simplified parametric representation of the lag structure to avoid collinearity among lagged variables; Fox (1958) decided to take the first difference of his time-series variables in order to steer clear of both inter-correlation and serial correlation of the time series; Lipsey (1960), in an attempt to better capture the dynamics and nonlinearity of the Phillips curve, chose the additional unemployment variables carefully, for example $\Delta U / U$ and U^{-2}, so that these would correlate little with the existing unemployment variable, U^{-1}, in a simple regression model (see Chapter 5). Such a practice was summarized as problem re-arrangement with the help of a priori information by Theil (1961; Section 6.2), which effectively meant model re-specification.

Theoretical research, on the other hand, divided into two almost orthogonal directions under the implicit assumption that structural models were well formulated and fixed a priori. The omitted-variable bias problem was discussed mainly in conjunction with model selection issues and dealt with by means of various information criteria (e.g. Pesaran, 1974); collinearity was handled in relation to parameter estimation of a given and selected model. One apparent route for the latter issue was to design estimators using orthogonalized regressors. The principal component method advocated by Massy (1965) was such a case. However, the statistically exploratory nature of the principal component method was unappealing to many econometricians because it was difficult to find a convincing economic interpretation for the principal components thus derived. Other methods were proposed, such as ridge regression (Hoerl and Kennard, 1970a, 1970b), or Stein-like estimators (Fomby and Hill, 1979); see also Judge et al. (1980: ch. 12). While statistical properties of the estimators and their variations were scrutinized and disputed, these methods all implied the basic need to use additional information to help reduce the correlation between parameter estimates of collinear explanatory variables.

An explicit analysis of the collinearity problem together with the omitted-variable bias problem was explored by E. Leamer (1973) (see also Chapter 2). From the Bayesian perspective, Leamer came to the view that collinearity was essentially a 'problem of interpreting multidimensional evidence' in a 'parameter-by-parameter fashion' and that its solution lay in effectively utilizing non-sample information, via well-defined Bayesian priors, such that the information would help to allocate the evidence to individual structural parameters. Leamer pointed out that the Bayesian approach formalized the ad hoc practice of restricting parameters to certain value ranges and that it transformed the collinearity problem into two aspects: the construction of Bayesian priors and evaluation of the resulting posterior distributions. The latter led to sensitivity analysis and a criterion on parameter definition. When the posterior showed high sensitivity with respect to the prior, the parameter definition should be reconsidered, as ‘regression coefficients should be defined about which prior opinions are independent’ (Leamer, 1985a). The analysis was found particularly useful in assisting specification searches in a situation where a modeller had a number of ‘focus’ variables dictated by economic theory and some other ‘doubtful variables’ to experiment with and choose from (Leamer, 1978a: 194). Here, sensitivity analysis was used to evaluate whether those parameters of the focus variables were invariant when the doubtful variables were added and whether and in what ways some of the doubtful variables should be added at all. Leamer (1985a) maintained that choice of parameterization was crucial in such specification searches whereas classical inference was silent on the issue and incapable of resolving the confusion over the collinearity problem.

What escaped Leamer’s criticism was an empirical study of the UK aggregate consumption model by Davidson et al. (1978) using the classical inference (see also Chapter 3). Unlike Leamer, Davidson et al. tackled the omitted-variable issue prior to collinearity. The specific omitted-variable problem they tried to tackle was the missing dynamics in static regression models using time-series data, and they chose dynamic models of type (7.5a) as a systematic solution to the problem (see also Section 7.4). To circumvent collinearity among lagged y_{t-i} and x_{t-j} in model (7.5a) and make the model economically interpretable, they transformed it into an error-correction model (ECM):

$$\Delta y_t = a_0 + \sum_{i=1} a_i \Delta y_{t-i} + \sum_{j=0} b_j \Delta x_{t-j} + \lambda (y - kx)_{t-1} + u_t. \qquad (7.8)$$

Compared with (7.5a), the regressors, that is, Δy_{t-i}, Δx_{t-j}, and $(y-kx)_{t-1}$, in (7.8) were far less correlated and their corresponding parameters easily interpretable. The transformation made it clear that due to ‘insufficient data information’ collinearity was only relative to a given parameterization, for example the way dynamics was specified in (7.5a), and the treatment lay in model reparameterization. Moreover, the transformation presupposed a data-congruent dynamic model where collinearity increased as a consequence of mending the omitted-variable problem. Davidson et al. (1978: 677) thus argued, ‘it is not universally valid to assume that a group of badly determined estimates indicates the presence of collinearity … rather than omitted variables bias’.

However, model reparameterization has not yet gained a formal place in the mainstream menu of remedies for mitigating collinearity (e.g. Belsley et al., 1980; Judge et al., 1980; Hill and Adkins, 2001). The resistance was essentially against the alteration of structural parameters set a priori in theoretical models, because estimates of the altered parameters could be ‘biased’ from those of the preset parameters (see Buse, 1994), for example the set of estimates for $\{a_i, b_j, \lambda\}$ of (7.8) would differ from those for the parameters of (7.5). On the other hand, it has become increasingly recognized that collinearity should not be viewed as an isolated estimation issue. Given the passive and non-experimental nature of data, modellers were advised to introduce nonsample information to 'bias' sample estimates if the information was considered 'good' (Hill and Adkins, 2001: 267). The advice effectively broke the taboo of keeping preset theoretical models fixed, although explicit model modification and design was still rarely treated as a proper task by mainstream econometricians. Few of them would step beyond the realm of devising estimators for given models and handle collinearity as a multitask involving modifying models with respect to data features. For applied modellers who found the need for model modification a routine task, most of the estimation remedies for collinearity appeared to serve cross purposes and were therefore seldom practised.

7.4 Specification and estimation of static parameters using time-series data

If collinearity is seen as a nuisance in estimating static multiple regression models using time-series data, what poses as a serious conundrum is the issue of how to deal with highly trended variables. The literature is commonly traced back to the early 1930s,[8] when Frisch and Waugh (1933) examined two commonly used methods, namely adding a time dummy in the regression model and de-trending the variables before running the regression. They argued that the structural parameter estimates generated by the two methods were identical on the assumption that the trend in each variable involved was simply linear and deterministic. However, their argument did not settle the disagreements on the two methods, owing primarily to different views on the deterministic linear trend assumption. The method of de-trending time-series variables prior to modelling remained popular in empirical business cycle analyses, as described in Chapter 6. For economists who were interested in structural parameters based on the equilibrium notion, trend removal was not considered sensible. For instance, Wold and Juréen (1953: 240–1) argued, in their empirical demand analysis, that an estimated elasticity from trend-free data would only be short-run whereas the parameter of interest should be long-run in nature even though the theoretical model was in a static form. Interestingly, Frisch had expressed similar ideas over two

[8] It is only very recently that an earlier work by Bradford Smith in the 1920s has been discovered, see Mills (2011). However, Smith's analysis on the matter remains almost totally ignored in the literature.

decades before in his 1930 Yale lecture notes, that is, a static relation should be seen as an equilibrium position in a dynamic model, although he had not associated the idea with static regression using time-series variables (see Bjerkholt and Qin 2010).[9]

Unfortunately, Wold's argument was, in the wake of CC Monograph 10, overshadowed by statistical concerns over residual autocorrelation, which was found to be widely present in static and also in some simple dynamic models fitted with time-series data. The collective research into the problem led by R. Stone at Cambridge University resulted in a first-order residual autocorrelation test (see Durbin and Watson, 1950) and a remedy estimation procedure (see Cochrane and Orcutt, 1949; Orcutt and Cochrane, 1949). Both were soon adopted in econometrics textbooks and became widely used among applied modellers.

The Cochrane–Orcutt procedure presumed to modify (7.1) by a first-order residual autoregressive scheme:

$$y_t = \alpha + \beta x_t + u_t; \qquad u_t = \rho u_{t-1} + \upsilon_t. \tag{7.9}$$

Model (7.9) was later named the COMFAC (common factor) model (Hendry and Mizon, 1978; see also Chapter 4). This model gained its popularity from its apparent preservation of the given structural model while alleviating the residual autocorrelation problem. It even rendered the error term of the model, u_t, a new structural status as being the 'structural autoregressive disturbance' (see Chapter 8). Nevertheless, the model was shown as a restricted form of a general dynamic model in a single-equation model context by Durbin (1960b) and in an SEM context by Sargan (1961). For example, a general dynamic model with respect to (7.9) should be a conditional model (7.5a) with one lag, that is:

$$y_t = a_0 + a_1 y_{t-1} + b_0 x_t + b_1 x_{t-1} + \upsilon_t, \tag{7.10}$$

whereas (7.9) amounted to imposing a nonlinear restriction on the parameters of (7.10): $b_1 = -a_1 b_0$. Sargan (1964) observed that the restriction was testable and that the test result could suggest changing the lag structure of the model from (7.9). But neither Durbin's nor Sargan's demonstration was heeded at the time, nor was the economic implication of the restriction.

On the applied front, model (7.9) was often simplified into a differenced-variable or growth-rate model (e.g. Stone, 1954):

$$\Delta y_t = \beta \Delta x_t + \upsilon_t \tag{7.11}$$

[9] Another method of implementing the idea was via choice of data. For example, Kuznets took ten-year average data to estimate the US consumption function while the estimated functions using shorter sample periods or cross-section data were interpreted as the short-run functions, see Thomas (1989).

since (7.11) was shown to be an expedient shortcut when the first-order autocorrelation parameter, ρ, was found to be close to one (e.g. Cochrane and Orcutt, 1949). However, the model was recognized as unsuitable for studying long-run parameters but possibly appropriate for short-run ones (e.g. see the short-run electricity demand model in Fisher and Kaysen, 1962: ch. 1). One alternative, proposed in the context of estimating long-run demand parameters, was to use the moving average of the explanatory variable to filter out the short-run information before running a static regression (e.g. Working, 1954). In other words, (7.1) was modified into:

$$y_t = \alpha + \beta\left(\frac{1}{m}\sum_{i=0}^{m} x_{t-i}\right) + u_t. \tag{7.1'}$$

But the method was rejected because of interpretation difficulties.[10] A new alternative was put forward by Nerlove (1958), who extended (7.1) by assuming that the explained variable followed a simple adjustment rule:

$$\left(y_t - y_{t-1}\right) = \gamma\left(y_t^e - y_{t-1}\right) \tag{7.12}$$

where $y_t^e = \alpha + \beta x_t$ in (7.1). Combining (7.12) into (7.1) resulted in a dynamic model, later known as the partial adjustment model:

$$\begin{aligned} y_t &= \alpha\gamma + (1-\gamma)y_{t-1} + \beta\gamma x_t + u_t \\ &= a_0 + a_1 y_{t-1} + b_0 x_t + u_t \end{aligned}. \tag{7.13}$$

The parameter of the static model, β, shown by Nerlove and Addison (1958), was the long-run elasticity in (7.13) and its estimate could be derived from the regression estimates of a_1 and b_0.

Many subsequent empirical demand studies adopted Nerlove's partial adjustment model. But it was not until the 1970s that his approach was generalized. The generalization formed an important part of the LSE dynamic specification approach led by D. F. Hendry (e.g. Pagan, 1987 and also Chapter 4). In an empirical study of UK building societies, undertaken jointly with Anderson, Hendry observed that 'economic theories remain for the most part of the long run equilibrium-comparative statics variety' and that a promising way to model the equilibrium was via 'a short-run control-theoretic model'.[11] It was from this standpoint that he rejected differenced-variable models such as (7.11) for removing long-run components altogether (Hendry and Anderson,

[10] The debates were mainly published in the *Journal of Farm Economics* and summarized by Nerlove and Addison (1958). However, the method of using moving averages to represent long-run tendencies has been used in other areas, such as international comparison of economic growth.

[11] Note that the method of studying the dynamic and long-run effects of an estimated dynamic model was developed earlier, e.g. Theil and Boot (1962) for the introduction of 'impact, interim, and total multipliers'.

1975). The viewpoint was clarified later in Hendry and Mizon (1978), where (7.11) was transformed into a short-run control-theoretic model:

$$\Delta y_t = a_0 + b_0 \Delta x_t + (a_1 - 1) y_{t-1} + (b_0 + b_1) x_{t-1} + \upsilon_t. \qquad (7.10')$$

Compared to (7.10′), model (7.11) was clearly a special case assuming the two level terms, that is, $a_1 = 1$ and $b_0 + b_1 = 0$, to be insignificant, an assumption implying the absence of any long-run relationship between x and y, and thus a rejection of model (7.1) in principle. The rejection was also demonstrated via an implicit inconsistency in the expected statistical properties of the error terms. When (7.11) was used as an alternative for (7.1) and its error term was expected to satisfy the usual classical assumptions, that could amount to a nonstationary error term in (7.1), since $\upsilon_t = \Delta u_t$ implied $u_t = \sum \upsilon_t + u_0$.[12] As for model (7.9), Hendry and Mizon offered an explanation of Hendry's (1971, 1974) earlier empirical finding that the COMFAC parameter nonlinear restriction seldom held—the model assumed implicitly an identical lag structure on the variables, when (7.9) was transformed into:

$$(y_t - \rho y_{t-1}) = \alpha + \beta (x_t - \rho x_{t-1}) + \upsilon_t. \qquad (7.9')$$

Such an assumption was obviously too restrictive for most pairs of economic variables. Hendry and Mizon therefore argued that applied modellers would be better off to start modelling from a dynamically unrestricted model such as (7.11), simplify it to what the data would permit and examine the model's long-run solution via (7.11′) with respect to the a priori equilibrium-comparative static theory concerned. A full exhibition of this modelling strategy was provided by Davidson et al. (1978) in an empirical study of UK aggregate consumption. As mentioned in Section 7.3, their study extended a simple static model based on the permanent income hypothesis into a dynamic model and transformed it further into an ECM in the form of (7.8) via (7.10′).[13] The term, $(y - kx)$, in the model, was interpreted as an error-correction term via a negative feedback parameter, λ. The term was also referred to as the disequilibrium-correction term, as opposed to the a priori postulated equilibrium relation of y conditioning upon x. In other words, parameter k, which was derived from all the parameters of a dynamic conditional relation, such as $k = \frac{b_0 + b_1}{1 - a_0}$ from (7.10), was seen as a dynamic modification of β in (7.1) and more appropriate than β in embodying the long-run equilibrium relation originally intended to fit the static model (7.1).

[12] Around the same time, time-series studies on the properties of the residuals of models using variable difference or detrending methods were provided by Chan et al. (1977).

[13] Davidson et al. (1978) acknowledged Sargan's (1964) wage-price model as the empirical origin of the ECM (see also Chapter 6).

The ECM was rationalized as an effective model of a dynamic control mechanism which dated back to its initial introduction into economics by Phillips (1954, 1958) (e.g. Hendry and von Ungern-Sternberg, 1981; Salmon, 1982; Nickell, 1985). But a much greater force to popularize the model came from the development of 'cointegration' theory (see Granger, 1981, 1983), as mentioned in Chapter 4. The theory provided a formal rationalization of the ECM using non-stationary (i.e. highly autocorrelated with unit roots) time-series data. In particular, a two-step estimation procedure devised by Granger jointly with Engle (1987)[14] showed how a static model like (7.1) could be linked to an ECM like (7.11) via u_t, the disequilibrium error term:

$$\begin{aligned} y_t &= \alpha + \beta x_t + u_t; \qquad \hat{u}_t = y_t - \left(\hat{\alpha} + \hat{\beta} x_t\right) \\ \Delta y_t &= a_0 + b_0 \Delta x_t + \left(a_1 - 1\right)\hat{u}_{t-1} + \upsilon_t \end{aligned} \tag{7.14}$$

where $\hat{u}_t$ was derived from OLS estimates of (7.1) in which both variables were found to be nonstationary, whereas $\hat{u}_t$ was stationary. The situation was termed as 'cointegration' of y_t and x_t with respect to $\hat{\beta}$, a consistent long-run parameter estimate. Since nonstationarity was exhibited widely in economic time series, 'cointegration' among them was frequently shown to be present and the finding reinforced the effectiveness of ECM not only among applied modellers but also theoretical econometricians and macroeconomists (e.g. Johansen, 2006).

While cointegration theory and the ECM had certainly improved a great deal of the specification and estimation of postulated equilibrium relations using time-series data, their practical significance was not so remarkable, as estimated cointegration relations of larger than a bivariate one were frequently inflicted by collinearity and the estimated feedback or disequilibrium correction impact via λ was often too small to explain much of the dynamic movements of the modelled variables of interest. Somewhat ironically, the major part of the explained dynamic movements was in fact driven by those short-run variables, for example Δy_{t-i} and Δx_{t-j} in (7.8) although the academic acceptance of ECMs relied mainly on the long-run term. That may explain why economists and econometricians who were interested in business cycle modelling moved their attention away from tightly parameterized structural models to models emphasizing the lag or dynamic structures of the variable of interest, such as VARs and dynamic factor models using detrended or differenced time-series variables, as described in the previous chapter.

[14] The draft version of their joint paper was circulated in 1985 (Hendry, 1986a).

7.5 History in retrospect

The foregoing historical study reveals that the field of research in structural parameter estimation has been a labyrinthine rather than a well-charted and self-contained one as the CC programme had intended to achieve. Moreover, the labyrinth has resulted from arbitrary and protean uses of the term 'structure'. In particular, a clear and commonly shared understanding of at least two fundamental questions is lacking: What is an adequate definition of 'structure'? Should the 'structural' attribute be defined at the level of a model/ equation or a parameter?

In the CC programme, 'structure' is primarily determined by a priori logic of causal and behavioural relations. Since there are not enough economic reasons to expect such relations to remain invariant, empirical evidence of non-constancy in parameter estimates finds an easy excuse in the interpretation of 'structural changes', 'structural breaks', or 'regime shifts', and the strategy of formalizing such changes into time-varying parameters encounters little suspicion or objection. However, even time-varying parameter models have to be dependent on certain assumed time-invariant parameters, albeit not 'structural' ones, such as θ and σ in (7.3). But more fundamentally, it is difficult to see how such models can be used for the empirical testing of theoretical models or for real-time forecasting, let alone for policy simulations.

When those pragmatic purposes of modelling are taken into consideration, 'structure' as described by the CC programme becomes obviously inadequate. In fact, the inadequacy has already been perceived by reformers of the CC programme. Perhaps the clearest statement is Sims' following remark while he was developing the VAR approach: 'Note that being causal … is only a necessary condition for an input-output mechanism to be structural. The mistake of treating this causality condition as sufficient for a relation to be structural is a version of the old *post hoc ergo propter hoc* fallacy' (Sims 1977a: 29). To correct the mistake, Sims resorts to the property of invariance: 'A better definition, I think, is that a structural model is one which remains invariant under some specified class of hypothetical interventions, and hence is useful in predicting the effects of such interventions' (Sims 1982b). The property is also endorsed by proponents of the LSE approach, as already described.

There is, however, a noticeable difference here between the VAR and the LSE approaches. In the VAR approach, invariance is attributed to models while the property of invariance is attached to individual parameters in the LSE approach. In retrospect, that position has been driven by a strong desire to make VARs accepted, as the empirical tool for policy analyses, by macroeconomists and especially those who have been brought up in the RE movement. Since the movement has drawn macroeconomists' attention away from setting individual structural parameters to the 'dynamic structures' of a set of variables, the VAR approach tries to accommodate that trend by advocating the use of 'loosely restricted multivariate time series models' particularly for conditional predictions on certain policy interventions. In Sims' (1982b) words, 'whether a model is structural depends on the use to which it is to be put—on what class of interventions is being considered'. A more precise definition of 'structure' can be found from his comments on a comparative study of VAR and structural models by Clement and Mizon: 'a claim that the consequences of some specific class of real world events or actions can be predicted by modifying one part of the model (changing the distribution of some disturbance or modifying some equation) while holding the rest of the model fixed' (Sims, 1991: 293). But, it is difficult to imagine how a modeller can possibly fix parts of a model without fixing all the relevant parameters.

In contrast, the LSE approach is immune to the difficulty by explicitly setting the claim of invariance on individual parameters. To support the claim, it has redefined the concept of exogeneity and revived Frisch's 'autonomy', an almost extinct concept after Haavelmo's (1944) discussion. Furthermore, it has prescribed collinearity as a problem of poor parameterization and advocates taking parameter design as an essential part of model specification and selection process (see Chapter 9). However, the claim has made it very difficult to find empirical models which would sustain such invariance as time elapses further beyond the data sample period used for the model specification and estimation. Moreover, it is difficult to furnish all those empirically designed parameters with elaborately proved behavioural foundations. It is thus unsurprising to see such a data-driven position being resisted by those who have either lost interest in scrutinizing individual parameters or are still adhering to the convention of treating structural parameters as being strictly devised a priori on the basis of certain max/min behavioural assumptions.

Nevertheless, the history shows just how costly maintenance of the convention could be for empirical research. It takes more than half a century for the idea of static relations as embedded equilibrium states in dynamic models to be developed into an ECM and the associated cointegration theory. There is a twenty-year gap between the first applied ECM (Sargan, 1964) and its theorization in cointegration (Engle and Granger, 1987). The situation is no better in the case of collinearity, as there is no established position reached between textbooks and practitioners. Theoretical discussion of the problem remains focused on estimation remedies within the confinement of structural parameters being a priori fixed and given, as shown from the reference lists in Hill and Adkins (2001). The diagnosis of inadequate parameterization from applied modellers such as Leamer (1973) and Davidson et al (1978) has received little attention there.[15] Nor is the discussion related to the issue of

[15] A citation search through Web of Science shows almost no references to Leamer (1973) and Davidson et al. (1978) as far as the collinearity issue is concerned.

omitted-variable bias. In applied studies, on the other hand, omitted-variable bias is of far greater concern than collinearity and hardly any of the estimation remedies for collinearity have been widely used.

The enduring gulf can be traced to the consolidation of the CC paradigm. By taking a priori postulated parametric models as maintained hypotheses and making parameter estimation a relatively self-contained technical issue, the paradigm has left no room for information from estimation to systematically loop back into better structural model building. Without such a feedback route, econometricians have taken lengthy detours in devising apparently innocuous remedies when empirical results have turned out to be undesirable. Various time-varying parameter methods and estimators for models allowing for autocorrelated residual errors are just a few of such examples. Problems with the remedies are not easily realized until they have been repeatedly proved counter-effective in real-world practice. As shown from our historical study, they could, empirically, impair the 'structural' function intended of the models and so, methodologically, camouflage incorrectly or incompletely built models for statistical purposes.

In fact, it is unrealistic to expect a priori postulated models to be complete. Their incompleteness has always been covered by the commonly used *ceteris paribus* assumption. It seems natural to leave econometricians to deal with the assumption when it comes to practice. But few have taken on the job directly or systematically. In a way, the labyrinth which they have circumnavigated can be seen as a result of tackling the job indirectly via parameter estimation. The *ceteris non paribus* situation is tucked into randomized structural parameters in the time-varying parameter approach, while much of it is hopefully absorbed into various estimators allowing for autocorrelated residual errors in the case of the COMFAC model. Either way, it is clear that the *ceteris non paribus* gap is unlikely to be covered by the naïve statistical assumptions of white-noise residual error terms in most circumstances. On the other hand, the practice of attending directly to possible omitted variables has continued to be regarded as ad hoc or 'measurement without theory', even when those omitted are lagged variables, as in the case of the VAR approach and the LSE approach. It is only until the development of a cointegration theory that the re-specification of an apparently static parameter, such as β in (7.1), into a long-run parameter, such as k in (7.8), has finally gained a theoretical status relatively widely. Such an upgrade has not yet reached other parameters, such as a_i and b_j in (7.8). In the situation of empirical studies using cross-section data, it is common practice to fill the *ceteris non paribus* gap tentatively by additional 'control variables'. But none of those associated parameters are considered explicitly as theoretical or structural ones.

It thus remains to be seen how and when the research community will establish a consensus on what constitutes the necessary properties of 'structural' parameters within a statistically well-defined model and to what extent the design and specification of such parameters should be shared by theoretical as well as applied modellers. What is learnt from the current historical examination is that segregation of duties between economists and econometricians as well as isolation of the core econometric task to device of estimators is far from the optimal research strategy, because parameter estimation has been shown to be much more useful as a medium in the search for better models than as the final step in measuring a priori specified parameters per se. Inadequate recognition of the inseparability of the measurement task from the task of model building may only extend the intellectual labyrinth. This is further demonstrated from the perspective of the error terms in the next chapter.

8

Evolving Roles of Error Terms

In contrast to structural parameters, the estimation of which has occupied the bulk of econometricians' attention, the status of error terms in econometric models remains 'residual'. These terms are often seen as a nuisance during the estimation process. This status of nuisance, however, makes error terms intriguing. If one delves into their roles, one may notice that major methodological shifts in the history of econometrics have coincided with changing interpretations of error terms. To a large extent, the history of econometrics may be seen as tidal oscillations between different and often disparate interpretations. Many econometricians view error terms broadly as what is omitted from statistical models, but some regard them as measurement errors in latent theoretical variables; for those who are deeply involved in theorizing business cycles, error terms are seen as meaningful stochastic shocks responsible for driving the dynamics of modelled variables. In fact, the varied interpretations are best reflected in the multiple names assigned to error terms, such as 'residuals', 'disturbances', 'aberrations', 'accidental variations', 'random shocks', 'errors in variable', and 'errors in equation'.

This chapter examines the evolving roles of error terms in econometric models, particularly during the post-1970 period.[1] The examination offers a mirror view of how the mainstream methods based on the CC tradition attracted increasing criticism and how a plethora of rival methodologies emerged from these debates, as already depicted in Chapters 2–4.By presenting these debates from a different angle (Sections 8.2–8.4), a view will emerge that should deepen our understanding of why those debates were sometimes exchanged at cross-purposes, what led modellers of different camps to choose their preferred standpoint, and what made the models that they built or maintained serve different purposes. Before we proceed, however, a summary is needed of the mainstream interpretations of the error terms formed

[1] This chapter is an extension of Qin and Gilbert (2001).

with the establishment of the CC paradigm. This need defines the task of the following section.

8.1 Structural approach and errors in equations

Error terms are commonly defined as ‘errors in equations’ in econometrics textbooks. The definition prevailed with a consolidation of the CC paradigm. Prior to the CC seminal contribution, interpretations of error terms varied, largely depending on modellers’ views of the theories of concern. For those who took theories formulated in a deterministic model form as correct and complete representations of reality, residuals from such fitted models were seen as measurement errors, or errors in variables due to ‘disturbance of data’ (e.g. see the empirical studies of American agricultural economists described in Morgan, 1990). But for those who were wary of theoretical models, residuals were viewed as representing the incorrect and incomplete aspects of the models and therefore as comprising two types of factors—‘disturbances’ caused by incorrectly excluded highly significant variables and ‘accidental variations’ caused by intentionally omitted negligible variables as part of the incomplete nature of model building (e.g. see Frisch’s discussion in Bjerkholt and Qin, 2010: ch. 3). However, the theory-based demarcation was crossed when dynamic and stochastic theoretical models were postulated. In the Slutsky–Frisch impulse-propagation macro dynamic model, for example, cyclical motions of modelled variables were shown to be driven by random shocks as part of explanatory variables in the model, see Frisch (1933) and Slutsky (1937). The shocks were later referred to as ‘stimuli’, to be differentiated from ‘aberrations’, that is, residuals from fitted models (Frisch, 1938). But Frisch stopped short of providing a working scheme of how the two types of errors should function in a unified empirical framework, since the Slutsky–Frisch model was presented without fitting to data at the time. Their approach was revitalized four decades later by macroeconomists in the RE movement. Meanwhile, measurement errors were also given behavioural interpretations. For instance, Tintner (1938), in the context of an errors-in-variables model, related the errors in a modelled variable to agents’ random failures to achieve the optimum indicated by economic theories and the errors in the explanatory variables to agents’ failures to forecast those variables correctly since, he argued, those variables were generally beyond their control. Such a behavioural interpretation helped strengthen the assumption, largely held implicitly by modellers, that given theoretical models were correct and complete.

The seminal works by the CC group played a significant role in strengthening that assumption. However, measurement errors faded out of their interpretations. Concerned with the task of estimating structural parameters of a priori given structural models, measurement errors or errors in variables were seen as having less of an effect on parameter estimation than errors due to omitted variables, and error terms were thus treated as solely due to the latter, that is errors in equations, as a simplification of the estimation task (e.g. Koopmans, 1937). The subsequent probability approach advocated forcefully by Haavelmo (1944) was largely built on models of the errors-in-equation type. Taking an SEM as the general form of a priori structural models, Haavelmo defined errors in equation as unsystematic disturbances with a jointly normal distribution. The 'joint' specification was to embody simultaneity while the unsystematic aspect connoted Frisch's 'accidental variations' with an implied correctness of the SEM. Haavelmo's specification was carried over in the CC work, where a general structural SEM was defined as:

$$Ay_t' + \Gamma w_t' = \varepsilon_t' \qquad \text{with } w_t = \begin{bmatrix} y_{t-1} & \cdots & y_{t-\tau} & 1 & z_t & z_{t-1} & \cdots & z_{t-\tau} \end{bmatrix} \quad (8.1)$$

where a set of variables $x = \{y, z\}$ were grouped into a vector of endogenous variables, y, and a vector of explanatory variables, z, and ε_t was the jointly and normally distributed error term. Moreover, ε_t was interpreted as 'disturbances (or shocks), which represent the aggregate effects of additional unspecified exogenous variables on the economic decisions expressed by each relation' (Koopmans and Hood, 1953: 115).[2] The interpretation was markedly different from Haavelmo's in that it suggested a structural function of these error terms. In fact, models like (8.1) were even referred to as 'shock models' for the explicit specification of ε_t (see Hood and Koopmans, 1953: Preface).

The property of simultaneity of (8.1) gave rise to the identification issue, that is, the issue of whether the parameters in A and Γ were uniquely estimable. Discussion on the identification conditions led to the derivation of the so-called 'reduced form' of (8.1) (e.g. see Mann and Wald (1943) and Koopmans (1950)):

$$y_t' = -A^{-1}\Gamma w_t' + A^{-1}\varepsilon_t' = \Pi w_t' + u_t'. \quad (8.2)$$

The discussion also led to a new interpretation of model 'completeness' with respect to an SEM: the SEM was considered complete when all its structural parameters were identifiable, see Koopmans (1950). The interpretation ignored the possibility of a priori postulated structural models omitting

[2] A more detailed explanation of the error terms was recorded in one of the Cowles documents: '*Disturbances* ordinarily include both "shocks" and "errors." Shocks are those factors a given theoretical system does not explicitly take into account or cannot explicitly take into account. They are usually small factors not separately noticeable. Errors are those in observation or measurement, which cannot in any case be made part of the economic theory underlying a system of equations. Unless otherwise stated, disturbances as used later will include only shocks, errors being assumed to be negligible' (Cowles Commission, 1947).

highly significant variables, a possibility that was closely associated with Frisch's idea of model completeness. From the perspective of ε_t in (8.1), the interpretation was embodied in the imposition of statistically classical conditions, namely that the assumed jointly normal distribution was also serially independent and identically distributed (IID). The imposition suggested Haavelmo's interpretation of ε_t but it was unclear how that could coexist with the interpretation of aggregate exogenous shocks.

However, there was still no unanimous agreement on the imposition of the IID conditions among the CC researchers. One particular condition that remained in doubt was serial independence. The doubt reflected particularly the impact of the ongoing business cycle modelling literature. For example, Hurwicz (1944) illustrated the possibility of having serially dependent or autocorrelated error terms by the following simple, dynamic SEM with respect to business cycle models:

$$\begin{cases} Y_t = bY_{t-1} + a_1 Z_{t-1} + a_2 Z_{t-2} + \varepsilon_t' \\ Z_t = \qquad\qquad cY_{t-1} + \varepsilon_t'' \end{cases}. \tag{8.3}$$

When (8.3) was transformed to a single autoregressive-moving average (ARMA) equation:

$$Y_t = bY_{t-1} + a_1 cY_{t-2} + a_2 cY_{t-3} + \eta_t, \tag{8.4}$$

the error term, $\eta_t = \varepsilon_t' + a_1 \varepsilon_{t-1}'' + a_2 \varepsilon_{t-2}''$, was obviously autocorrelated even if ε_t' and ε_t'' in the original structural model (8.3) satisfied the IID conditions. Hurwicz thus referred to η_t as 'composite disturbance' and concluded that 'it might well be that some of the distrust with which the "literary" economists have viewed the "mathematical" business-cycle theory has arisen from their opposition to the unrealistic postulate of non-autocorrelated "disturbances"'.

Hurwicz's illustration implied that the serially independent restriction on error terms might be merely a simplifying assumption to ease the task of estimation but the restriction could be at odds with some structural models, particularly those characterizing business cycles. Hurwicz's message was sympathetically shared by his contemporaries. In the introductory chapter of CC Monograph 14, Marschak maintained that possible autocorrelation in the disturbances 'must be considered part of the structure' on the grounds that there were no economic reasons to rule out the possibility of their forming a 'stochastic process, each shock depending on one or more of its predecessors', especially when the observation frequency became high, for example 'weekly or even quarterly instead of annual time series' (Marschak, 1953: 21–3).

Interestingly, the CC structural interpretation of possible autocorrelation in error terms is resonant of Frisch's efforts to handle both aberrations and stimuli in one business-cycle model framework. But further progress was deterred by the focus of the CC group on estimation. Since Hurwicz's model showed autocorrelation in the error terms of the 'reduced form', a model form which was regarded merely as a convenient vehicle for estimation, the feature of residual autocorrelation was subsequently seen as a problem to be dealt with by estimator designs. That perception soon became widely accepted as more and more empirical evidence confirmed the prevalence of residual serial correlation in macro-econometric models or models fitted by time-series data. Few tracked the problem to the wide gap between the generality of the abstract structural model (8.1) and the frequently oversimplified formulation of structural models in practice, for example (8.3), not to mention that many such models were of the static type.

8.2 Errors in equations during the consolidation years

To a large extent, the consolidation of the CC paradigm was embodied in the prevalence of the view among econometricians that the key role of the error terms was to assist the design of statistically best estimators of a priori given structural parameters. The classical IID conditions were expedient for acquiring such estimators. Nevertheless, error terms were treated as a necessary intermediate means of parameter estimation, otherwise devoid of any significant economic meaning.

One line of research which contributed substantially to the prevalence of this view was the design of estimators for static structural models augmented by autocorrelated errors. The research was motivated by widespread findings of autocorrelated residuals from fitted structural models using time-series data. A leading figure in this research was G. Orcutt. As described in Chapters 3 and 4, Orcutt started the research by looking into the autocorrelation in each of the fifty-two economic time series used in Tinbergen's (1939) model. Assuming an autoregressive model of order two, AR(2), Orcutt (1948) found that many of the series could be characterized by the following AR(2) with a unit root:

$$y_{i,t} = y_{i,t-1} + 0.3\Delta y_{i,t-1} + \varepsilon_{i,t} \,. \tag{8.5}$$

If (8.5) correctly modelled the dynamics of those series, the OLS method used by Tinbergen would lose its statistical optimality. Assisted by D. Cochrane, the search for a general solution resulted in the Cochrane–Orcutt (CORC) estimation procedure (Cochrane and Orcutt, 1949) (see also Sections 4.1 and 7.4). Actually, two possible routes to a solution were proposed in that joint paper. One was 'to change some of the variables, add additional variables, or modify the form of the relation until a relationship involving what appear to be random error terms is found'; the other was to develop more elaborate estimation methods than the OLS while leaving the structural specification fixed. The latter route was chosen because economists typically specified structural models in terms of 'the most reasonable choice of variables and form of relation' (Cochrane and Orcutt, 1949) prior to data inspection. To reconcile the fixed structural models with autocorrelated error terms, Orcutt and Cochrane came up with a compromise solution—to augment the models with an autoregressive error scheme. In the case of a static structural model:

$$y_t = \beta z_t + u_t, \tag{8.6}$$

the simplest augmentation was an AR(1) of u_t:

$$u_t = \rho u_{t-1} + \varepsilon_t. \tag{8.7}$$

The CORC estimator was essentially an iterative least-squares estimation procedure based on (8.6) and (8.7).

However, Cochrane and Orcutt did not totally ignore the other route. They observed that the CORC estimator of ρ tended to be biased towards unity when the a priori models took the static form of (8.6), especially when z and y are trending variables, and recommended that it would probably be more convenient to estimate β from the following growth-rate model:

$$\Delta y_t = \beta \Delta z_t + \Delta u_t = \beta \Delta z_t + e_t. \tag{8.8}$$

However, the change in residual definitions from (8.6) to (8.8) was left out of the discussion. Their recommendation was adopted by Stone (1954) as well as the Netherlands Central Planning Bureau (see Theil, 1961; see also Section 7.4 of the previous chapter).

Noticeably, the auxiliary equation (8.7)—which distinguishes between autocorrelated disturbances u_t and white-noise residuals ε_t in a somewhat similar manner to Frisch's distinction between 'stimuli' versus 'aberrations'—suggests the possibility of reviving Frisch's ideas. However, this turns out to be illusory because the error terms analysed by Cochrane and Orcutt lacked theoretical interpretation, either in respect of the structural model (8.6) or in relation to the generation of dynamics, as in Hurwicz (1944). In other words, the addition of (8.7) to (8.6), or the move to first-differenced series tended to reinforce the prior status of theoretical models, and concentrated attention on the statistical issue of efficient estimation. The consequence was that the majority of econometric research was oriented towards developing more complicated estimation tools for fixed, frequently static, theoretical models with relatively little attention devoted to the way more complicated dynamic specifications might explain economic fluctuations.

The consequence was probably best illustrated from the subsequent works by Sargan (1959, 1961) and Durbin (1960b), following the CORC procedure.

However, they showed independently, through combining (8.6) into (8.7), that augmenting a static model by an AR(1) error process was equivalent to dynamic respecification of the original model (8.6) with a particular restriction on the coefficient of z_{t-1}:

$$y_t = \beta z_t + \rho y_{t-1} - \rho z_{t-1} + \varepsilon_t = \beta z_t + \rho y_{t-1} + (-\beta\rho) z_{t-1} + \varepsilon_t. \tag{8.9}$$

But neither Sargan nor Durbin pursued the implications of such respecification. Instead, both focused on how to achieve the optimal estimation for β in a more general way than the CORC procedure, although Sargan suggested in passing the possibility of testing the residual autoregressive setting against the more general dynamic model, that is one relaxing the coefficient restriction on z_{t-1} in (8.9). It was well over a decade before that suggestion was fully explored, as described in Section 7.4.

Meanwhile, researches into the complications of model residuals failing the other IID conditions followed a similar methodology, that is to design new estimators to restore the loss of statistical optimality in the existing estimators owing to the failures. The case of how to improve the efficiency when heteroscedasticity was observed from the residuals served as one such example (e.g. Theil, 1951; Goldberger, 1964; Lancaster, 1968). The seemingly uncorrelated regression estimation (SURE) method proposed by Zellner (1962) was another example, one which took into account possible cross-equation error correlations in a set of regression equations, each having different sets of exogenous variables. These research studies helped to strengthen the CC paradigm of taking a priori structural models as correct irrespective of dubious signs revealed from statistically non-purified residuals.

Research in applied modelling, on the other hand, followed far less uniformly from such a pedantic position. Early applied modellers were familiar with the frequent need for ad hoc adjustments to a priori postulated structural models. Theil was one of the first econometricians who tried to highlight the problems associated with such practices. Theil (1961: ch. 6) argued that such practices amounted to changing the maintained hypotheses and were therefore contrary to standard Neyman–Pearson hypothesis testing methodology. He advocated that changes in the maintained hypothesis be made explicit through what he referred to as 'specification analysis'. Statistical properties of the model residuals played an essential role in specification analysis. This implied two extensions to the previously existing framework. First, Theil proposed minimization of the residual variance as the main criterion for specification choice. He recognized that 'there is no law in economics which states that such proportions [of the disturbance] are small or even "as small as possible"', but he attempted nevertheless to justify minimization as the criterion for choosing between different model specifications (Theil, 1957). Second, Theil included both residual autocorrelation and cross-equation dependence as potential problems to be dealt with in specification analysis (Theil, 1958: ch. 6), thereby reviving the question of how model specification should relate to the statistical properties of the error terms. Theil's specification analysis effectively claimed to make it routine to respecify models and test between different (often non-nested) theoretical models since it presumed that applied modellers 'do not know the "true" specification in general' (Theil, 1958: 215). Such a position was potentially subversive to the CC paradigm of taking structural models as a priori known and fixed.

8.3 Error terms as manoeuvrable unobservables

The subversion remained largely dormant for nearly two decades until the mid 1970s. As described in Chapter 2, Leamer's (1978a) book on specification searches posed an open challenge to the CC approach. Despite the difference of statistical approach between the classical and the Bayesian standpoints, Leamer's view was surprisingly close to that of Theil. In fact, Leamer referred to Theil's view on the necessity of specification analysis and regarded it as a rejection of 'classical inference as unworkable' but nevertheless viewed it as inadequate to 'offer a procedure that would allow valid inferences in the context of a specification search' (Leamer, 1978a: 5).

Chapter 2 has shown that Leamer's specification searches were targeted at filling the substantial gap between the theoretical and applied positions in econometric modelling. Interestingly, he related such searches to applied modellers' choice in specifying the error terms, which he regarded as manoeuvrable 'unobservables'. Starting from a basic textbook regression model similar to (8.6),[3] Leamer demonstrated that such a given model amounted to 'a tautological definition' of the error term, since by definition u was 'all of those things that determine y, excluding' $\beta z : u \equiv y - \beta z$. This point was further presented explicitly as

$$u = \sum_i \gamma_i x_i, \tag{8.10}$$

where the x_i defined all the possible explanatory variables omitted in the original model. When a given structural model took the form of (8.6), z became a 'fixed' explanatory variable while those x_i in (8.10) were able to 'to vary within the confines of some more-or-less well-defined experimental conditions'. Noticeably, (8.10) enabled Leamer to show that the conventional practice of asserting the IID conditions on u was equivalent to asserting those conditions

[3] Leamer's equation actually covered the case of multiple regression without specific reference to time-series data.

on $\sum_i \gamma_i x_i$ and further that the designation of z being fixed would require that $E\left(\sum_i \gamma_i x_i \middle| z\right) = 0$, that is, the condition to rule out the possibility of omitted variable bias. By representing the two conditions in such a way, Leamer made it obvious how 'unlikely' or enormously restrictive these conditions were 'in nonexperimental research' and likewise how restrictive was the conventional position of taking the original model as 'well defined' (1978a: 65). It therefore became necessary to undertake a data-instigated specification search.

Leamer also used the ideal situation of $u = 0$ to define 'a complete theory'. Since it was such an unlikely situation in reality, the task was boiled down to finding an estimate for β through minimizing the residuals so as to 'make the theory appear as complete as possible'. The minimization made the variance of u 'a measure of the completeness of the theory' (Leamer 1978a: 66). It should be noted that there was a marked difference between Leamer's conception of completeness and Frisch's, or Koopmans's (1950) discussion on 'when is an equation system complete for statistical purposes'. Leamer's conception was essentially driven by the need for estimation of a priori given structural parameters.

Indeed, Leamer's discussion on most of the six varieties of specification searches that he classified from applied studies was centred on the posterior estimators of β resulting from experimenting with various priors, $p(\beta)$, and the corresponding sensitivity analyses of the posterior estimates (see Chapter 2). Little of the discussion was focused on the error terms and their properties. Even in the case of 'data-selection searches' which was intimately involved with the statistical properties of the residuals, such as autocorrelated or non-normally distributed disturbances, the discussion was still focused on selecting the corresponding priors and deriving the corresponding posterior estimators. The possibility of treating those non-IID symptoms in the residuals as signs of model mis-specification was totally ruled out, on the ground that data-selection searches were regarded as situations where the data evidence could not be precisely defined and 'the interpretation of the data evidence thus remains elusive'. Therefore, 'a researcher can only report features of the mapping of prior and data distributions into posteriors' (Leamer 1978a: 260). Consequently, data-selection searches were reduced to designing estimators which would accommodate those distributional complications in the residuals, in virtually the same spirit as what those non-Bayesian econometricians did following the CC paradigm. Specification searches for 'postdata model construction' was another case where discussion relating to the error terms resurfaced. However, the discussion was focused on the omitted variable bias in the estimates of β. Leamer referred to the bias as '*experimental bias*' since it effectively designated (8.6) as a '*false model*' for excluding those x_i which were correlated with z. Furthermore, he saw the bias as the only form of 'misspecification uncertainty' induced by the failure, provided that the sample size was not too small, on the ground that 'in the nonexperimental sciences, the possibility of improving an "experiment" is, by definition, excluded' (Leamer 1978a: 296–9). In the event that a modeller decided to take

$$y = \beta z + \gamma x + e \tag{8.11}$$

as the new model instead of (8.6), the decision amounted to postulating a new experiment, no matter whether it was 'data-instigated' (Leamer 1978a: section 9.4). The econometrician's job consequently became the estimation of the parameters in (8.11).

The unwavering attention that Leamer devoted to parameter estimation shows that he was allied to the CC paradigm much more than his Bayesian veneer suggested. Unlike Theil's specification analysis which encouraged applied modellers to experiment with revising model formulations and specifications when the residuals of the original model were found to deviate from the IID conditions, the treatment that Leamer's specification searches advocated was essentially to revise the priors for the originally given structural parameters under the situation. Leamer's interpretation of residuals as manoeuvrable unobservables, for example as determined by a non-uniqueness set of x_i 's in (8.10), made it effectively impossible to diagnose those statistical deviations as model misspecifications. Moreover, the Bayesian subjective stand helped him tolerate any arbitrariness in any a priori postulated structural models and hold them as maintained. To a large extent, Leamer's specification searches could be regarded as 'the prior' searches. Nevertheless, the credibility of structural models being indiscriminately maintainable was weakened by Bayesian subjectivity, as well as by the arbitrarily manoeuvrable residuals and frequently fragile results of posterior parameter estimates from sensitivity analyses.

8.4 Error terms as innovative structural shocks

In contrast to Leamer's Bayesian approach, error terms drew much more attention in the VAR approach and that attention was even at the expense of the estimation of structural parameters. Chapter 3 has shown how the VAR approach benefited from a cross-fertilization of the RE movement in macroeconomics and advances in time-series statistics. Both directions involved new aspects of interpreting error terms.

An overarching theme of the RE movement was to improve macroeconomics by designing dynamically testable and behaviourally optimizable models. Its realization was through the introduction of latent 'expectation' variables, originally proposed by Muth (1961), as a key dynamic driver of an otherwise conventional static model. These variables not only offered a powerful economic justification for the wide and general use of lagged variables in the name of the available past information, but also transformed them into a series of shocks (see the example shown in equations (3.2)–(3.7) of Section 3.2). The RE models thus revived Frisch's vision of 'stimuli' as a means of characterizing the dynamic effects of random shocks on business cycles. The revival brought back the 'shock' interpretation of the error terms by the CC group. The interpretation effectively upgraded the role of error terms to structural status by treating them as important as certain exogenous variables and possibly even more vital for being the driving processes of business cycles (cf. Whiteman, 1983: ch. 1).[4] But RE modellers were not unaware of a certain ambiguity in such a structural interpretation. For example, Lucas and Sargent (1979) acknowledged that 'restrictions … governing the behavior of the error terms … are harder to motivate theoretically because the errors are by definition movements in the variables which the *economic* theory cannot account for'; and Sargent (1978) wrote,

> optimizing, rational-expectations models do not entirely eliminate the need for side assumptions not grounded in economic theory. Some arbitrary assumptions about the nature of the serial-correlation structure of the disturbances and/or about strict econometric exogeneity are necessary in order to proceed with estimation. (Sargent 1978a: 1025)

However, their concerns were somehow disregarded when it came to the econometric implementation of the RE models.

Attentive handling of the error terms was also a salient feature of the time-series statistical approach, albeit based on quite different reasoning. They were regarded purely as 'residuals' from model building and their statistical properties were thus used as a key yardstick for model construction and evaluation. The time-series approach also attributed a new property to the error terms—'innovation' with respect to forecasting, in addition to the conventional IID conditions. With the popularization of the approach by Box and Jenkins's (1970) monograph, the econometric community became increasingly keen to combine the time-series approach into econometric modelling, including having innovational residuals as a necessary condition (e.g. Hendry, 1974; Granger and Newbold, 1974b, 1977a; Sargent and Sims, 1977).

As described in Chapter 3, the VAR approach advocated starting macroeconometric models from a dynamically unrestricted VAR of all the variables, say a set of x_t, which were considered important and indispensible on the basis of economic arguments:

$$A(L)x_t = \varepsilon_t, \tag{8.12}$$

[4] Actually, Muth's (1961) RE model was not the first to 'structuralize' error terms. An earlier example was Solow's (1957) proposal to interpret the residuals from production functions as a measure of productivity.

where $A(L)$ is a matrix polynomial of parameters in lag operator L of order n, the magnitude of which should be chosen in such a way as to ensure that the model-derived $\{\varepsilon_t\}$ should be an innovation process, that is $E(\varepsilon_t | x_{t-1}, x_{t-2}, \cdots) = E(\varepsilon_t | X_{t-1}) = 0$. Provided that $A(L)$ was invertible (which requires that each component of x is stationary), the VAR was then transformed into the moving average representation (MAR):

$$x_t = A(L)^{-1}\varepsilon_t, \tag{8.13}$$

so as to enable model-based shock simulations, known as impulse response analysis. The duality of (8.12) versus (8.13) brought about a dual interpretation of the error term, ε_t. Viewed from (8.12), ε_t was merely statistical residuals since, solely for estimation purposes, it was tacked onto a theoretical model which was 'silent about the nature of the residuals'. Through the mapping of (8.12) onto (8.13), however, ε_t acquired the interpretation of innovational 'shocks' relating to each modelled variable in the VAR, for example the error term of a money-demand equation was called 'money innovation' or monetary shock (Sims, 1980a). Sims (1978) even referred to such impulse response analysis as 'innovation accounting'. The latter aspect of a structural interpretation played a significant role in popularizing the VAR approach as a powerful means of policy analysis for macroeconomists. Noticeably, the VAR-based MAR (8.13) actually resonated, in a multivariate context, the Slutsky–Frisch impulse-propagation scheme, making Frisch's idea of 'stimuli' empirically possible. In other words, once the series $\{\varepsilon_t\}$ were interpreted as stochastic economic shocks, (8.13) was seen as a general representation, or a structural model, of how the impact of these shocks was transmitted via $A^{-1}(L)$ to generate business cycles.

However, impulse response analysis based on (8.13) requires the imposition of a causal ordering (i.e. a recursive structure) on the VAR system. This amounts to triangularization of the leading term of $A^{-1}(L)$, or orthogonalization of the error terms. But there is no unique way of doing this—alternative triangularization schemes are effectively equivalent to alternative identification assumptions within an SEM framework. VAR analysis simply reverses the traditional sequencing of model identification and estimation. The identification issue under this veil was thus seen as a new technical challenge to the VAR approach and gave rise to its revision into the SVAR approach as well as various methodological disputes concerning the structuralness of the VAR or SVAR approach referred to in Section 3.5.

Meanwhile, the practice of innovation accounting through impulse response analysis encountered little disagreement. In fact, it was extensively adopted in the RBC models, which, interestingly, were seen as being distinct from VAR models, or following a methodology in opposition to the VAR approach. RBC modellers would generally specify the error terms of their models as arising from two sources: shocks from exogenous variables, which were commonly treated as evolving from AR processes with random shocks, and measurement errors, which were actually introduced because of the necessity of model estimation (see Kim and Pagan, 1995). When significant discrepancy occurred between the values of their model simulation and the actual data, RBC modellers tended to explain the discrepancy as arising from unimportant or uninteresting aspects of the economy from which their models had been abstracted (e.g. Kydland and Prescott, 1991). Explanations of this sort aroused great distrust of RBC models among econometricians (e.g. Quah, 1995; Gregory and Smith, 1995).

In spite of the seemingly highly incompatible methodologies, the two camps actually share a common perception of the error terms—as exogenous stochastic shocks driving the cyclical movements of the endogenous variables. Essentially, such a 'structural' perception rules out the possibility of model mis-specification, particularly mis-specification due to omitted variables. In the case of the VAR approach, the practice of impulse response analysis based on (8.13) presupposes that the underlying VAR (8.12) is an economically valid and statistically autonomous characterization of the dynamics of the x variables, that is, the VAR is treated as the maintained hypothesis since the practice effectively disregards the existence of structurally unexplained residuals. The apparent lack of any restrictions in (8.12) might appear to guarantee that (8.12) is sufficiently general to embed all a priori postulated theoretical models. However, that actually depends on the choice of variables to be included in the variable set x. In practice, limited data samples tend to result in extremely small sets, frequently with fewer than ten variables. Empirical VAR models are thus not much less 'incredible' than those conventional macro structural models, if compared to Sims's (1980a) criticism of those models.

8.5 Error terms and error-correction models

The lack of credibility of innovation accounting through impulse response analysis in practice could merely be due to the over-strong assertion needed for the analysis that the VAR models were correct and complete. Nevertheless, the advocacy of the VAR approach to target innovational errors in econometric model building struck an approving chord among empirical modellers who were sharply aware of the pitfalls of taking a priori structural models as fixed. In particular, the target was adopted independently by modellers who were developing the LSE approach, as described in Chapter 4. However, the LSE modellers did not describe model residuals as structural shocks. Instead, they tried to build economically interpretable shocks within dynamically well-specified models in the form of EC model specification.

As shown in Chapter 4, the EC model specification was used to bridge time-series data and theoretical models of the partial equilibrium type, a type most often used in practice. In a stochastic framework, the type amounted to posing a conditional distribution, $D(y \mid z)$, through decomposing the variable set, $x = \{y, z\}$. If viewed from the joint distribution $D(x)$ underlying (8.12), this conditional postulate implied discarding the marginal distribution $D(z)$ from the factorization of $D(x) = D(y \mid z)D(z)$. Under the time-series approach, the conditional postulate was commonly approximated by models of the autoregressive distributed lag (ADL) class:

$$y_t = \sum_{i=0}^{n} \Gamma_i z_{t-i} + \sum_{j=1}^{n} \Phi_j y_{t-j} + \varepsilon_t \,. \tag{8.14}$$

The error term, ε_t, in (8.14) was regarded as a model-derived mean-innovation process, the same as the first side of the VAR dual interpretation. But in contrast to the second side, ε_t was considered as 'a compound of many reduction losses' relative to the information available during model specification and therefore 'cannot sustain claims to be a "demand" shock or a "monetary innovation"' in the sense that it could not be identified as an autonomous economic entity (Hendry, 1995: 61 and 359). It should be noted that ε_t was acknowledged to include the errors not only from omitting $D(z)$ but also from omitting other potential variables from the initial variable selection of the set x. However, these errors should be insignificant if the theoretical postulate was to be empirically viable.

Starting from (8.14), an LSE modeller would search for, or design, a statistically optimal ADL with the minimum number of lags, n, by minimization of the variance ε_t. Once an optimal ADL was found, it was reparameterized into an EC model mainly for interpretation purposes. For example, a first-order (i.e. $n = 1$) ADL of (8.14) could be transformed into:

$$\Delta y_t = \Gamma_0 \Delta z_t + (\Phi_1 - 1)(y - Kz)_{t-1} + \varepsilon_t \quad \text{where } K = \frac{\Gamma_0 + \Gamma_1}{1 - \Phi_1}. \tag{8.15}$$

As described in Section 5.3, parameter K in (8.15) was regarded as the key parameter of interest, since it embodied the long-run coefficient of the static, equilibrium condition:

$$y = Kz. \tag{8.16}$$

Comparing (8.16) with the EC term in (8.15), that term could be interpreted as an error or long-run disequilibrium variable: $u_{t-1} = (y - Kz)_{t-1}$. In other words, Equation (8.15) could be rewritten as:

$$\Delta y_t = \Gamma_0 \Delta z_t + (\Phi_1 - 1)u_{t-1} + \varepsilon_t. \tag{8.15$'$}$$

Notice that (8.15′) can be interpreted as modelling the dynamics of y_t in terms of its short-run Δy_t by three types of shocks: short-run shocks from exogenous variables Δz_t, the lagged long-run disequilibrium shocks u_{t-1}, and the innovative error ε_t, with only the first two potentially having a structural interpretation. However, if one follows the traditional approach, the interpretation is confined to K alone, that is, to the a priori postulated parameter in theoretical model (8.16), albeit its estimate is now redefined by the long-run relationship $K = \frac{\Gamma_0 + \Gamma_1}{1 - \Phi_1}$. Under this narrow interpretation, the corresponding residuals will generally be autocorrelated, since (8.15′) reveals:

$$u_{t-1} = (\Phi_1 - 1)^{-1}(\Delta y_t - \Gamma_0 \Delta z_t - \varepsilon_t). \tag{8.17}$$

Equation (8.17) shows that even though $\{\varepsilon_t\}$ are serially independent by construction through (8.14), $\{u_t\}$ are not. This interpretation provides a certain rationalization to the conventional practice of estimating a priori structural models in a simple static form allowing for residual autocorrelation. It also implies that it may not always be appropriate to select an estimator for a structural model like (8.16) by imposing the classical IID conditions on its tacked error term, u_t. Notice, however, that (8.17) also defines a general function for the autocorrelation of u_t derived from the ADL, making any other ad hoc imposition of the autocorrelation forms on u_t testable.

If one goes beyond the traditional approach bounded by (8.16), all the estimated parameters in (8.15) can be interpreted as structural. This is a much broader interpretation in that it not only considers the short-run exogenous shocks Δz_t as structural but also allows implicitly for K to be data-instigated rather than a priori given (this latter case is described as 'error correction' instead of 'equilibrium correction' in Hendry, 1995). Notice that the innovation property of ε_t becomes one prerequisite of this interpretation. But more interestingly, this interpretation indicates a possible way out of the difficulties inherent in Frisch's attempt to classify random shocks according to their dynamic roles in structural models. Following Frisch, we can transform (8.15) into a type of final form:

$$y_t = y_0 + \Gamma_0 \sum_{j=0}^{t-1} \Delta z_{t-j} + (\Phi_1 - 1)\sum_{i=1}^{t} u_{t-i} + \sum_{j=0}^{t-1} \varepsilon_{t-j}. \tag{8.18}$$

Equation (8.18) shows that we can now decompose the input shocks into a set of model-derived, non-structural, innovational shocks $\{\varepsilon_t\}$, and another set of structural shocks which further divide into the short-run exogenous shocks $\{\Delta z_t\}$ and the long-run disequilibrium shocks $\{u_{t-1}\}$. This decomposition has two advantages over Frisch's (1938) distinction between 'aberrations' and 'stimuli'. From the economic standpoint, it enables us to see the distinct dynamic impacts of the short-run and long-run shocks after filtering out the non-structural shocks designated as innovational and model-dependent, and to conceive the long-run equilibrium path as a latent structure imposing a negative feedback on the dependent variable. From the econometric viewpoint on the other hand, we need only require weak stationarity of the structural shocks, that is, they are not necessarily free of autocorrelation. Furthermore, the separation of the short-run Δz_t from the long-run relation $(y - Kz)$ reduces the likelihood of high collinearity among the parameter estimates (see also Section 7.4). Note also that if we compare the EC reparameterization (8.15) of an ADL model (8.14) with the MA representation (8.13) of a VAR model (8.12) in the previous section, we find a close resemblance except that, in the first case, ε_t, (the innovation residuals) has not been granted a structural shock status as it has in the second. The comparison shows clearly that error terms are by no means the only and unique entity to represent the idea of structural shocks for the purpose of simulating their effects.[5]

8.6 History in retrospect

The previous sections provide us with an interesting perspective—pitfalls of the traditional modelling approach and reformative alternatives can all be reflected by various ways to manoeuvre the unobservable error terms. Since these unobservables represent gaps between theoretical postulates and empirical models, as well as between models and reality, error term manoeuvres arise naturally from attempts to narrow these gaps. To a large extent, discontentment with the traditional approach derives from the observation that the gap between a priori postulated models and the relevant data information is simply too large to be filled by error terms bounded by IID conditions.

The Bayesian approach pioneered by Leamer perceives the problem mainly as one of classical statistics for failing to generate the 'best' adaptable estimators. The Bayesian prescription effectively tries to narrow the gap by proposing more flexible estimators than the classical approach allows, together with explicitly freeing the error terms from the IID straitjacket. The remedy maintains the traditional position of confining the econometricians' duty to estimation of a priori postulated structural parameters, but at the expense of exposing how arbitrary or fragile the estimates could be without model re-specification.

The VAR approach, on the other hand, adopts the Box–Jenkins time-series methodology to allow room for data-instigated model specification. The position of IID error terms is retained and strengthened to innovational shocks in the process, but at the expense of abandoning the position of taking all structural parameters as a priori given. This expense, however, has caused hardly any anxiety to macroeconomists of the RE breed, since they

[5] For more arguments on this point, see e.g. Hendry (2002).

are too preoccupied with experimenting with latent shock variables in macroeconomic models. The shock interpretation of the MAR-based impulse response analysis fits their need timely. It serves as an empirical gloss over those newly invented RE models and makes them appear correct and complete. Unfortunately, data information offers no guarantee of rendering to any VAR models innovational error terms which will also generate economically feasible shock stories. The dual interpretation of the error terms—being simultaneously 'aberrations' and 'stimuli' using Frisch's terminology—simply demands too great a faith in models. A structural interpretation of equation errors presupposes that they are autonomous, that is, that they have the same status as other variables in structural models. But in that case, there is no general argument that these shocks must surely exhibit those statistical properties embodied in the classical assumptions. Nor is there any probability left for those structural models to be incorrect or incomplete.

In contrast, the LSE approach takes a more moderate position by seeking to specify and estimate data-permissible models with both economically interpretable structural parameters and astructural IID error terms. The approach can thus be seen as a middle course between a traditional structural model approach and the VAR approach. Under the LSE approach, structural shocks or 'stimuli' are identified within a well-specified dynamic model, that is, after an innovational error term is filtered out as the model residuals. The identification is realized through reparameterization of the model into an EC type, where explanatory variables are turned into shocks which are economically interpretable and can be statistically autocorrelated. But the approach has met with resistance from many economic theorists because the reparameterization effectively denies the usefulness of postulating a priori tightly formulated models. The validity of a full structural interpretation of EC models, such as (8.15), is thus disputed not only for its lack of uniqueness but also for its data sample contingency, especially with respect to those short-run shocks.

In retrospect, disparate and sometimes confusing interpretations of the error terms essentially reflect various and changing views of models and especially of empirical model building. Many of the reformative ideas were driven by a desire for statistically better performing models to enhance the CC legacy. As a result, innovational residuals have become the benchmark for macro-econometric model searches and data-instigated model selection procedures have gained increasing popularity. Such practice has encouraged an instrumentalist viewpoint of modelling (e.g. Rissanen, 1987, 1989).[6]

[6] Rissanen proposes viewing models as an algorithmic encoding of data, which amounts to the imposition of constraints on the data. Following the theory of algorithmic complexity, he develops a theory of stochastic complexity to capture the essence of modelling as searching for the minimum description length (MDL) or the shortest code length with respect to data information.

The viewpoint abandons the quest for models with 'true' or inherently meaningful parameters and regards, instead, any statistical models as mathematical descriptions of certain regularities in data. The modellers' task is thus reduced to searching for the most compact models with innovational error terms using available information.[7] In this instrumentalist approach, error terms may be seen as playing the role of demarcating, from given data information, what is unknowable from what is knowable, the demarcation being achieved by requiring the unknowable errors to be innovations with respect to the knowable part (i.e. the model itself). Any more extended interpretation of the error terms would appear redundant. But although a completely instrumentalist view of modelling is sustainable so long as the econometrician is concerned only with representation and prediction, policy intervention requires valid external reference, and this could make the thoroughgoing empiricism underlying the instrumentalist stand unattractive to theoretical economists. Nevertheless, the insight from examining the history via the error terms should help us better understand the complexity of model building and avoid the pitfalls of unconditionally taking models as reality, or as the correct and complete representation of reality.

[7] For more discussions in the econometric community, see e.g. Phillips (1996), Phillips and Ploberger (1996), Phillips and McFarland (1997), Chao and Chiao (1998), and Reschenhofer (1999).

9

Calibration of Model Selection and Design Procedure

To a great extent, confusions and misinterpretations of the error terms described in the last chapter are reflective of difficulties in model specification, selection, and design. Likewise, those problematic aspects of parameter estimation, as described in Chapter 7, are effectively symptoms of inadequate model specification and design. In comparison to the estimation of structural parameters, however, research in model selection and design has been disparate and slow to develop as well as to become widely accepted in practice, somewhat similar to the gap between the econometric modelling of the Phillips curve and econometric studies of business cycles, as shown in Chapters 5 and 6. This relatively slow development is probably due mainly to the consolidation of the CC paradigm. In the paradigm, the job of formulating testable theoretical models is delegated to academic economists and the central task of econometricians is to provide statistically best estimators for the unknown parameters of those models; the task frequently entails a bridging chore—model specification—basically to attach to those models a set of error terms assumed to satisfy certain probabilistic assumptions which are statistically desirable for the estimation. It took several cohorts of econometricians and applied economists to fully realize and openly admit that such a delegation of jobs was but an illusion in practice, and to become actively involved in sharing the task of model formulation, selection, and design. This is perhaps the most salient feature which has distinguished the three reformative enterprises described earlier in Chapters 2, 3, and 4.

Although a great deal of the earlier description and discussion has been devoted or related to issues pertinent to model specification, selection, and design, the intricacy of the topic makes it worthy of a separate treatment, especially with respect to the topics discussed in Chapters 7 and 8. Hence, the present chapter tracks the uneven research path in model specification, selection, and design, and shows how the issue moved from a marginal concern to a central position during the reformative years. It starts from a summary description of the sparse studies on the issue during the consolidation years (Section 9.1). The subsequent section moves on to the rising focus on model evaluation and on developing specification tests mainly during the 1970s–1980s. Section 9.3 discusses and compares the model selection and design procedures proposed by the three reformative approaches described in Chapters 2–4. The subsequent development up to the early 1990s forms the topic of Section 9.4. The last section concludes with methodological reflections of the history.

9.1 Model specification and selection in the consolidation years

As described above, econometricians have been obliged to impose certain stochastic assumptions on the error terms in model specification as an essential prerequisite for the design of best estimators for certain given structural parameters. The most commonly used assumptions in regression models are known as the classical assumptions for regression analysis, that is assuming that the error terms are identically, independently, and normally distributed. To a large extent, the CC programme can be seen as setting a new standard in how to modify those assumptions in line with the economic properties of given structural models and to resolve the consequent estimator design problems. The particular property in need of modification then was simultaneity, which was translated, in the model specification step, with the assumption that all the error terms in an SEM were jointly distributed. This modification gave rise to a correlation between the error term and those endogenous regressors in each individual equation of the SEM, a problem which makes the OLS estimator lose its statistical optimality.

Empirical experience from macro-econometric model building subsequent to the CC programme soon revealed that the simultaneity-derived estimation problem was actually mostly insubstantial in practice (e.g. Christ, 1952b, 1960; Waugh, 1961; Epstein, 1989). Meanwhile, however, two other acute and recurring problems surfaced. The first was significant residual serial correlation from models estimated using time-series data. This problem would violate the classical assumption of independence. The second was larger than expected out-of-sample residual values, which mirrored model forecasting failures. The failure implied a rejection of the identical assumption in model specification. Solutions to the first problem led to latent revisions of the a priori model concerned into a COMFAC type of model, as already described in Sections 4.2 and 7.4, while the second problem was often dealt with by explicit model revisions through data-driven experiments, and two early examples can be found in Christ (1952b) and Klein and Goldberger (1955).

The search for solutions to both problems effectively extended the realm of model specification from simply assuming stochastic properties of the error terms to model re-specification, namely modification of the initial model formulation so as to have error terms of the modified model fulfilling the assumptions. In so doing, applied modellers were inevitably confronted with and involved in the vexing task of model selection.

In fact, applied modellers had rarely been given structural models which would not require any modifications when being fitted with data. Since model selection was formally assumed to be beyond the duty of econometricians, the topic was excluded from econometrics textbooks, except perhaps for limited and covert discussion under the disguise of alternative specification conditions to the classical assumptions. Consequently, applied modellers were left to commit 'sins' (see Leamer, 1978a: Preface) routinely in carrying out the task of model selection and modification in an ad hoc manner. Econometrics has abounded with such sinners since its very early days. It may be particularly interesting to look at some of the examples described in Chapter 5, where even the most theoretically minded practitioners, Lucas and Rapping, were involved in ad hoc data-driven experiments after their elaborate derivation of the inverse Phillips curve model based on a priori labour market equilibrium theory. The situation then was probably best summarized by Howrey et al. (1974):

> Economic theory and knowledge of the world around us can suggest a list of variables and some general parametric relationships. It cannot, however, set out in advance what the precise lag structure is nor all the departures from linearity. That is a matter of sample experimentation and not purely a matter of specification. (Howrey et al., 1974: 379)

It should be noted, however, that for many applied modellers the connotation of specification was already much wider than what the term implied in the quote from Howrey et al., which was basically derived from the CC paradigm. For example, specification was defined 'as the process of deciding on the hypothetical structure of a model, preparatory to testing this with observed data and finally measuring its parameters', and comprised two major steps—choosing variables and functional forms for each equation in a multi-equation model context (Brown, 1970: 49).

During the consolidation years, arguably the most significant contribution to the widening of the connotation through active model construction was due to Henri Theil, mainly through his pioneering work on specification analysis (1957 and 1961: ch. 6). From a statistical standpoint, Theil interpreted the structural model approach as virtually assuming the a priori model given to applied modellers as the 'maintained hypothesis', that is, the hypothesis which was not meant to be subject to test. He seriously questioned the practical validity of such an assumption as it greatly overestimated 'the economic theorist's knowledge and intellectual power' (Theil, 1958: 206). In his view, applied modellers should treat theoretical models as testable hypotheses and therefore be allowed to experiment with various hypotheses using data information in model selection. Specification analysis thus formed a cornerstone in his attempt to systematize the ad hoc practice in applied model selection and design. In his 1958 book, Theil spelt out three key principles for model selection: (a) predictive power, which was 'the' criterion for a good econometric model, (b) plausibility, which could help raise the probability of forecasting success and which would primarily rely on the support of economic theory, and (c) simplicity, a criterion somewhat contradictory to 'plausibility' but one which was crucial to keep the selected models useful for practical purposes (Theil 1958: section 6.2). These principles led to Theil's setting of applied model selection and design on a two-fold target. Economically, all estimates of the parameters of a selected model must be plausible and, statistically, the residuals must not only satisfy the desired stochastic assumptions but also have the smallest residual variance with respect to other possible model variants. To achieve this target, extensive specification analyses came into play. These analyses were to iron out a number of common specification errors in regression models, such as residual serial correlation, correlation between the regressors and the error terms, and severe collinearity among the regressors on the one hand and omitted variable bias on the other. Concrete tests and diagnostic procedures were recommended for each type of specification error, and possible ways of modifying models were introduced with respect to the available econometric methods of the time. In particular, Theil suggested an adjustment of the conventional goodness of fit statistic, R^2, by the number of regressors so that the revised statistic could be used as a model selection measure to incorporate the simplicity criterion (Theil 1961: 211–14). This measure of Theil's was a precursor to the information criteria proposed by statisticians over a decade later, such as the Akaike information criterion (AIC) (1973) and the Schwarz criterion (1978), and adapted into econometrics as measures of model parsimony towards the end of 1970s (e.g. Sawa, 1978; Zellner, 1978; and also Amemiya, 1980).

Before then, however, textbook econometrics had remained largely indifferent to Theil's approach to model selection and specification analyses. Apart from the rise to dominance of the CC structural approach, the indifference probably came from the negative image of 'data mining', which had lingered on from Friedman's sharp criticism of Tinbergen's (1939) model and subsequently Klein's (1950) model. In Friedman's view, one should not be taken in by the 'goodness of fit' statistics of those empirical models because they were built and selected for yielding 'high coefficients of correlation' 'after an extensive process of trial and error' (Friedman 1940). In other words, 'the fact that the equations fit the data from which they were derived is a test primarily of the skill and patience of the analyst; it is not a test of the validity of the equations for any broader body of data'. The only appropriate test of the validity of a model is 'the adequacy with which it predicts data not used in deriving it' (Christ, 1952b: Friedman's comment, pp. 108–9).

9.2 Data-based model evaluation and the rise of specification tests

It has already been described in the previous chapters how the severe macro-econometric model forecasting failures of the 1973 oil-crisis-led recessions was a major impetus for the rise of the post-1970 reformative movements and that the movements led to a certain shift of position from the CC tradition towards an approach grounded more on statistical theories. The shift was well reflected in the growing research attention and emphasis on model specification, selection, and design issues. The growth brought about a revival of Theil's approach to model selection and specification analyses.

A major herald of the revival was a joint work by Phoebus Dhrymes, Philip Howrey, Saul Hymans, Jan Kmanta, Edward Leamer, Richard Quandt, James Ramsey, Harold Shapiro, and Victor Zarnowitz (1972), which aimed towards developing a framework for the systematic evaluation of macro-econometric models. The work resulted from a project for the Seminar on Criteria for the Evaluation of Econometric Models of the NBER-sponsored Conference on Econometrics and Mathematical Economics, a programme which commenced in 1970. It should be noted that 'model evaluation' was effectively a misnomer for model selection and design in Dhrymes et al. (1972), in which the operational procedures that they proposed were mostly statistical-test based model specification analyses. In other words, 'evaluation' in their discussion was regarded as omnipresent and necessary during the model selection process, a viewpoint which had clearly forsaken the textbook assumption of taking structural models as a priori given and fixed. Moreover, they acknowledged that there was no uniquely appropriate way to carry out such a model evaluation process. In order to make the task manageable, Dhrymes et al. classified the issues pertinent to model evaluation into two categories: 'parametric' versus 'non-parametric' evaluation: 'An evaluation procedure is said to be parametric if it relies on a formal statistical test based on the stochastic specification assumed to apply to the econometric model. Non-parametric evaluation is concerned with specialized and descriptive procedures' (Dhrymes et al., 1972: 293). Here, it is interesting to see that parametric relationships defined by economic theory are used as the fundamental divide in the classification. The classification allocates to 'parametric evaluation' almost all the apparently general issues relating to econometric tests and the verification of plausible economic hypotheses. However, it segregates issues concerning the statistical properties of the residual error terms from those general issues and groups them into 'non-parametric evaluation', alongside other practical issues, such as measurement issues concerning the chosen individual variables of interest and the suitability of the model to specific purposes as well as the time-series properties of individual variables. Such a classification could significantly under-utilize the diagnostic function of the statistical properties of the residual errors during the model evaluation process. Noticeably, forecasting tests were assigned to the second division under the heading of 'tracking measures'.

In the first division, parametric evaluation was further divided into two steps—model construction and post-model evaluation via prediction. The latter was somewhat awkwardly separated from those 'tracking measures' of the second division. Dhrymes et al. (1972) devoted most of their discussion to the first step, particularly the issue of how to choose the appropriate functional forms, a key prerequisite for parametric estimation. Since 'economic theory gives preciously few clues' here, statistical techniques were resorted to for the choice. It was in this context that the then recently developed Ramsey's (1969) test was recommended. Widely known nowadays as the RESET test, Ramsey's test helps provide diagnoses of the appropriateness of a linear regression model against a non-linear one by testing whether non-linear combinations of the explanatory variables have any significant power in explaining the modelled variable. Dhrymes et al. further related the conceptualization of Ramsey's test to the general issue of discriminating between multiple possible models, and brought in the statistical classification of 'nested hypotheses' (for cases when one hypothesis could be a special case of another more general one) versus 'non-nested hypotheses' (for the case when neither of two hypotheses is a special case of the other) for the analysis. Specifically, they introduced a test procedure for non-nested hypotheses proposed by Cox (1961, 1962) through the generalization of the likelihood ratio test. Interestingly, the similarity of the Cox procedure to the Bayesian approach to the problem of model selection was discussed in terms of the likelihood ratio, but the discussion left out the close relation of the ratio to the relative magnitudes of the residual error terms and hence the idea of selecting a model by the relative smallness of its residual errors.

A fuller adoption of the Cox procedure into econometrics was proposed by M. H. Pesaran (1974). Apparently unaware of the joint work of Dhrymes et al. (1972), as it was absent from the reference list of the paper, Pesaran saw the Cox procedure as a promising way to improve Theil's adjusted R^2, especially in the case of non-nested model selection. Based on Cox's procedure, Pesaran proposed a specification test to assist model choice between two non-nested linear single-equation models. The test procedure was subsequently extended to the case of non-nested multivariate nonlinear models, see Pesaran and Deaton (1978). Other emulative works on the topic flourished around the turn of 1980, as illustrated by the special issue on 'Non-Nested Models' in *Journal of Econometrics* in 1983 and also by the survey papers by McAleer (1987) and MacKinnon (1992). Most of these works were focused on technical issues such as how to design computationally simple tests, how to improve the finite-sample properties of a test, and how to adapt such tests for models of special data types such as models of limited dependent and qualitative variables.

In fact, the thriving research in devising tests for non-nested models was rooted in an environment of expanding interest in tackling various specification issues through designing appropriate specification tests, as summarized by Hendry's (1980) 'three golden rules of econometrics'—'test, test, test'. It was in this environment that Hausman's (1978) specification test received wide attention and was soon branded 'the Hausman test', even though the same test procedure—to test the validity of conditioning a modelled variable upon a set of explanatory variables in a regression—had been proposed by J. Durbin (1954) over two decades before.[1] However, the majority of model specification tests were developed for the purpose of diagnosing possible mis-specifications through violations of the classical assumptions on the residual errors. These included a generalized approach proposed by Godfrey (1978) for testing residual serial correlations, tests of residual heteroscedasticity devised by White (1980) and independently by Nicholls and Pagan (1983), and a test primarily for checking whether the residual errors were normally distributed (see Jarque and Bera, 1980).

While applications of those newly devised specification tests were on the increase in the empirical literature, there was still a great deal of disarray in practice when it came to applied model selection and design. The growing supply of diagnostic tests had indeed greatly facilitated the exposure of commonly occurring model mis-specification problems and enhanced awareness of those problems among applied modellers. But, methodologically, the tests provided no conclusive solutions as to what the correct model form and specification should be once the problems were exposed, since the causes of the problems were seldom directly exposed by the tests. Moreover, multiple failures uncovered by different diagnostic tests might not necessarily indicate more than one cause of those failures. For example, concurring rejections of the classical assumptions such as homoscedasticity and normality could well be the result of the model under test being unable to explain certain unexpected data volatility of the modelled variable in a small part of the data sample used. In other words, what the test statistics directly revealed was

[1] I am grateful to David Hendry for this historical fact.

the symptoms rather than the causes of the mis-specifications. The causes remained to be diagnosed by the modellers when the models which they initially fitted failed one or more specification tests. The causes that they had diagnosed would play an important role in their decisions about how to proceed with modifying and improving the initial models. Such decisions were also found to be intimately related to the criteria and the purposes underlying the model specification search, selection, and design process.

9.3 Major alternatives to specification searches and model selection

It was in the milieu of thriving research into data-based model testing that the three reformative enterprises described in Chapters 2, 3, and 4 arose and distinguished themselves by explicitly offering alternative procedures to resolve the conundrum of the choice of model left aside by the CC group. Let us review and compare these major reformative approaches here, mainly from the perspective of the model selection criteria and purposes underlying these procedures.

In Leamer's (1978) *Specification Searches*, the Bayesian approach was prescribed as a fundamentally effective treatment of the model selection problem. The prescription was based on the following presupposition inherent in the CC tradition: measurement of economic theory was the principal driving force for applied model building and the measurement task amounted to obtaining the statistically best estimates of the a priori postulated structural parameters of interest. Accordingly, the Bayesian priors were seen as the most convenient means of representing the a priori nature of these theory-driven parameters and of exposing the subjectivity of modellers' decisions during model selection and specification searches. The traditional research focus on the estimation of structural parameters was technically undiluted. Under the maintained absolute priority of the economic theory at hand, any additional data-based information from statistics relating to the residual error terms, and from trial additions of variables during specification searches should be regarded as trivial or auxiliary in the sense that the information was considered useful as far as it would help improve the estimates of the parameters of interest. In cases where post-data model evaluation indicated that the estimates were statistically too fragile, it would ultimately be up to the particular needs and requirements of the economists concerned to decide how the theoretical model should be revised and improved.

Leamer's position of retaining the dominance of economic theory is illustrated quite clearly in his schematic diagram of model inference shown in Figure 2.1 of Chapter 2. Furthermore, the position could also be seen from his emphasis on weaving bits of the relevant data evidence together for the purpose of coming up with a 'plausible story' in his analogy between 'Sherlock Holmes inference' and his Bayesian specification searches (Leamer, 1974; the analogy also appeared in Dhrymes et al., 1972). The resort to reaching one particular 'plausible story' implies that it is beyond the modellers' task to come up with a model which could fully explain all the features of the available data. In Leamer's view, available samples of economic data were usually too weak to supply any sustainable stories without prior information (Leamer, 1978a). Technically, the Bayesian approach concentrated researchers' attention on the probabilistic properties of the non-trivial parameter space, which was derived from the a priori defined variable space by a given theoretical model. Model selection issues which would entail modification of the variable space should therefore be left open to economic considerations. Further developments of the Bayesian approach, especially those utilizing Bayesian decision theory, were mainly concerned with the reformulation of various model selection criteria and tests based on the classical statistic theory into comparable Bayesian ones (e.g. Kadane and Dickey, 1980, and Klein and Brown, 1984).

In contrast, the VAR approach explicitly denied the capacity of economists to provide a priori correct and tightly formulated parametric structural models. The VAR was thus regarded as the most general form of econometric model specification for macroeconomic models using time-series data. In spite of their atheoretical appearance, VAR models adhered to the fundamental principle of general equilibrium in economics by maintaining the CC tradition of mapping economic theories onto a general system of interdependent equations. However, by starting the modelling procedure from a general VAR, the approach effectively upgraded, as one important model selection criterion, the desired statistical properties of the residual error terms—that they should be innovative and white-noise processes. This criterion was believed to be directly advantageous for raising the predictive power of empirical models. But when the performance of VARs used for forecasting purposes turned out to fall much below that of the VARs built for policy debates, those residual error terms acquired a new interpretation as unexpected structural shock variables. As described in Chapters 3 and 8, the interpretation stemmed from the rise of RE models and gained its popularity via the operational mapping of the VAR impulse–response analysis to the impulse–propagation dynamics of a set of interdependent variables emphatically examined in RE models. It was for this type of macroeconomic model that the VAR approach shifted econometricians' conventional task from parameter estimation to latent shock simulation. Consequently, the need for the careful evaluation and interpretation of individual structural parameter estimates was heeded much less than in Leamer's approach or the CC tradition.

On the other hand, the VAR approach agreed with Leamer's Bayesian approach in maintaining the position that the choice of variables in model formulation and selection should be left to economic theorists, especially for VAR models which were used for policy analyses. In those VAR models, the a priori given chosen variables were virtually treated as the maintained hypothesis because there were usually fewer prior requirements on the individual parameters linking the chosen variables than on those desired statistical properties of the error terms, and also because the latter properties were likely to be met with a generous inclusion of the lags. Although an abstract VAR had the potential to include as many variables as desired, empirical VAR models were frequently severely constrained by the lack of degrees of freedom due to available finite data samples. Hence, in practice, most of the VAR models contained fewer than ten variables, making their correspondence to a general-equilibrium scheme somewhat far-fetched.

The model selection and design procedure of the LSE approach was arguably the most elaborate of the three methodological alternatives. Its development could be summarized into three crucial stages. The first was the introduction of Anderson's (1971) sequential testing procedure, in which all the testable hypotheses were arranged in an order of increasing restrictiveness so as to achieve the uniformly most powerful results. As demonstrated by Mizon (1977a, 1977b) in a single-equation model context, Anderson's procedure entailed a general-to-specific modelling approach, which enabled modellers to conduct specification searches and select models according to specification test results in a rigorous statistical manner. The second stage came with the three principles for applied model research put forward by Davidson et al. (1978), as described in Section 4.2 of this book. The principles effectively evoked a set of composite model selection and design criteria which not only covered Theil's three principles of predictive power, plausibility, and simplicity, and made their connotations richer and more meticulous, but also included a requirement on any new models to have the capacity of outperforming the existing ones, a principle which later became known as 'encompassing'. The criteria essentially raised the goal of empirical model building above the conventional yardsticks of theoretical relevance and precision in terms of parameter estimates to the level of having the overall model capacity of explaining all the salient data features better than other available model results. In fact, the overall model capacity was shown to be the essential prerequisite of theoretical relevance and precision of parameter estimates. Note that a priori economic theory was designated to a more limited but also more specific role than the traditional approach—providing mainly hypotheses about long-run equilibrium relationships. It became the task of applied modellers, during the general-to-specific dynamic specification searches, to supplement the theory with data-instigated variables to capture the salient data features, especially those of the short-run type. The searches also involved reparameterization in order to make the model economically interpretable at the individual parameter level and facilitate verification of the relative constancy of the parameter estimates. Failure to meet any of the multiple specification criteria during the searches could lead to a complete overhaul of the whole model selection process. The overhaul usually implied widening the choice of variables to capture idiosyncratic features of the data samples, which were not originally considered in the theory. The widened variable coverage entailed another round of general-to-specific dynamic specification searches, making the searches an iterative and progressive process (see also Chong and Hendry, 1986). As described in Section 4.3, the LSE strategy of model specification, selection, and design was eventually summarized into a reduction theory (see Hendry, 1995: ch. 9).

In view of the previous two approaches, the LSE modelling strategy is probably the most comprehensive in having the strategy built under relatively the widest set of model selection criteria. It is certainly the most explicit in actively promoting the need for applied modellers to take model design as their primary task. Similar to Leamer's Bayesian approach, the LSE strategy maintains the 'structural' principle of the CC tradition through the estimation of the a priori postulated parameters of interest. But unlike Leamer's approach, the LSE approach explicitly conditions such estimation upon the attainment of a well-specified model first and interprets those parameters mostly as the long-run equilibrium parameters. To facilitate the attainment, a general-to-specific route is adopted in the LSE procedure, similar to the VAR approach. However unlike the VAR approach, applied modellers following the LSE approach are expected to be more innovative in variable choice and the innovation is heavily dependent on exploring sample data and related background information. In particular, the choice covers not only the lags of those theory-nominated variables but also other vernacular variables which best embody idiosyncratic features of the data samples, as shown from the examples described in Section 4.4. Moreover, the choice includes transforming existing variables to assist reparameterization, that is to minimize the degree of collinearity through active parameter design. In that respect, the usefulness of theoretical models is marginalized, especially those dynamically elaborate models. That may explain why the LSE strategy has met strong resistance from modellers with a staunch conviction of the value of a theory-based modelling approach.

9.4 Diversified approaches and formalization of selection procedures

Although controversial, the three alternative modelling approaches helped greatly in raising econometricians' awareness of the substantive gap between economic theories and empirical models, and particularly of the high degree of dependency of parameter estimates on model specification in the sense that any accuracy gained in the estimates from specification improvement was generally significantly larger than that which an elaborate choice of estimators could deliver. The post-1980 period saw increasingly vibrant research into topics pertinent to model selection and specification searches and, with that research, more diversified viewpoints.

Methodologically, one hot and controversial topic was how much data information should be allowed in the empirical model selection and specification as opposed to a priori theoretical information (e.g. Pesaran, 1988). Among those more theoretically minded economists, the conviction that econometrics was useful only for confirmative purposes was so strong that any endeavours to reorient econometrics towards data-exploratory purposes was met with serious resistance. In particular, the VAR approach—as well as the LSE approach—were objected to for their heavily pro-data-driven strategy in exploring and selecting causal relations, which were traditionally assumed to be the territory of theoretical derivation (e.g. Miller, 1987). The rise of the RBC and the related DSGE modelling approaches, as described in Chapter 6, could be seen as the most extreme reaction to any of those modelling strategies which allowed data any explicitly active role in formulating or selecting causal relationships.

Within the econometric research circle, scepticism to those pro-data-driven modelling strategies was raised, particularly by M. Lovell's (1983) investigation into the consequences of data-mining activities by means of simulation exercises. Methodologically, Lovell's demonstration and viewpoint were not new. Friedman made the point in his criticisms of Tinbergen's model over four decades before, as mentioned in 9.1; and it was also reiterated by Leamer (1974). Technically, Lovell's demonstration accentuated the problem of over reliance on the 'goodness of fit' test statistics as the key model selection criterion, since those statistics had not been discounted because of the double-counting problem. In that respect, his demonstration widened the awareness of the problem of 'pre-test estimation', a problem concerning the design of estimators which were conditioned upon statistics from hypothesis testing used for the model selection purposes prior to the estimation of the selected model (e.g. Wallace, 1977; also Judge and Bock, 1978). For many applied modellers, the discussion was simply perceived as a taboo against being actively engaged in 'sinful' data-mining activities. But for theoretical econometricians, the discussion indicated new research possibilities—how to devise estimators which were explicitly conditioned on those pre-test statistics so that the double-counting effect could be reflected in the sampling distributions of the parameter estimators of the model eventually selected (e.g. Pötscher, 1991; Kabaila, 1995). Interestingly, data-mining activities were actually taken for granted in those pre-test researches, as they formed the very prerequisite of the specifically devised post-model-selection estimators. On the other hand, the researches maintained the CC tradition in that the focus was kept on devising estimators so as to concentrate and transpose model selection issues onto parameter estimation issues.

Meanwhile, those pro-data-driven modelling strategies also reoriented the attention of theoretical econometricians from devising more elaborate specification tests towards formalizing those newly proposed model selection procedures (e.g. Granger et al., 1995). For example, White (1990) formalized the general-to-specific specification procedure of the LSE approach by a theorem stating that such a model search procedure will converge on the true model asymptotically. The theorem was subsequently extended to a situation where the true model was outside the range of models considered under the specification search process, and the extended theorem demonstrated that the LSE procedure should result in a model that was closest to the true model (Sin and White, 1996). From the perspective of Bayesian inference, Phillips (1995a, 1996) formally justified the data-instigated econometric model selection procedures using time-series data by means of the asymptotic theory of Bayesian inference that he developed jointly with Ploberger (Phillips and Ploberger, 1996). The theory effectively facilitated the use of model prediction as a key model selection criterion. Phillips's formalization approach was heavily influenced by the views of Jorma Rissanen, an information theorist. Rissanen regarded models as essentially the algorithmic encoding of data and model selection to search for the shortest code length which could capture the desired data information (Rissanen, 1987, 1989; see also Section 8.6 of the previous chapter). Such a point of view opened up the possibility of conducting data-led model selection in an automated way.[2] On the other hand, it downplayed the substantive meaning of models, especially that of structural parameters.

Indeed, there was a discernible trend to grant data-based criteria the dominant status in the wave of attempts to formalize model selection procedures by mathematical statistics. Criteria such as the AIC and the Bayesian information criterion (BIC) were given a fundamental role, which effectively depreciated the role of a priori available theoretical models. Economics, the subject matter of econometric models, was hardly or merely perfunctorily considered in many of the formalization studies. Such depreciation was probably most noticeable in the erosion of the substantive connotation of structural parameters. With the advance of time-series econometrics, parameters of solely

[2] Phillips (1995b) pioneered the automation of a model selection procedure guided primarily by the BIC principle and the automation was carried out in terms of the GAUSS programme. One of the first pieces of automated econometric modelling software is PcGets developed by Hendry as an extension of PcGive (Hendry and Krolzig, 2003). However, these developments are beyond the span of the history examined here.

descriptive time-series models were treated as 'structural' (e.g. Section 6.3), and time-varying parameter models became a fashionable technique to circumvent 'structural breaks', as described in Section 7.2.[3]

Those statistical theory-based formalization studies have clearly helped to widen the divide between the econometric modelling approach and the essentially economic theory-based approaches by means of DSGE and CGE models. Unlike the latter, however, most of the formalized econometric modelling procedures have been kept as armchair strategies by the majority of applied modellers, who, while faced with a specific subject matter of concern, on the one hand, and incomplete theories, on the other, have opted to follow a pragmatic, middle-of-the-road approach in model building in that their model selection process has largely remained ad hoc, especially in their choice of mixing the relevant theoretical and data information.

9.5 History in retrospect

It is clearly evident from the above historical investigation that substantial advances in model specification and selection have been made during the reformative period. Primarily, the importance of model specification has been widely acknowledged and it has been increasingly appreciated by the econometrics community that data information has to be considered in model specification and selection and the guilty sentiment accompanying data-mining activities has largely been dispelled. Consequently, we see a widening of the scope of specification searches and with it an expansion of core econometric research from devising estimators into developing specification tests and consistent model selection and specification procedures. Models have become an increasingly more important research object than certain a priori defined parameters within a model, as model uncertainty is recognized to be more serious and fundamental than the sampling uncertainty inherent in parameter estimation.

However, what has remained unresolved is how to balance the information mixture between theory and data during the model specification, selection, and design process. Such a balance is dependent upon which model selection criteria are used and the relative weights or the sequences of priority that are allocated to each chosen criterion. The choice in turn depends ultimately on the purposes of model building. As the purposes vary a great deal, it is thus

[3] Much later, the development has led to the advocacy of abandoning the tradition of parametric models altogether in favour of nonparametric or semi-parametric models, which are selected mainly on the basis of the estimation method of general method of moments (GMM) (Hansen, 2005).

practically impossible to reach a predominant agreement on what the best mixture is and consequently what the best modelling strategy should be.

From the joint work of Dhrymes et al. in the early 1970s to those attempts at formalized model selection procedures in the early 1990s, there is a notable shift of attitude in empirical model selection—from primarily leaving the issue to the design board of theoretical economists towards actively engaging in it by a growing number of data-based criteria (see also Hendry, 2009). It is probably fair to attribute the shift to the most essential theme of the reformative movements. With respect to the CC paradigm, the shift can be seen naturally as an 'error correction' of the naïve assumption that a priori reasoning alone is capable of coming up with adequately and correctly formulated theoretical models.[4] Significantly catalysed by advances in mathematical statistics and information theories, the shift has highlighted the fact that it is indispensable to have an element of exploratory data analysis incorporated into the conventional duty of confirmative analysis in econometric modelling. This enables a better bridge between theoretical information and data information. In fact, the shift suggests quite strongly that it is rarely possible to conduct reliable and robust confirmative analyses in econometric studies unless the analyses are preceded by carefully designed exploratory data analyses.

It is natural not to expect such a suggestion to be adopted easily nor to be widely accepted by the economics profession. The task of conducting data-exploratory oriented analysis and relying heavily on it for model selection has invoked too much apparently ad hoc decision-making for the liking of many economists, especially those armed with a solid training in textbook econometrics. While an increasing amount of research has been undertaken to formalize model specification and selection procedures and to automate them computationally as much as possible, the course of this research has run into deadlock and exposed its own limitations. Almost all of the formalized procedures are fundamentally built on statistical selection criteria, whether these are based on a multiple-hypothesis testing scheme or the optimization of some kind of penalized goodness of fit criteria. The reliance on economics built upon the CC tradition has been weakened to a minimum. There is a discernible, and possibly increasing, gap between the ways that theoretical models are rigorously derived and the ways that these models are used or regarded as useful in econometric modelling. In spite of that, however, formalization has remained a commonly shared driver of research. Models under discussion from both sides are frequently treated as abstract

[4] It should be noted that the shift was also accompanied by a change of attitude among economists. They have increasingly taken into consideration data features and econometric evidence in theoretical model formulation (e.g. Eichenbaum, 1995).

entities, stripped of any idiosyncrasy, that is, the particular economic context which has supposedly motivated the modelling activities in the first place. On the econometric side, studies in formalizing modelling procedures, other than those motivated by the automation purpose, do not directly serve the concrete purposes of empirical modelling, unlike those conventional studies which devise estimators and tests. Rather, they are heavily involved in elaborating metaphysical justifications of whatever practice has already been found empirically effective under particular circumstances. It is therefore difficult for such studies to engage a large audience from the applied front. Indeed, most of the consequent studies have ended up in a limited circle of academics who are more interested in methodologies and the philosophy of social sciences than empirical modelling for purely economic purposes (e.g. Mäki, 2002; Stigum, 2003; Spanos, 2009). Unfortunately, such a route to formalization could be fundamentally flawed as statistical inference forms only part of the empirical modelling process. The process on the whole entails a synthesis of information from multiple sources, not just from statistical data or theories but also from the local circumstances from which the data and economic issues of interest are generated. The need for a fusion of knowledge from a holistic approach in empirical modelling is well beyond whatever formalized mathematical frameworks could ever deliver.

In fact, the limitation of formalization alone has been brought to open attention during discussions on model specification and selection issues (e.g. Phillips, 2003). But it remains to be seen when and how the wave of 'error corrections' of the conventional theory-based modelling approach by a data-based one during the reforms will reverse to curb the tendency of over reliance on the data-based approach and the associated statistical formalization such that the mainstream strategy of model selection and design will converge towards or oscillate around a more balanced path.

10

The Impact of the CC Programme through Citation Analysis

Most of the previous chapters are concerned with attempts to reform the CC paradigm during the two decades after 1970. The question naturally arises as to how much the CC paradigm has been weakened or abandoned by the academic community as a result. This chapter tries to provide answers to the question in terms of certain statistical measures via citation analysis. It should be noted that these answers may mirror answers to the question of how great the impact of the reforms has been. It would be desirable to conduct a direct citation analysis for the latter question, but the time is not quite ripe since it takes decades for citations to accumulate into decent sample sizes in economics.

Citation analysis has grown rapidly in recent years, thanks to fast-growing computing technology and the internet. Most existing citation studies are on the impact and research trends of science subjects (e.g. van Raan, 2004). Citation analyses of economic research are few and far between. Little has been done to assist historical studies. Actually, combining citation analyses with historical studies should be beneficial for both sides. While it is self-evident how much citation data could enhance our information about the general impact of selected works, factual historical accounts would certainly help reduce the 'unavoidable uncertainty' well-known in citation information alone. Specifically, motives for making references vary considerably, some for cronyism or authority, others refutation (e.g. Brooks, 1986; MacRoberts and MacRoberts, 1989). As far as the objects of our historical investigation are not derived from citation analyses, such uncertainty should not pose a serious problem. In fact, the uncertainty should not affect the broad reliability of citation-based impact measures in general so long as they are properly interpreted, as argued by van Raan (1998).

The rest of this chapter is organized as follows. Section 10.1 briefly describes the citation database, a series of impact measures to be used, and some key summary statistics of the database. A more detailed description of the database and the specification of the particular root document bases are given in the final section (10.5). Section 10.2 examines various impact measures of the CC paradigm to help evaluate its historical significance. To furnish the impact of the CC paradigm with a comparative perspective, a number of impact statistics from alternative bases and sources are provided and discussed in Section 10.3. Section 10.4 summarizes the main findings.

10.1 Citation database, impact measures, and key summary statistics

The citation database has been constructed on the basis of over 1300 root documents dating from 1913 up to 1969, inclusive. These root documents are considered as forming the core of the econometrics literature of the pre-1970 era. Around 33,400 citations of these root documents are collected in the database. The citations for the period 1970–2005 are extracted mainly from Web of Science; those of the pre-1970 period are mainly from JSTOR. Since JSTOR covers a narrower journal range and its citation information is less comprehensive than Web of Science, the present citation analysis is focused on the post-1970 period. Clearly, the database suffers from a bias towards journal-based documents at the expense of book-based documents, although some of the latter have been added manually. On the other hand, the database enjoys the advantage of being extensively classified. All the publication sources are categorized into three groups: academic, non-academic, and educational; all the documents are categorized by a modified JEL system (see Section 10.5 for a more detailed description of the database including the definition of the modified classification system). The size of the database and its classification enables us to examine the evolution of the citation impact of selected groups of root documents in a number of ways, such as topic transfer, diffusion, and diversification; publication sources; and their relative degrees of impact.

Most of the impact measures to be used here are derived from those described in Mann et al. (2006). In that joint paper, the main object of the measures is ‘topic’. The primary object in the present chapter is a ‘citation base’, that is, a selected group of root documents. The measures are defined mostly in a time-series format so as to facilitate the illustration of the evolving paths of the measured impact.

Let us denote the total number of documents in the database at year t by D_t, a citation base by i, and its citation count by $\Gamma_{i,t}$ for year t. The impact factor of the base is defined as:

$$I_{i,t} = \frac{\Gamma_{i,t}}{D_t}, \qquad t = 1970, \cdots, 2005. \tag{10.1}$$

Since all the documents are classified into topics using English letters (see Section 10.5), we can also obtain the impact factor of a particular topic, τ:

$$I(\tau)_{i,t} = \frac{\gamma(\tau)_{i,t}}{d(\tau)_t}, \qquad \tau = D, E, \cdots, T, U \tag{10.1a}$$

Note that $\Gamma_{i,t} \leq \sum_\tau \gamma(\tau)_{i,t}$ and $D_t \leq \sum_\tau d(\tau)_t$ because it is common for a document to have more than one topic in the JEL classification. In fact, we can exploit this feature to build a measure for the degree of width of the topic coverage:

$$\mu_{i,t} = \frac{\Gamma_{i,t} \Big/ \sum_\tau \gamma(\tau)_{i,t}}{D_t \Big/ \sum_\tau d(\tau)_t}. \tag{10.2}$$

We can then interpret $\mu_{i,t} < 1$ as base *i* being relatively effective in encouraging the succeeding research to cover a wider range of topics on average, with $\mu_{i,t} > 1$ indicating the opposite.

It may not be so straightforward to compare different citation bases by the impact factor series. For one thing, the document size and form differs between citation bases, though the difference could reflect the effect of the critical mass in research. A popular way of impact comparison is to use a single statistic measure, such as the *h*-index of Hirsch (2005) for measuring the impact of an individual scholar. We can produce a single number from (10.1) simply by taking the ratio of the total citation counts to the total document number of the whole sample period. But that ratio suffers from being too dependent on the database. Here, we choose Silagadze's (2009) *s*-index and adapt it to account for the document number of base *i*, n_i. In particular, our *s*-index is defined as:

$$s_i = \frac{1}{2}\sqrt{\frac{\sum_{t=y_0}^{y_T} \Gamma_{i,t}}{n_i} \frac{\varepsilon_i}{\ln(y_T - y_0)}}, \qquad t = y_0, \cdots, y_T \tag{10.3}$$

where $\varepsilon_i = -\sum_{t=y_0}^{y_T} p_{i,t} \ln(p_{i,t})$ is the entropy with respect to $p_{i,t} = \frac{\Gamma_{i,t}}{\sum_{t=y_0}^{y_T} \Gamma_{i,t}}$.[1] Similar to the impact factor in (10.1), we can also calculate an *s*-index for each topic within base *i*:

$$s(\tau)_i = \frac{1}{2}\sqrt{\frac{\sum_{t=y_0}^{y_T} \gamma(\tau)_{i,t}}{n_i} \frac{\varepsilon(\tau)_i}{\ln(y_T - y_0)}}, \tag{10.3a}$$

[1] Note that $\ln(y_T - y_0)$ is the maximum entropy of the citation series and that we take a natural logarithm in the entropy calculation. When the index is calculated for the period 1970–2005, the maximum entropy is $\ln(36)$.

if we want to find out a comparative ranking of the degrees of impact that various topics have received.

Notice that Silagadze's *s*-index takes into consideration the effect of the cross-time citation distribution by means of the relative entropy. In fact, another important aspect of the impact measurement is the citation distribution across different topics, known as topic diffusion and/or diversity through topic transfer. Following Mann et al. (2006), the relevant measures are based on the ratio:

$$r(\tau)_{i,t} = \frac{\gamma(\tau)_{i,t}}{\sum_{\tau} \gamma(\tau)_{i,t}}. \tag{10.4}$$

When the topic under consideration does not overlap with the topics covered by the base, $r(\tau)_{i,t}$ serves as a topic transfer indicator. We can then define the diffusion index by:

$$\delta_{i,t} = \sum_{\tau \notin \tau(\beta)} r(\tau)_{i,t}. \tag{10.5}$$

$\tau \notin \tau(\beta)$ in the above equation denotes topics not covered by the those of base *i*. Obviously, the complement of the diffusion index is a within-base impact indicator, measuring how much a topic evolves within its own field of research.

Finally, a summary measure for topical diversity is defined by the entropy with respect to $r(\tau)_{i,t}$:

$$\varepsilon_{i,t} = -\sum_{\tau} r(\tau)_{i,t} \ln\left[r(\tau)_{i,t}\right]. \tag{10.6}$$

Let us now look at some summary statistics of the database. First, the entropy of the roughly 33,400 citation documents for the period 1970–2005 is calculated to be 3.574, which is extremely close to the maximum entropy of 3.584. The closeness shows that the citations are distributed quite evenly across the 36-year period without discernible signs of decay. In other words, it indicates that the pre-1970 core econometrics literature is still well cited on the whole over three decades later, without obvious signs of impact decline. However, the distribution of the citations clearly becomes uneven when their dispersion across the topics of the JEL classification is considered, as shown by the pie chart in Figure 10.1. Remarkably, the non-economics topic (U) occupies the largest component of all the topics (see Section 10.5 for further description of the 'U' classification). Two series of topic diversity measures are also plotted in Figure 10.1: one is the ratio, $D_t \big/ \sum_{\tau} d(\tau)_t$, from equation (10.2) and the other the topic diversity index of (10.6). Both series exhibit a noticeable decline, with the first series showing a sharp drop in the

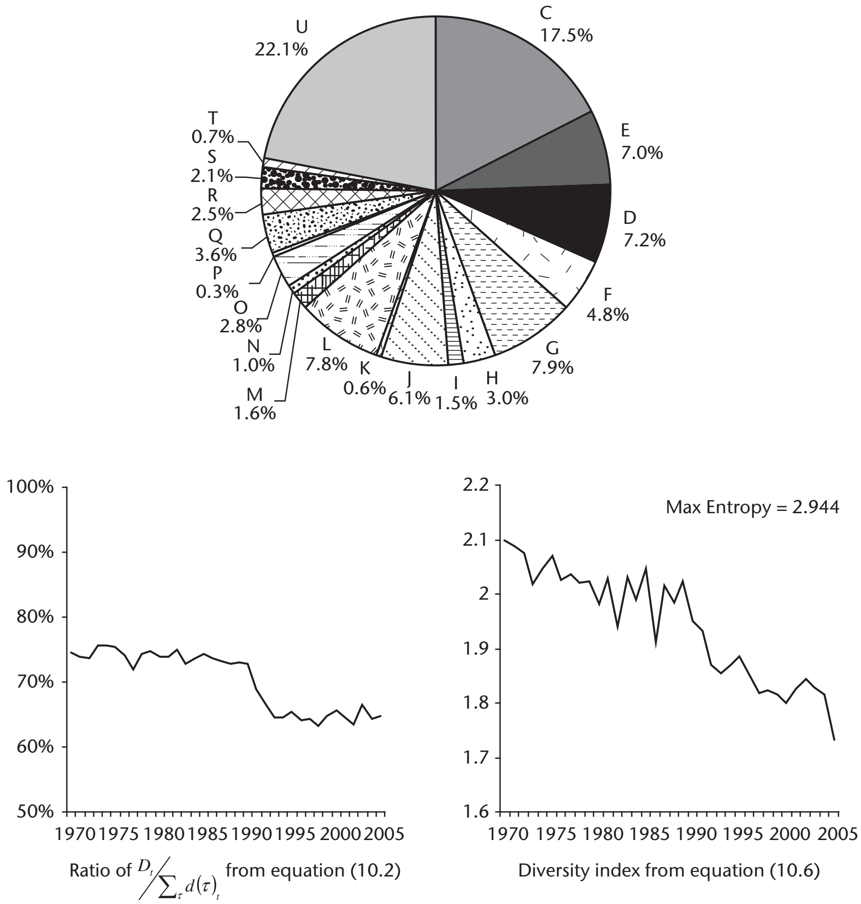

Figure 10.1. Topic composition and diversity measures of the database

post-1990 period while the second shows a steeper decline over the whole sample period. These suggest a gradual shrinking of the audience base of the core root literature.

Further examination into the topic ratios based on equation (10.3) tells us that the decline is most discernible from microeconomics (D) and labour economics (J), and to a lesser degree macroeconomics (E). On the other hand, the econometrics topic (C) ratio is clearly on the increase (see Figure 10.2). The finding suggests that the core root literature has encouraged specialization and compartmentalization in later economic research, and that their non-decay impact has mainly been maintained by the econometrics profession itself, while it has been gradually left behind by the mainstream economics profession. It is now interesting to see how the impact of the CC programme has evolved against these general trends.

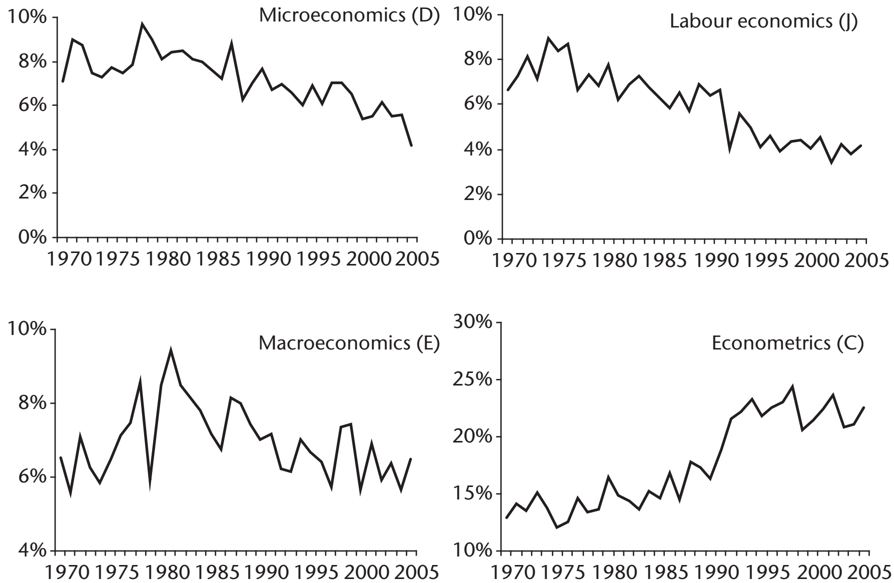

Figure 10.2. Impact factors of four major economic topics
Note: These series are calculated by equation (10.1a).

10.2 Citation analysis of the CC paradigm

Let us now turn to the construction and analysis of the impact measures of the CC paradigm. Since the database does not have extensive secondary citations or citation trees by design, two sequential sets of citation bases are constructed here. The first set is to capture the CC paradigm in its original form. The set consists of four bases. The first one is a wide base covering both the key works of Haavelmo and the CC group, referred to as 'CCWide'; the second is a subset of it, referred to as 'CCNarrow', focusing on the estimation and identification techniques developed for SEMs; the third is an applied base, referred to as 'CCApp', consisting of two of Haavelmo's papers and the early modelling works of Klein; and the last one is a methodological base containing several debates, referred to as 'CCMethy', before the CC programme became dominant. The second set is made up of four bases which emulated the CC paradigm—an estimator base, referred to as 'Estim', including the invention of two-stage least squares (2SLS), three-stage least squares (3SLS), instrumental variable (IV) estimators, and a Tobit estimation method; an identification base, referred to as 'Ident', following the monumental Cowles Commission Monograph 10 (Koopmans, 1950); and two applied bases of macro and micro modelling, referred to as 'AppMac' and 'AppMic' respectively, which are selected with reference to Bodkin et al. (1991) and Griliches and Intriligator (1983, 1986). In addition, two textbook bases are also constructed, one covering the eleven textbooks of the 1950–70 period, as listed in Chapter 1, and the other a subset of that base consisting of five popular textbooks of the post-1960 period. The two bases are referred to as 'Textbks' and 'Textbks60s' respectively. All the documents in the ten bases are listed in the Appendix Table in Section 10.5. The bases are also numbered in sequence as B1 to B10 in the tables and graphs in this chapter.

Table 10.1 reports the modified *s*-indices using equation (10.3) and the impact factors of the whole sample for all ten bases, along with their classified topics, n_i, and total citation counts. It is discernible that the two sets of statistics do not always yield the same ranking orders and that a certain scale effect is discernible in the impact factor statistics. Nevertheless, 'Estim' (B5) takes the clear lead among the first eight bases, but it is overtaken by 'Textbks60s' (B10). The base which exerts the least impact is 'AppMac' (B7), followed by 'Ident' (B6). On the whole, the citation impact of the three empirical bases, 'CCApp' (B3), 'AppMac' (B7), and 'AppMic' (B8), is overshadowed by that of

Table 10.1 Impact *s*-indices and impact factors of citation bases

Base (topics; n_i; citation count)	*s*-index *I* by (10.1)	Top five ranking topics (the topic *s*-indices) Their whole-sample *I* by (10.1a)				
1 (C, D, E, S; 25; 1275)	3.5548 3.8%	C (2.469) 7.5%	U (1.913) 3.6%	S (1.313) 18.7%	E (1.221) 4.7%	D (1.026) 3.3%
2 (C; 11; 616)	3.7160 1.8%	C (3.125) 5.3%	U (2.184) 2.1%	E (1.136) 2%	S (0.891) 4.3%	D (0.698) 0.8%
3 (C, E, D; 5; 196)	3.0750 0.6%	C (2.356) 1.4%	E (1.591) 1.8%	S (0.986) 2.7%	U (0.871) 0.2%	D (0.827) 0.6%
4 (C, E, S; 11; 539)	3.4810 1.6%	C (1.757) 1.7%	U (1.565) 1.1%	S (1.536) 11.5%	D (1.227) 2.2%	L (1.126) 1.7%
5 (C; 9; 2729)	8.6520 8.2%	C (5.290) 12.3%	U (4.621) 7.7%	D (2.994) 9.6%	L (2.963) 8.8%	G (2.671) 7.3%
6 (C; 8; 284)	2.9230 0.9%	C (1.933) 1.5%	U (1.833) 1.1%	E (0.678) 0.7%	D (0.620) 0.5%	J (0.467) 0.4%
7 (C, E; 13; 235)	2.0114 0.7%	C (1.451) 1.5%	E (1.114) 2.2%	S (0.624) 2.7%	U (0.623) 0.2%	L (0.445) 0.4%
8 (C, D, L, Q; 13; 574)	3.2990 1.7%	C (2.037) 2.7%	L (2.030) 6%	D (1.495) 3.6%	Q (1.376) 6.1%	U (0.801) 0.4%
9 (C, E, D; 13; 2453)	6.5358 7.4%	U (4.271) 9.9%	C (3.246) 7.3%	D (1.889) 6.3%	G (1.870) 5.8%	L (1.742) 4.9%
10 (C, D; 5; 2089)	9.6940 6.3%	U (6.463) 8.7%	C (4.718) 6%	D (2.880) 5.7%	G (2.813) 5.1%	J (2.479) 5.2%
11 (C, E; 5; 467)	4.6890 1.4%	E (3.662) 8.9%	C (3.193) 2.9%	S (1.589) 6.2%	N (1.502) 11.3%	F (1.396) 2.1%
12 (C, U, J; 10; 1929)	6.9070 5.8%	C (4.530) 10%	U (3.538) 4.9%	E (2.402) 7.3%	G (2.281) 5.8%	J (2.175) 6.8%

Note: The topics in the bracket of the first column include all the topics which have been assigned to the documents in each base.

Table 10.2 Subsample *s*-indices of citation bases

1971–90	1986–2005	1971–90	1986–2005	1971–90	1986–2005	1971–90	1986–2005
CCWide (B1)		CCNarrow (B2)		CCApp (B3)		CCMethy (B4)	
2.784	2.542	2.949	2.609	2.556	1.911	2.724	2.517
Estim (B5)		Ident (B6)		AppMac (B7)		AppMic (B8)	
6.252	6.925	2.443	1.919	1.807	1.050	2.675	2.210
Textbks (B9)		Textbks60s (B10)		NBER (B11)		Test (B12)	
8.248	7.512	7.592	6.976	2.617	4.223	4.993	5.537

the other bases. The subsample *s*-indices in Table 10.2 reveal that the impact of all the bases, that is B1–B10, with the exception of 'Estim' (B5), recedes to various degrees in the latter part of the sample period.

The evolution of the impact of all the bases is better shown by the time series of the impact factors plotted in Figure 10.3. Discernibly, the series of the three empirical bases are on the relatively low side with a downward trend. But surprisingly, 'Ident' (B6) shows the lowest impact factor series. The series which decline the most rapidly are from the two textbook bases. They have lost their lead to 'Estim' (B5) since the 1980s. The other interesting feature is that there is no discernible decline in the impact factor series of the three non-applied CC bases in the first set, that is, 'CCWide', 'CCNarrow', and 'CCMethy', in addition to 'Estim', demonstrating a remarkable longevity of impact of the CC paradigm in its original form.

Let us now look into the disaggregate *s*-indices using equation (3a) in Table 10.1. We see that the leading rank is occupied by either the 'econometrics' topic (C) or the 'non-economics' topic (U). Major topics of economics, such as 'macroeconomics' (E) and 'microeconomics' (D) rank among the top five, but not those on more specialised topics, such as 'health economics' (H) and 'regional economics' (R). What is surprising perhaps is the absence of topic 'E' in 'CCMethy' (B4), 'Estim' (B5), and the two textbook bases, 'Textbks' and 'Textbks60s' (B9, B10). Note that topic 'S' (history and methodology) gets ranks in the top five with all four CC bases of the first set, indicating the historic prominence of the CC paradigm.[2]

To examine the evolution of topical distribution, we plot in Figure 10.4 the topic diversity index series using equation (10.6). It is immediately noticeable from the figure that base 'Estim' (B5) retains its top position throughout the 36-year span, corroborating the argument made in Chapter 1 that the essential heritage of the CC paradigm lies in the device of statistically optimal estimators. It can also be seen from the figure that the diversity series of the two textbook bases, 'Textbks' (B9) and 'Textbks60s' (B10), remain rather high,

[2] Topic 'S' covers the field of the history and methodology of economics in the classification. But documents on the history and methodology of econometrics are also classified into topic 'C'.

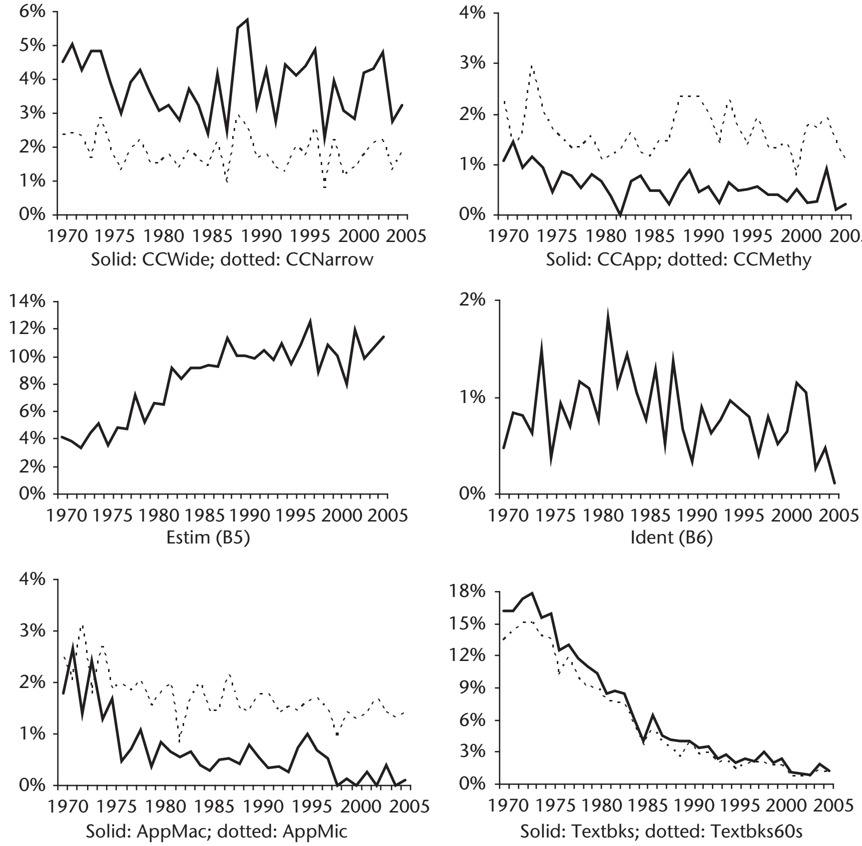

Figure 10.3. Impact factors of ten citation bases
Note: The scales of vertical axes vary across plots.

although there are signs of a gradual decline. Other declining series include base 'Ident' (B6) and the three empirical bases, 'CCApp' (B3), 'AppMac' (B7), and 'AppMic' (B8), with the trend in 'AppMic' (B8) the mildest. Again, the series of the three original CC bases other than 'CCApp' (B3) show no sign of weakening. These results are in broad agreement with those illustrated by the impact factor series plotted in Figure 10.3.

The next two figures present more detailed information on topic transfer and diffusion. Figure 10.5 plots the series of $\mu_{i,t}$ using equation (10.2) for all ten bases. As seen from the figure, there is a discernible trend of narrowing topic coverage from the time series of most of the bases in Figure 10.6, that is, a rising $\mu_{i,t} > 1$ over time, with the two textbook bases leading the trend. The exception is with the two empirical bases, AppMac (B7) and AppMic (B8). The series are in general agreement with the trend found from the bottom panels

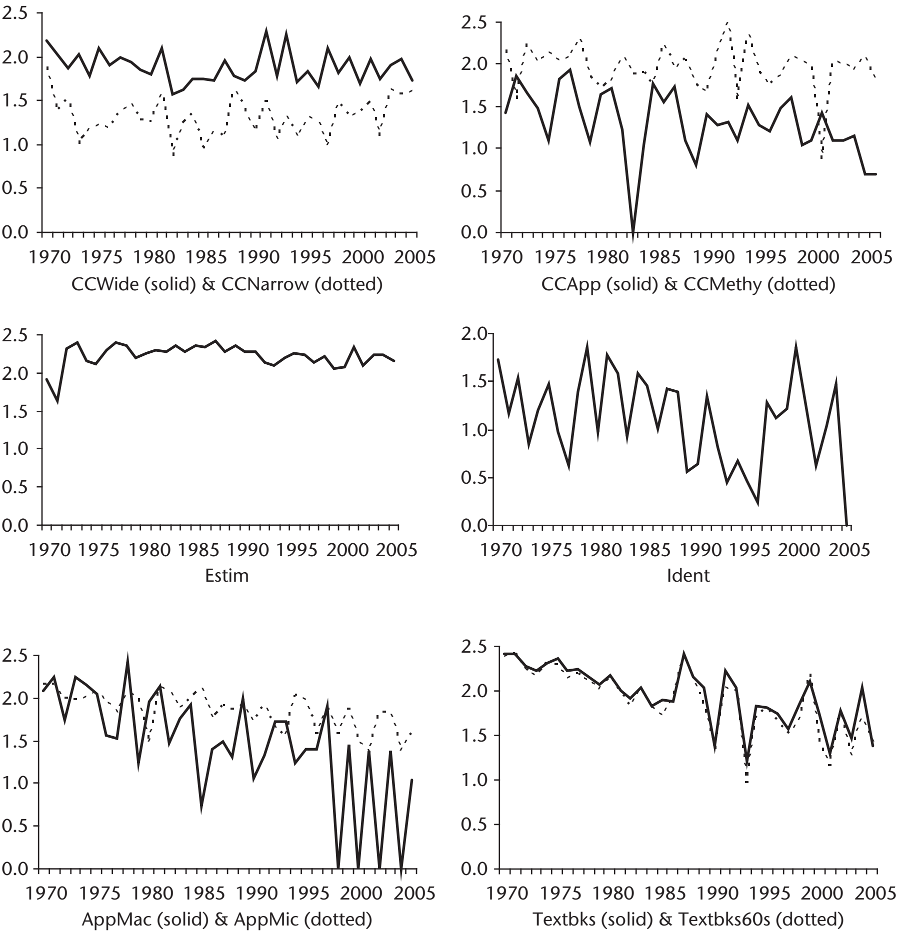

Figure 10.4. Topic diversity indices of ten bases

Note: Some curves swing to zero because there are zero citation counts at the observations for certain years.

of Figure 10.1, namely that the CC paradigm and its consolidation through textbooks have, on the whole, encouraged specialization.

Figure 10.6 provides three diffusion series, δ_t, for each base calculated by equation (10.5). The smallest one (lowest curve) is the diffusion indicator within economics but excluding the topics 'methodology and history of thought' (S) and 'survey' (T), the two topics are included in the next one (the dotted curve), and the non-economics topic (U) is added on in the largest indicator (the top curve). Note that the area above the third indicator up to 100 per cent represents the share of within-topic citations. It is interesting to find that base 'AppMic' (B8) has the lowest topic diffusion impact and its three diffusion indicators

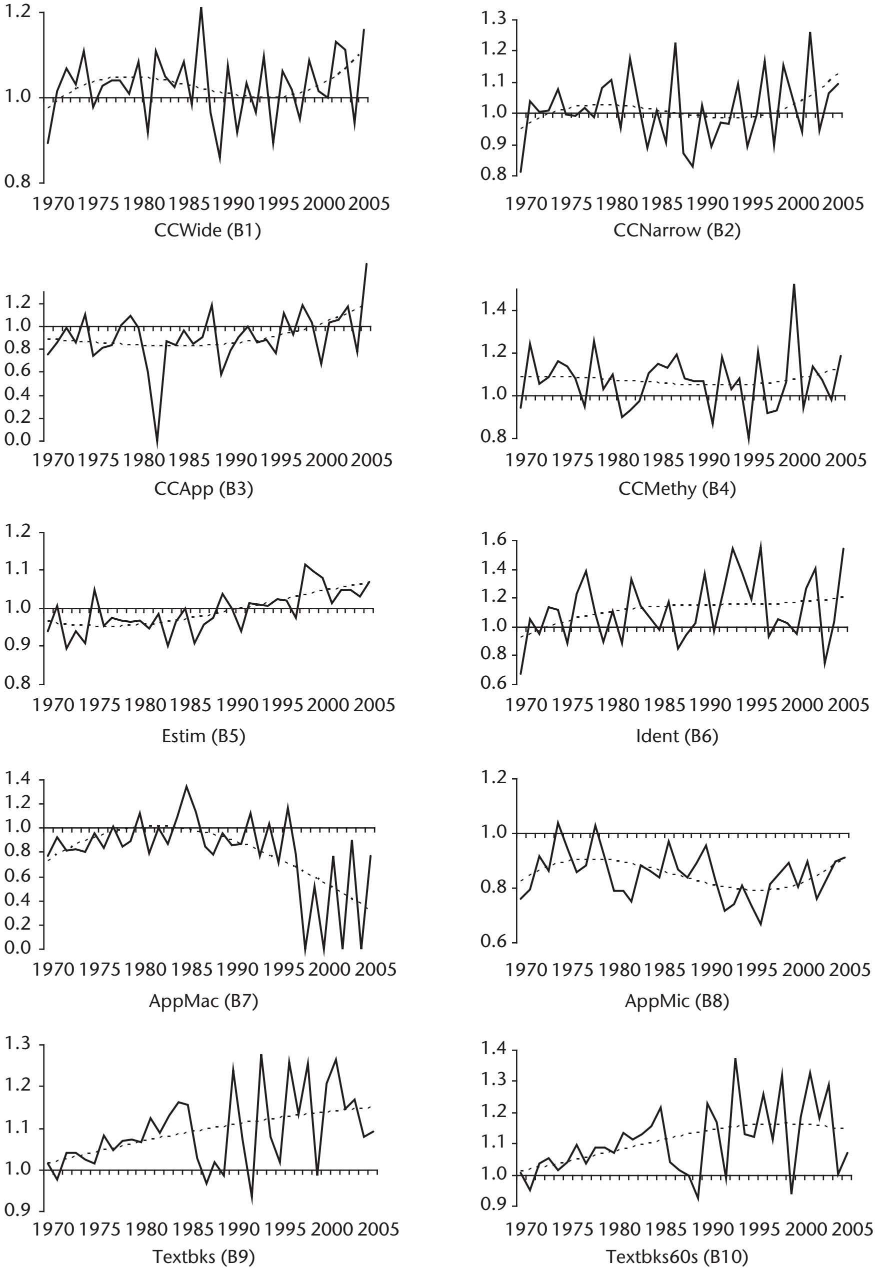

Figure 10.5. Series of μ_t using equation (10.2)
(Dotted curves are the polynomial trends based on μ_t)

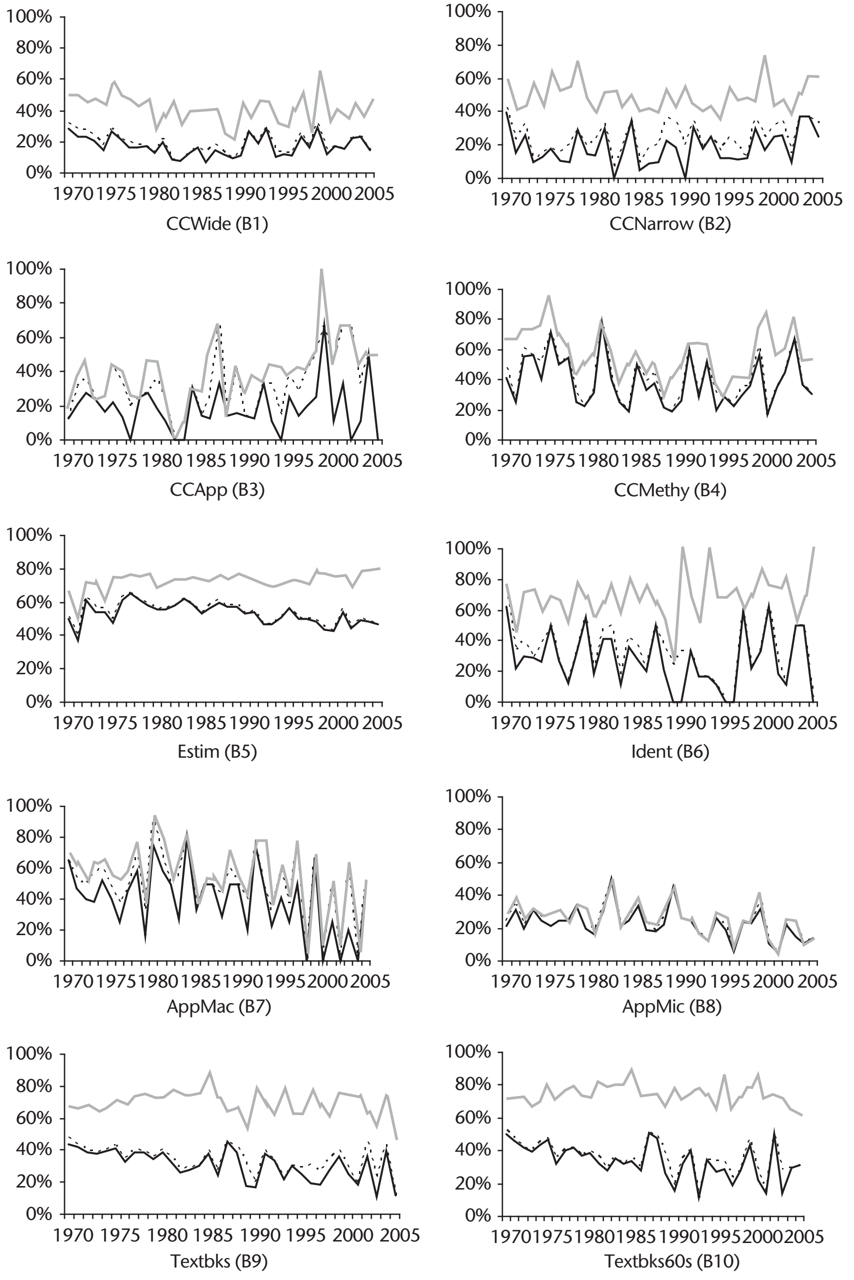

Figure 10.6. Diffusion indices using equation (10.5)

Solid curve at bottom—narrow set of economic topics (exclude topics 'S' and 'T'); dotted curve—the narrow set plus 'S' and 'T'; top grey curve—the dotted curve plus 'U' (non-economics) show hardly any difference. In comparison, base 'AppMac' (B7) demonstrates a higher diffusion impact, at least during the pre-2000 period. 'Estim' (B5) is probably the best performing base with relatively large and steady diffusion indicators. However, its diffusion into the non-economics field is overtaken by the two textbook bases (B9, B10). It is remarkable that the original works of Haavelmo and the CC group have also attracted quite a lot attention outside economics, that is, 'CCWide' (B1) and 'CCNarrow' (B2).

10.3 Citation analysis of comparative alternatives

Section 10.2 illustrates how the CC paradigm has been received through citations. But it does not tell us how strong the reception is in comparison with other alternatives, especially competing research approaches. In order to address this issue, two additional bases are constructed here as examples of alternatives to the CC paradigm. One is base 'NBER' (B11) containing the classical works of W. C. Mitchell and A. F. Burns on business cycle research at the National Bureau of Economic Research; the other is base 'Test' (B12) containing a group of papers proposing various model tests during the 1960s (see Appendix Table for the detailed list). The choice of base 'NBER' (B11) becomes easy and natural since the rivalry between the NBER and the CC camps is well-known following the 'measurement without theory' debate (see Koopmans (1947) and Vining (1949) listed in Appendix Table). Base 'Test' (B12) is constructed as a particular contrast to 'Estim' (B5), based on the observation that estimation of given structural models figured so centrally on the research agenda of the CC programme that hypothesis testing was not much heeded until the 1970s, when econometric modellers, especially macro modellers, began to shift their attention towards more rigorous testing of structural models, as described in previous chapters.

Let us look at the citation impact statistics of 'NBER' (B11) first. What is striking from Figure 10.7 is that its impact factor series has been on a rising trend since the mid 1980s and has caught up with CCWide (B1) since the late 1990s, in spite of the significant difference in base size, n_i (5 versus 25, Table 10.1). When the size factor is accounted for by the modified s-index, the impact of 'NBER' overtakes those of both 'CCWide' (B1) and 'CCNarrow' (B2), as shown in Table 10.1. Furthermore, the overtaking occurs during the post-1985 period, as shown from the subsample s-indices in Table 10.2 (comparing the index pair of B11 with those of B1 and B2). These citation impact statistics suggest that the NBER business cycle research programme has enjoyed longer research vitality than the CC programme, strengthening the case study in Chapter 6. If we look into the disaggregate s-indices in Table 10.1, we find the impact of 'NBER' on econometrics (C) is slightly higher than

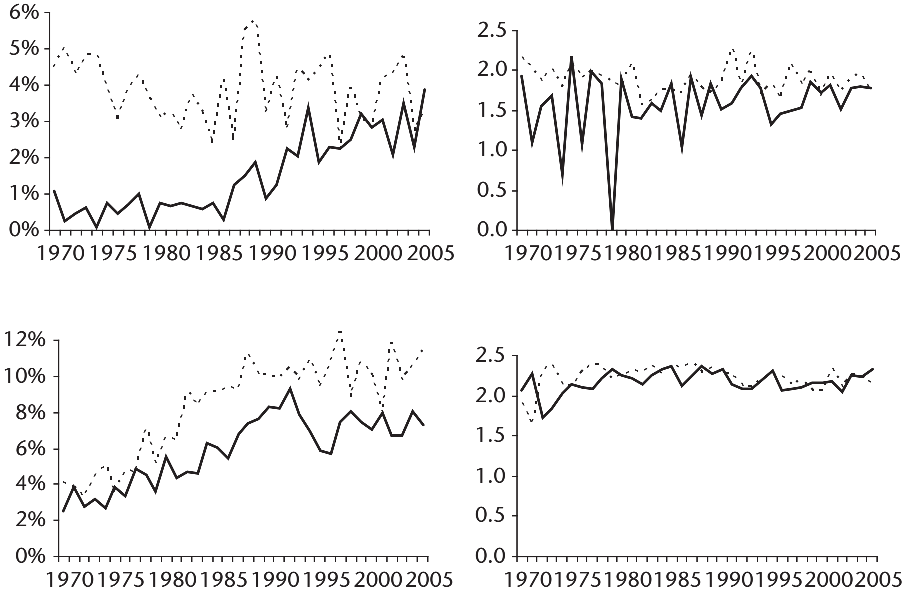

Figure 10.7. Impact measures of alternative bases with comparable measures from the previous citation bases

that of 'CCWide' and 'CCNarrow' (3.193 versus 2.469 and 3.125) while the impact of 'NBER' on macroeconomics (E) is far stronger (3.662 versus 1.221 and 1.136). Besides, 'NBER' has a fairly strong impact on economic history (N) and international economics (F) while the CC bases exert a rather strong impact on topics outside economics (U). The difference is more discernible from the comparison of the relevant topic transfer and diffusion graphs in Figures 10.8 and 10.6. There we see that 'NBER' (B11) has kept a higher rate of topic diffusion within economics than the CC bases on the whole but the latter enjoy a much wider margin of topic diffusion outside economics (U). In fact, the stronger diffusion in 'U' of 'CCWide' appears to have compensated its relative narrower diffusion within economics so that there is virtually little difference in the aggregate topic diversity series between the two bases (see Figure 10.7). However, 'NBER' (B11) outperforms the CC bases in the topic coverage indicator, if we compare its $\mu_{i,t}$ series in Figure 10.8 with those of 'CCWide' (B1) and 'CCNarrow' (B2) in Figure 10.5. Overall, the citation-based impact measures suggest that the CC paradigm has eventually lost the upper hand in the 'measurement without theory' debate.

Let us now turn to the citation statistics of base 'Test' (B12). Comparison with those of 'Estim' (B5) results in many similarities. There is an upward trend in both impact factor series (see the left middle panel of Figure 10.7) only the 'Estim' series leads a few years ahead and also maintains a higher position. The

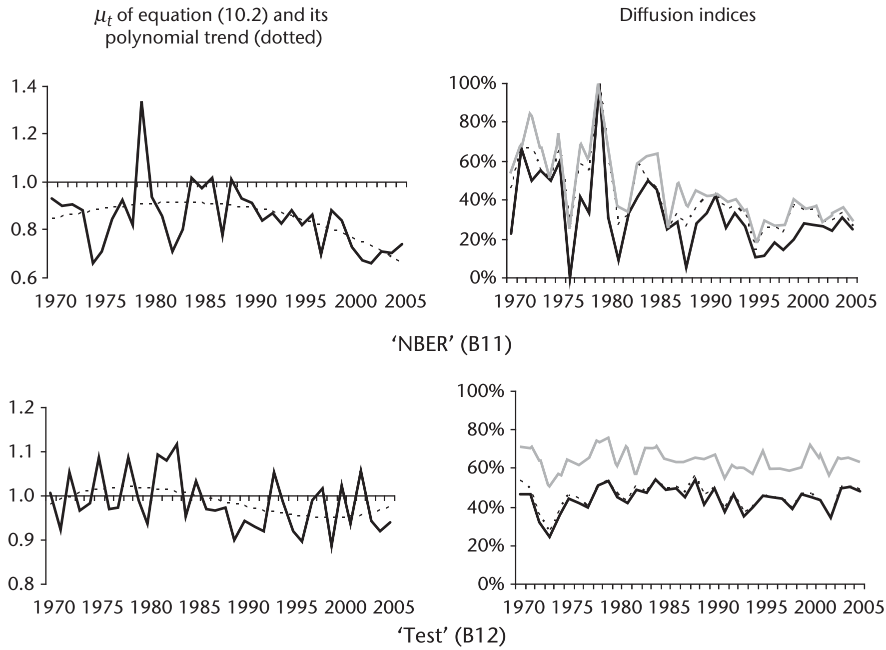

Figure 10.8. Impact measures of alternative bases
Note: See Figure 10.6 for the definition of the diffusion series.

leading position of 'Estim' is verified by the aggregate impact *s*-indices in Table 10.1. In fact, 'Estim' is the only one of the first eight bases whose subsample *s*-index increases in the second period (see Table 10.2). These findings confirm that there has been rising research attention on model testing since the mid-1980s but estimation still retains its popularity. The impact measures with respect to topic coverage, diffusion, and diversity are also extremely similar between 'Test' and 'Estim', if we compare the relevant panels in Figures 10.5–10.8. Probably the only interesting difference between the two bases lies in the pattern of their disaggregate *s*-indices (see Table 10.1). While 'Estim' has microeconomics (D) as its topic of highest impact within economics (other than econometrics), what 'Test' impacts on most is macroeconomics (E). The difference confirms the observation made in earlier chapters that reforms of the CC paradigm have occurred mostly in macro- rather than micro-econometrics.

10.4 Concluding remarks

The CC paradigm was able to withstand the post-1970 reformative movements and the above citation analysis illustrates its enduring and extensive impact almost half a century after its original contribution. This longevity is primarily sustained by emulous research into devising estimators for various types of models, the impact of which is still growing and widening through topic diffusion and diversity. The economic topics most receptive to diffusion are closely related to microeconomics, a finding which corroborates Heckman's (2000) vision of the CC tradition. The longevity is also helped by the publication of textbooks, which played an important role in topic diffusion, notably outside economics, although their impact has significantly declined over time.

Through its diffusive impact on various topics within and beyond economics, the CC paradigm is shown to have helped promote the specialisation and compartmentalisation of economic research. During the process, academic research interest in empirical modelling, especially applied macro modelling, has noticeably declined. Another area with a notable decline of interest is 'identification'. The relatively poor citation measures of 'Ident' (B6) confirm Hoover's (2004) argument on 'lost causes'.

Citation measures of alternative routes illustrate that the modelling community has paid increasing attention to methods of model testing, attention which has become almost comparable to that on estimation. But, more interestingly, the NBER business cycle research programme, an old-time methodological rival of the CC paradigm, is found to have maintained greater stamina not only in its overall citation impact but also in topic diffusion within economics, particularly econometrics. The lack of dominance of the CC paradigm and its gradual dwindling impact could indicate that the reforms have been effective, but have not yet reached a sufficient degree to revolutionize the discipline. There is not yet enough evidence for a Kuhnian (1962) type of 'paradigm shift'.

10.5 Appendix: database description

The citation database for the history of econometrics is built upon over 1300 root documents dating from 1913 up to 1969, and consists of around 34,700 entries in total. The root documents include early econometric works, deriving initially from bibliographies of relevant historical studies, for example Epstein (1987), de Marchi and Gilbert (1989), Morgan (1990), Bodkin et al. (1991), Qin (1993), and various econometrics textbooks as well. The emphasis is on the evolution of econometrics since its formalization during the 1940s, as shown from the extension of those initial root documents—an issue-by-issue search of econometrics papers from eight journals via JSTOR:

1. *Econometrica*, period: 1933–1969;
2. *American Economic Review*, period: 1950–1969;

3. *Economic Journal*, period: 1950–1969;
4. *Journal of American Statistical Association*, period: 1950–1969;
5. *Journal of Royal Statistical Society*, period: 1950–1969;
6. *Journal of Political Economy*, period: 1950–1969;
7. *Quarterly Journal of Economics*, period: 1950–1969;
8. *Review of Economics and Statistics*, period: 1950–1969.

Furthermore, works of key authors, such as Frisch, Haavelmo, Koopmans, Morgenstern, Wold, are also checked and added if not already covered in the above cited literature.

Most of the citation documents have been collected from Web of Science, which starts in 1970. The citation period ends at 2005. Note, however, that not all the journals in Web of Science have their reference lists linked as early as 1970 because these journals used to have an incompatible format of referencing.[3] A small number of citation documents for the pre-1970 period have been collected from JSTOR.

All the documents come from over 2600 publication sources, more than 2500 of which are journals. Other sources include proceedings, monographs, and books. Some journals have changed name during the sample period and have been unified by their latest name to avoid double counting. All the documents are classified by a modified JEL classification system as follows:

C—Econometrics and other quantitative methods

D—Microeconomics

E—Macroeconomics

F—International economics

G—Financial economics

H—Public economics

I—Health, education, and welfare

J—Labour and demographic economics

K—Law and economics

L—Industrial organization

M—Business economics, marketing, and accounting

N—Economic history

[3] One such example is de Marchi and Gilbert (1989). As a special issue of *Oxford Economic Papers*, the volume adopted the format of consolidated references, making it impossible for Web of Science to pick up citations from individual papers. Citations in the issue have been manually added into the database due to its special nature.

O—Economic development, technological change, and growth

P—Economic systems

Q—Agricultural and natural resource economics; environmental economics

R—Urban, rural, and regional economics

S—Methodology and history of thought

T—Surveys

U—Non-economics topics

When the document does not already specify its own JEL, the classification has been done manually according to the title and sometimes the abstract of each document, not the publication source. For example, when a document has been published in a statistics journal but deals with an economics-related issue, it is classified by an economics topic at least; otherwise it is classified into the non-economics category. Note that the non-economics category covers a wide range of fields such as medicine, engineering, physics, politics, and environmental science. Note also that each document can have more than one topic following the JEL classification convention. For instance, Bodkin et al. (1991) is assigned to the topics of C, E, and S.

Table 10.A1 List of documents in citation bases used in Sections 10.2 and 10.3

Root document	Base
Anderson and Rubin (1949) Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations	1, 2
Anderson and Rubin (1950) The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equations	1, 2
Girshick and Haavelmo (1947) Statistical Analysis of the Demand for Food Examples of Simultaneous Estimation of Structural Equations	1, 3
Haavelmo (1943) Statistical Testing of Business-Cycle Theories	1
Haavelmo (1943) The Statistical Implications of a System of Simultaneous Equations	1, 2
Haavelmo (1944) The Probability Approach in Econometrics	1
Haavelmo (1947) Methods of Measuring the Marginal Propensity to Consume	1, 3
Hood and Koopmans (1953) Studies in Econometric Method	1, 2
Hurwicz (1944) Stochastic Models of Economic Fluctuations	1
Klein (1947) The Use of Econometric Models as a Guide to Economic Policy	1, 3
Klein (1950) Economic Fluctuations in the United States 1921–1941	1, 3
Koopmans (1941) The Logic of Econometric Business-Cycle Research	1
Koopmans (1945) Statistical Estimation of Simultaneous Economic Relations	1, 2
Koopmans (1947) Measurement Without Theory	1, 4
Koopmans (1949) Identification Problems in Economic Model Construction	1, 2
Koopmans (1949) The Econometric Approach to Business Fluctuations	1, 4
Koopmans (1950) Statistical Inference in Dynamic Economic Models	1, 2
Koopmans (1950) When Is an Equation System Complete for Statistical Purposes	1
Koopmans (1957) Three Essays on the State of Economic Science	1, 4
Koopmans, Rubin, and Leipnik (1950) Measuring the Equation Systems of Dynamic Economics	1, 2

(Continued)

Table 10.A1 (Continued)

Root document	Base
Koopmans and Reiersøl (1950) The Identification of Structural Characteristics	1, 6
Mann and Wald (1943) On the Statistical Treatment of Linear Stochastic Difference Equations	1, 2
Reiersøl (1950) Identifiability of a Linear Relation between Variables Which Are Subject to Error	1, 2
Simon (1952) On the Definition of the Causal Relation	1, 6
Simon (1953) Causal Ordering and Identifiability	1, 2, 6
Klein and Goldberger (1955) An Econometric Model of the United States 1929–1952	3
Klein (1960) Single Equation Vs Equation System Methods of Estimation in Econometrics	4
Liu (1960) Underidentification, Structural Estimation, and Forecasting	4
Orcutt (1952) Toward Partial Redirection of Econometrics	4
Vining (1949) Koopmans on the Choice of Variables to Be Studied and of Methods of Measurement	4
Waugh (1961) The Place of Least Squares in Econometrics	4
Wold (1954) Causality and Econometrics	4
Wold (1955) Causality and Econometrics Reply	4
Wold (1956) Causal Inference from Observational Data: A Review of Ends and Means	4
Balestra and Nerlove (1966) Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model The Demand for Natural Gas	5
Basmann (1957) A generalized classical method of linear-estimation of coefficients in a structural equation	5
Sargan (1958) The Estimation of Economic Relationships using Instrumental Variables	5
Sargan (1959) The Estimation of Relationships with Autocorrelated Residuals by the Use of Instrumental Variables	5
Theil (1953) Estimation and simultaneous correlation in complete equation systems	5
Tobin (1955) The application of multivariate probit analysis to economic survey data	5
Tobin (1958) Estimation of Relationships for Limited Dependent Variables	5
Zellner (1962) An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias	5
Zellner and Theil (1962) Three-Stage Least Squares Simultaneous Estimation of Simultaneous Equations	5
Fisher (1959) Generalization of the Rank and Order Conditions for Identifiability	6
Fisher (1961) Identifiability Criteria in Nonlinear Systems	6
Fisher (1963) Uncorrelated Disturbances and Identifiability Criteria	6
Fisher (1965) Near-Identifiability and the Variances of the Disturbance Terms	6
Fisher (1966) The Identification Problem in Econometrics	6
Adelman and Adelman (1959) The Dynamic Properties of the Klein–Goldberger Model	7
Barger and Klein (1954) A Quarterly Model for the United States Economy	7
Christ (1951) A Test of an Econometric Model for the United States 1921–1947	7
Duesenberry, Klein, and Kuh (1965) The Brookings Quarterly Econometric Model of the United States	7
Duesenberry, Klein, and Kuh (1969) The Brookings Model: Some Further Results	7
Evans (1966) Multiplier Analysis of a Post-War Quarterly US Model and a Comparison with Several Other Models	7
Evans and Klein (1968) The Wharton Econometric Forecasting Model	7
Fox (1956) Econometric Models of the United States	7
Gallaway and Smith (1961) A Quarterly Econometric Model of the United States	7
Goldberger (1959) Impact Multipliers and Dynamic Properties of the Klein-Goldberger Model	7
Klein and Shinkai (1963) An Econometric Model of Japan, 1930–59	7
Klein, Ball, Hazlewood, and Vandome (1961) An Econometric Model of the United Kingdom	7
Nerlove (1962) A Quarterly Econometric Model for the United Kingdom	7

(Continued)

Table 10.A1 (Continued)

Root document	Base
Barten (1964) Consumer Demand Functions under Conditions of Almost Additive Preferences	8
Barten (1967) Evidence on the Slutsky Conditions for Demand Equations	8
Barten and Turnovsky (1966) Some Aspects of the Aggregation Problem for Composite Demand Equations	8
Barten, Theil, and Leenders (1962) Farmers' Budgets in a Depression Period	8
Jorgenson and Stephenson (1967) The Time Structure of Investment Behavior in United States Manufacturing, 1947–1960	8
Jorgenson and Stephenson (1967) Investment Behavior in US Manufacturing 1947–1960	8
Jorgenson and Stephenson (1969) Anticipations and Investment Behavior in U. S. Manufacturing 1947–1960	8
Kmenta (1964) Some Properties of Alternative Estimates of the Cobb–Douglas Production Function	8
Kmenta and Williamson (1966) Determinants of Investment Behavior United States Railroads 1872–1941	8
Nerlove (1956) Estimates of the elasticities of supply of selected agricultural commodities	8
Nerlove (1965) Estimation and Identification of Cobb–Douglas Production Function	8
Summers (1959) A Note on Least Squares Bias in Household Expenditure Analysis	8
Zellner, Kmenta, and Drèze (1966) Specification and Estimation of Cobb–Douglas Production Function Models	8
Brennan (1960) Preface to Econometrics	9
Christ (1966) Econometric Models and Methods	9
Dhrymes (1970) Econometrics Statistical Foundations and Applications	9, 10
Fox (1968) Intermediate Economic Statistics	9
Goldberger (1964) Econometric Theory	9, 10
Johnston (1963) Econometric Methods	9, 10
Klein (1953) A Textbook of Econometrics	9
Leser (1966) Econometric Techniques and Problems	9
Malinvaud (1966) Statistical Methods in Econometrics	9, 10
Tinbergen (1951) Econometrics	9
Tintner (1952) Econometrics	9
Valavanis (1959) Econometrics	9
Wonnacott and Wonnacott (1970) Econometrics	9, 10
Bases of Alternative Routes	
Burns and Mitchell (1946) Measuring Business Cycles	11
Mitchell (1913) Business Cycles and Their Causes	11
Mitchell (1927) Business cycles: The Problem and Its Setting	11
Mitchell (1951) What Happens During Business Cycles: A Progress Report	11
Mitchell and Burns (1938) Statistical Indicators of Cyclical Revivals	11
Chow (1960) Tests of Equality Between Sets of Coefficients in Two Linear Regressions	12
Durbin and Watson (1950) Testing for Serial Correlation in Least Squares Regression I	12
Durbin and Watson (1951) Testing for Serial Correlation in Least Squares Regression II	12
Durbin (1957) Testing for Serial Correlation in Systems of Simultaneous Regression Equations	12
Glejser (1969) A New Test for Heteroskedasticity	12
Goldfeld and Quandt (1965) Some Tests for Homoscedasticity	12
Quandt (1960) Tests of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes	12
Ramsey (1969) Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis	12
Sargan (1964) Wages and Prices in the UK: A Study in Econometric Methodology	12
Theil and Nagar (1961) Testing the Independence of Regression Disturbances	12

Epilogue

Chapter 10 concludes from citation evidence that the CC paradigm is far from being refuted despite the reformative movements during the period 1970–1990. This is further corroborated by the citation impact indices, given in Table 11.1, of three leaders of those movements together with a few prominent econometricians selected mostly from the authors listed in the Appendix Table of Chapter 10.[1] It can easily be seen from Table 11.1 that the three leaders have not outpaced, in terms of citation impact, econometricians whose works fall broadly into the traditional CC approach, especially those whose research mainly falls into the area of micro-econometrics.

More corroborating evidence can be found from currently popular econometrics textbooks. If we scan through the lists of contents of the books, such as Greene's *Econometric Analysis*, Wooldridge's *Introductory Econometrics: A Modern Approach*, or *Introduction to Econometrics* by Stock and Watson, we can easily notice that the essential structure of the textbooks of the 1960s, as described in Chapter 1, has been maintained. Most of the contents are devoted to derivations of estimators for a priori given structural parameters. Introduction of statistical tests, especially diagnostic tests, is often conveyed as the prerequisite for the prescription of more sophisticated estimators. Little attention is given to how to use statistical methods for research which contains a strong element of exploratory data analysis or empirical discovery, or for issues concerning applied model specification, selection, or design, let alone to major methodological debates or alternative modelling strategies. Econometrics is still taught as a universal statistical toolbox to furnish theories postulated by economists with numbers.

Why, then, has there not yet been a 'paradigm shift' in spite of all those reformative movements during the 1970–1990 period? What can be regarded as the major methodological advances of those movements? How effective are the advances in reshaping econometrics? What is the key insight or lesson that can be drawn from this historical investigation? After years of toiling at this history project, it is finally time for me to venture some tentative answers.

[1] For a detailed description of the data source, see the note to Table 11.1.

Table 11.1 Citation impact indicators for individual authors (1970–2005)

Author:	E. E. Leamer	C. A. Sims	D. F. Hendry
Root document	44	25	42
Total citation count	1027	2726	2108
h-index	16	16	22
s-index	2.379	5.106	3.433
Author:	L. R. Klein	F. M. Fisher	H. Theil
Root document	48	53	93
Total citation count	164	878	496
h-index	9	17	12
s-index	0.898	2.021	1.142
Author:	D. Jorgenson	Z. Griliches	A. Zellner
Root document	58	56	49
Total citation count	2028	2770	948
h-index	24	29	21
s-index	2.923	3.293	2.130
Author:	M. Nerlove	J. Kmenda	J.J. Heckman
Root document	60	23	58
Total citation count	780	212	5367
h-index	16	7	32
s-index	1.793	1.491	4.656

Note: Data collected from Web of Science. However, it is impossible to get any comparable data for Haavelmo, or members of the CC group other than Klein, or many other early econometricians, because the citation statistics from Web of Science start from 1970 and cover only journal-based root documents. The authors other than the three leaders are selected by their relatively high *h*-indices among the authors in the Appendix Table of Chapter 10. The root documents exclude works of a non-research type, such as book reviews. The *h*-indices are taken directly from Web of Science; The *s*-indices are calculated by equation (10.3) of Chapter 10.

One feature of the reforms which immediately comes to mind is the widening of research issues other than an estimation of a priori given structural parameters targeted primarily by the CC methodology and the related fact that contentions and diverse approaches were concentrated on those widened stretches.[2] To delve further into the feature, I find Cox's (1990) classification of three broad roles or purposes of models in statistical analysis and their subdivisions particularly useful:

$$\textit{Substantive}\begin{cases}\textit{Directly substantive}\\ \textit{Substantive hypothesis of interdependence}\\ \textit{Retrospective discovery of substantive issues}\end{cases}$$

$$\textit{Empirical}\begin{cases}\textit{Estimation of effects and their precision}\\ \textit{Correction of deficiencies in data}\end{cases}$$

$$\textit{Indirect}\begin{cases}\textit{Calibration of methods of analysis}\\ \textit{Automatic data reduction}\end{cases}$$

[2] This finding is not new. It corroborates the view of the editors of the *Palgrave Handbook of Econometrics*, as expressed in the 'Editors' Introduction' by Mills and Patterson (2006).

Note that these three roles 'may occur in combination in a specific application' and that 'quite often parts of the model, e.g., those representing systematic variation, are based on substantive considerations with other parts much more empirical', as emphasized by Cox (1990: 172).

In the light of Cox's classification, mainstream econometrics has evolved around the 'estimation of effects and their precision' of 'directly substantive' models. Most of the methodological battles have been fought on the crossing of the three variants within substantive models. The CC enterprise was built on the estimation of 'directly substantive' models, referred to as 'structural models'. Leamer highlighted a number of commonly occurring practical problems in fitting such structural models, diagnosed them as due to the lack of informational precision in estimators designed by means of classical statistics and prescribed the Bayesian approach as a systematic remedy. The VAR approach moved away from those directly substantive models to models comprising mainly substantive hypotheses of interdependence. Other than issues from the econometric aspect which triggered it, the move corresponded to a similar shift in macroeconomics led by the RE movement. The LSE approach explored the role of the retrospective discovery of substantive issues as a main device to treat those problems complained about by Leamer as well as to resist the lessening of attention on structural parameters, especially those parameters embodying the long-run equilibrium relations in directly substantive models. Both the VAR approach and the LSE approach met with strong resistance from those 'fundamentalists' who took structural models as belonging solely and entirely to the directly substantive type, in spite of the fact that the rise of those alternative approaches has not really diluted the emphasis of core econometric research on the estimation-related 'empirical' aspect of substantive models.

Nevertheless, those reformative enterprises have helped to break the textbook confinement of econometrics to the style of dominantly confirmatory analysis of given directly substantive models and made exploratory data analysis an increasingly essential element in modelling research (e.g. Hendry, 2011). Through highlighting that it is a cul-de-sac to camouflage what are fundamentally model formulation and selection issues and to tackle them within the realm of improving estimation precision for fixed directly substantive models, the reforms have urged econometricians to venture into the territory of substantive model formulation and specification. Consequently, various types of data-instigated parameters and error terms have been introduced into econometric models. Interestingly, the introductions are frequently furnished with 'structural' interpretations and justified by the need to improve estimation precision in broad allegiance to the CC paradigm. Such practice has given rise not only to multiple, often confusing 'structural' connotations but also to cross-purpose arguments about how to define and conduct structural modelling.

Expansions of substantive modelling research in economics have certainly contributed to the confusion. Economists have become increasingly receptive to data features and empirical results in theoretical model formulation, as clearly shown from RE models (see also Eichenbaum, 1995). Within econometrics territory, however, the confusion could be attributed to the cost or the limit of the reforms. They have never explicitly revolted against the primary role, assumed in the CC paradigm, of maintaining substantive models via measurement provision. Most of the fragile model specifications exposed by Leamer are treatable by the classical route, while Bayesian tools have been adopted widely, irrespective of apparent methodological differences, because of their usefulness mostly for the parameter-based 'estimation and the related precision' purpose via focusing on the probability space of the parameter domain. The VAR approach has, irrespective of its archaic 'astructural' label, been assimilated as an indispensible toy in generating shock-instigated dynamic stories in macroeconomic discourses. Furnished by measured empirical supports, these stories significantly strengthened the maintained position of whatever a priori hypotheses because the structuralized residual errors as substantive shocks have virtually immunized the hypothetical models from serious mis-specification contention. The popularity of the LSE approach rests crucially on the connection of the error-correction or cointegration component in ECMs to established long-run equilibrium relations in traditionally formulated directly substantive models. The component has virtually become the hallmark of the models produced by the LSE approach in spite of the fact that the estimated long-run effects are frequently much smaller than most of the short-run effects, dependent on certain idiosyncratic factors and susceptible to institutional or historical based regime shifts. Viewed in this light, the reforms have actually helped support the 'maintained position' of substantive models, and the dominance and primacy of the confirmative role has not been significantly altered through the augmentation of the exploratory role. The hard core of the CC paradigm has been further consolidated.

The limited nature of the reforms is probably best reflected in the fact that their achievement shines much more noticeably within the academic community than outside in resolving real-world economic issues. Since most substantive models are partially formulated under the *ceteris paribus* assumption, it is not difficult to produce empirical verification of them through arbitrarily controlling for *ceteris non paribus* factors. In other words, when the subject matter merely involves postulates of causes which partially explain certain modelled variables, it becomes logically unnecessary to rely on stringent tests of the model residuals for possibly inadequate model design. Empirical verifications of a particular partial cause are not difficult to come by, only they tend to vary with different choices of the additional control variables needed to fill the partially unexplained gap. However, these control variables are of little direct concern from the point of view of those empirical modellers whose attention is focused on the given subject matter. What becomes a main concern is that the consequent empirical verifications are not unambiguous or conclusive enough to help settle theoretical and policy related disputes, as illustrated in the case study of the Phillips curve in Chapter 5. The case study shows that such empirical studies have seldom succeeded in adding conclusively quantifiable exactness to a priori given directly substantive themes, as expected of econometrics from the textbook teaching. What are, then, the new understandings that those studies provide us?

To thoughtful and experienced applied modellers, those studies simply show how incomplete most of the a priori given directly substantive models are for statistical purposes. In fact, attempts to rectify the incompleteness from the data end form the essence of the reforms. For instance, the VAR approach and the LSE approach promote the route of a general dynamic extension of the a priori partial causes so as to close the gap between the *ceteris non paribus* factors and white-noise residual errors. While the route has yielded notable empirical improvements, albeit being labelled 'measurement without theory' by critics, the task of selecting and designing empirically complete models extended from subject-matter based partial models has turned out to be unexpectedly recalcitrant. The task essentially requires a shift of emphasis from the interpretive role of substantive models to prediction. Econometric models still fall short of providing timely predictions of major economic downturns, as shown from the case study in Chapter 6, in spite of general improvements in dynamic model specification. The task of business cycle monitoring and forecasting would require modelling research to be focused on the discovery of substantive issues from topical concerns in a real-world environment, an area still much under-developed in econometrics. Mainstream econometrics is still dominated by the belief that modelling for theory-based or policy-related issues is more important and superior, or philosophically more appropriate,[3] than for forecasting.

Meanwhile, econometric research as a whole continues to derive its meaningfulness from the hermeneutics of data in line with known substantive models. The incomplete nature of the subject matter provides econometricians with seemingly endless measurement tasks and fertile ground for academic research. The recurring interest in the identification of causal

[3] For a general and philosophical argument that research in the social sciences should be focused on explanation and understanding rather than prediction due to the non-deterministic nature of social events, see Manicas (2006).

parameters serves as a best example here (e.g. Heckman, 2000, 2008). The usually meagre practical gain is largely overshadowed by a steady growth of new supplies in estimators and tests backed by rigorous probability and distribution discussions. The growth is so vibrant that it appears impossible to see any significant change to the situation sneered at by Frisch as 'playometrics' and criticized by Leontief as the unhealthy state of our discipline over four decades ago. In fact, similar criticisms have been voiced repeatedly by econometricians, economists, and scholars of the history and methodology of economic thought, especially in the wake of the recession triggered by the 2008 financial crisis (e.g. Buiter, 2009; Krugman, 2009). The formalization trend led and fostered by the CC paradigm has been simply unstoppable, irrespective of its incomparably low and small empirical achievements. As long as econometrics is widely perceived as mainly useful for the role of supporting 'directly substantive' models, any formalization endeavours may find themselves serving primarily to reinforce and enrich the deeply rooted ideology-based tradition in economics, than to enhance the chance of discovering empirical regularities, given the nature of non-experimental and inaccurate data and of the mutable contingencies of the economic environment.

From the angle of philosophy of science, the enduring scientific attraction of the CC approach can be regarded as lying in its provision of a methodological closure for applying analytical statistics to the service of economic measurement. Such a closure comprises three elements: (i) a set of self-contained variables of substantive interest; (ii) the existence of certain, expected regularities among these variables representing a certain structural mechanism; and (iii) the possibility of separating the regularities to the level of individual input variables (e.g. Olsen and Morgan, 2005). Notice that (i) and (ii) closely resemble Frisch's notion of 'autonomy'. Although periodic critiques of the CC approach have erupted subsequent to notable cases of serious incompatibility between the methodology and reality, the majority of the academic community have maintained their faith in the methodology by opting either to stay closer to the areas where they have greater confidence in conducting relatively closed empirical investigations or to improve the CC approach in the hope that the improved methodologies would relate better to reality. The latter is embodied in the reformative movements described in Chapters 2–4. Essentially, the reforms are centred on two points. First, economic theorists should no longer be shouldering job (i) alone, that is, the provision of adequate sets of self-contained variables in the form of directly substantive models; and second, it is more important for econometricians to attend to elements (i) and (ii) than (iii) in empirical modelling. Nevertheless, those alternative approaches are reformative as their methodological improvements are formalized in a close resemblance of the CC tradition. The Bayesian approach resorts to the subjective Bayesian stance to justify the gap between a CC-style methodological closure and closure in reality; the VAR approach advocates VARs as the ultimate model type to bridge the gap; the LSE approach encourages the active use of data-driven specification searches to supplement the incompleteness of a priori given sets of self-contained variables. The advocacy and promotion of the alternative approaches rests heavily on the presentation of methodological closures similar to that of the CC.

Needless to say, historical evaluation of any research methodologies presupposes that the methodologies have already distinguished themselves among the academic community. To achieve such a state, pedantic persuasion becomes a primary and necessary task, overtaking that of practical concerns, even though these are frequently used as initial motivating justifications. The history thus abounds with cases of research trends being driven by powerful rhetoric, a phenomenon described as one of 'more heat than light' by Mirowski (1989). In an environment where econometrics has been widely taught as a universal statistical toolbox, the attraction of any newly emerged methodology would depend heavily on its 'universal solvent' appeals (Swann, 2006: Preface), appeals which are sustained by a strong desire for certain, clear-cut, and simple methodological closure. It is thus no wonder that the VAR approach has gained popularity more than the other two approaches so far, especially among macro modellers, and that the LSE approach, in comparison, has been the least successful in converting believers (e.g. Magnus and Morgan, 1999). The desire in the profession to make universalistic claims following certain standard procedures of statistical inference is simply too strong to embrace procedures which explicitly rely on the use of vernacular knowledge for model closure in a contingent manner. More broadly, such a desire has played a vital role in the decisive victory of mathematical formalization over conventionally verbal based economic discourses as the principle medium of rhetoric, owing to its internal consistency, reducibility, generality, and apparent objectivity. It does not matter that 'as far as the laws of mathematics refer to reality, they are not certain'. What matters is that these laws are 'certain' when 'they do not refer to reality'.[4] Most of what is evaluated as core research in the academic domain has little direct bearing on concrete social events in the real world anyway.

However, it could be over-simplistic to dismiss such academic accomplishments purely for their lack of immediate relevance to real-world problems. The need to build a 'superstructure' for any established discipline is shown, at least, from two aspects in many historical studies. First, scientific research

[4] These quotes are from Einstein (1922), where he writes, 'as far as the laws of mathematics refer to reality, they are not certain, and as far as they are certain, they do not refer to reality' (Part II Geometry and Experience).

is generally full of advocacy and subjective biases, but scientific objectivity evolves, in the long run, from heated, intense, and biased confrontations between biased ideas of biased individuals (e.g. Mitroff, 1972, 1974). Therefore, 'more heat' should be seen as a necessary short-run phase in the search for the eventual 'light'. Secondly, the maturity of a discipline entails its branching out into refined sub-disciplines and an associated division of labour. In econometrics, this is embodied in the forming of a hierarchy where academic econometricians have gained the freedom to converse in abstruse language on 'theological' issues because a critical mass of professionally trained practitioners has been reached and these practitioners have been shouldering the responsibility of dealing with every day 'parish' matters in non-academic institutions such as government agencies and policy banks. The search for the eventual 'light' and better research strategies does not call for a radical overthrow of the hierarchy. Rather, it requires frequent exchanges of opinion and heated debates across its hierarchical ranks to prevent institutional decay with respect to the evolving external world. In econometrics, such debates are particularly needed to help keep modellers constantly reminded of the difficulties in applying statistical methods which are predicated on closure and regularity to an open economic reality, and to guard them against the tendency, in applied research, of conflating methodological closure with the closure of reality as well as the tendency, in theoretical research, of wilfully pursuing atomistic studies following the hard science oblivious to the prominent non-separable feature of the social and economic environment. It is also in this respect that historical investigations like the present one aim to make a constructive contribution.

References

Adelman, F. and I. Adelman (1959) The dynamic properties of the Klein-Goldberger Model, *Econometrica*, **27**, 596–625.

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. Proc. 2nd International Symposium on Information Theory, Budapest, pp. 267–81.

Akaike, H. (1981) Likelihood of a model and information criteria, *Journal of Econometrics*, **16**, 3–14.

Aldrich, J. (1989) Autonomy, *Oxford Economic Papers*, **41**, 15–34.

Aldrich, J. (1993) Cowles exogeneity and CORE exogeneity, Southampton University Economics Discussion Paper.

Allen, P. G. and R. Fildes (2005) Levels, differences and ECMs—principles for improved econometric forecasting, *Oxford Bulletin of Economics and Statistics*, **67** (suppl), 881–904.

Altonji, J. and O. Ashenfelter, 1980. Wage movements and the labour-market equilibrium hypothesis. *Economica,* **47**, 217–45.

Amemiya, T. (1980) Selection of regressors, *International Economic Review*, **21**, 331–54.

Anderson, T. W. (1971) *The Statistical Analysis of Time Series*, New York: Wiley.

Armstrong, J.S. (1978) Forecasting with econometric methods: folklore versus fact, *The Journal of Business*, **51**, 549–64.

Armstrong, J. S. (1979) Advocacy and objectivity in science, *Management Science*, **25**, 423–28.

Baba, Y., Hendry, D. F., and Starr, R. M. (1992) The demand for M1 in the U.S.A., 1960–1988, *Review of Economic Studies*, **59**, 25–61.

Ball, L. and G. N. Mankiw (2002) The NAIRU in theory and practice. *Journal of Economic Perspectives* **16**, 115–36.

Barger, H. and L. R. Klein (1954) A quarterly model for the United States economy. *Journal of the American Statistical Association*, **49**, 413–37.

Barro, R. J. (1977) Unanticipated money growth and unemployment in the United States. *American Economic Review,* **67**, 101–15.

Barro, R. J. (1978) Unanticipated money, output, and the price level in the United States. *Journal of Political Economy,* **67**, 549–80.

Barro, R. J. and M. Rush (1980) Unanticipated money and economic activity, in S. Fischer (ed.), *Rational Expectations and Economic Policy,* Chicago: University of Chicago Press, pp. 23–73.

Basmann, R. (1957) A generalized classical method of linear estimation of coefficients in a structural equation, *Econometrica,* **25**, 77–83.

Basmann, R. (1974) Exact finite sample distributions for some econometric estimators and test statistics: a survey and appraisal, in M. Intriligator and D. Kendrick (eds), *Frontiers of Quantitative Economics*, New York: North-Holland, ch. 4, pp. 209–71.

Bean, C. (2010) Joseph Schumpeter lecture: The great moderation, the great panic, and the great contraction. *Journal of the European Economic Association*, **8,** 289–325.

Belsley, D. A., E. Kuh, and R. E. Welsch (1980) *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, New York: John Wiley & Sons.

Bentzel, R. and H. O. A. Wold (1946) On statistical demand analysis from the viewpoint of simultaneous equations, *Skandinavisk Aktuarietidskrigft*, **29**, 95–114.

Bergstrom, A. R. and C. R. Wymer (1976) A model of disequilibrium neoclassical growth and its application to the United Kingdom, in A. R. Mergstrom (ed.), *Statistical Inference in Continuous-time Economic Models*, Amsterdam: North Holland, pp. 267–328.

Bernanke, B. S. (1986) Alternative explanations of the money–income correlation. *NBER Working Papers*, 1842; (also) *Carnegie-Rochester Conference Series on Public Policy,* **25**, 49–99.

Bernanke, B. S. and A. S. Blinder (1992) The Federal funds rate and the channels of monetary transmission. *American Economic Review,* **82**, 901–21.

Berndt, E. R. (1991) *The Practice of Econometrics: Classic and Contemporary,* Reading MA: Addison-Wesley.

Beveridge, S. and C. R. Nelson (1981) A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle', *Journal of Monetary Economics*, **7**, 151–74.

Bjerkholt, O. (2007a) Writing 'The Probability Approach' with nowhere to go: Haavelmo in the United States, 1939–1944, *Econometric Theory*, **23**, 775–837.

Bjerkholt, O. (2007b) Ragnar Frisch's business cycle approach: The genesis of the propagation and impulse model, *European Journal of the History of Economic Thought*, **14** (3), 449–86.

Bjerkholt, O. and D. Qin (eds) (2010) *A Dynamic Approach to Economic Theory: Lectures by Ragnar Frisch at Yale University* (Routledge Studies in the History of Economics), Abingdon: Routledge.

Blanchard, O. (2000) What do we know about macroeconomics that Fisher and Wicksell did not? *The Quarterly Journal of Economics*, **115**, 1375–409.

Blanchard, O. and D. Quah (1989) The dynamic effects of aggregate demand and supply disturbances, *American Economic Review*, **79**, 655–73.

Blanchard, O. and M. Watson (1984) Are business cycles all alike? *NBER Working Papers*, 1392.

Blanchard, O. J. and M. W. Watson (1986) Are business cycles all alike? In R. J. Gordon (ed.), *The American Business Cycle: Continuity and Change*, Chicago: University of Chicago Press, pp. 123–56.

Blinder, A. S. (1997) Is there a core of practical macroeconomics that we should all believe? *American Economic Review,* **87**, 240–3.

Bodkin, R. G., L. R. Klein, and K. Marwah (1991) *A History of Macroeconometric Model-Building*, Aldershot: Edward Elgar Publishing Co.

Boughton, J. M. (1993) The demand for M1 in the United States: A comment on Baba, Hendry and Starr. *Economic Journal,* **103**, 1154–7.

Box, G. E. P. and G. M. Jenkins (1970) *Time Series Analysis, Forecasting and Control*, San Francisco: Holden-Day.

Brennan, M. J. (1960) *Preface to Econometrics*, Cincinnati, OH: South-Western Publishing Company.

Brooks, T. A. (1986) Evidence of complex citer motivations, *Journal of the American Society for Information Science,* **37**, 34–6.

Brown, T. M. (1970) *Specification and Uses of Econometric Models*, London: Macmillan & Co.

Bry, G. and C. Boschan (1971) *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*, New York: NBER.

Buiter, W. (2009) The unfortunate uselessness of most 'state of the art' academic monetary economics, *Financial Times*, 3 March.

Burns, A. F. (1969) *The Business Cycle in a Changing World*, New York: NBER.

Burns, A. F. and W. C. Mitchell 1946. *Measuring Business Cycles*, New York: NBER.

Buse, A. (1994) Brickmaking and the collinear arts: a cautionary tale, *Canadian Journal of Economics*, **27**, 408–14.

Cagan, P. (1956) The Monetary Dynamics of Hyperinflation, in M. Friedman (ed.), *Studies in the Quantity Theory of Money*, Chicago: University of Chicago Press, pp. 25–117.

Canova, F. (1994) Statistical inference in calibrated models, *Journal of Applied Econometrics*, **9**, S123–44.

Canova, F. (1995) Vector autoregressive models: Specification estimation, inference, and forecasting, in M. H. Pesaran and M. R. Wickens (eds), *Handbook of Applied Econometrics: Macroeconomics,* vol. 1, Oxford: Blackwell, pp. 73–138.

Cargill, T. F. (1974) Early applications of spectral methods to economic time series, *History of Political Economy*, **6**, 1–16.

Carnap, R. (1950) Empiricism, semantics, and ontology, *Revue Internationale de Philosophie*, **4**, 20–40.

Cartwright, N. (1988) A case study in realism: Why econometrics is committed to capacities, *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, vol. 2: Symposia and Invited Papers, pp. 190–97.

Cartwright, N. (1995) Probabilities and experiments, *Journal of Econometrics*, **67**, 47–59.

Chamberlain G. and Z. Griliches (1975) Unobservables with a variance-components structure: Ability, schooling, and the economic success of brothers', *International Economic Review*, **16**, 422–50.

Chan, K. H., J. C. Hayya, and J. K. Ord (1977) A note on trend removal methods: The case of polynomial regression versus variate differencing, *Econometrica*, **45**, 737–44.

Chao, J. C. and C. S. Chiao (1998) Testing the expectations theory of the term structure of interest rates using model-selection methods, *Studies in Nonlinear Dynamics and Econometrics*, **12**, 95–108.

Chetty, V. K. (1968) Bayesian analysis of Haavelmo's models. *Econornetrica,* **36**, 582–602.

Chong, Y. Y. and D. F. Hendry (1986) Econometric Evaluation of Linear Macro-Economic Models, *Review of Economic Studies*, **53**, 671–90.

Chow, G. C. (1960) Tests of equality between sets of coefficients in two linear regressions, *Econometrica*, **28**, 591–605.

Chow, G. C. and G. H. Moore (1972) An econometric model of business cycles, in B. G. Hickman (ed.), *Econometric Models of Cyclical Behavior*, New York: NBER, Columbia University Press, pp. 739–812.

Christ, C. F. (1952a) History of the Cowles Commission, 1932–1952, in *Economic Theory and Measurement: A Twenty Year Research Report 1932–1952*. Chicago: Cowles Commission for Research in Economics, pp. 3–65.

Christ, C. F. (1952b) A test of an econometric model for the United States, 1921–1947. *Cowles Commission Papers*, new series No. 49.

Christ, C. F. (1960) Simultaneous equations estimation: Any verdict yet? *Econometrica*, **28**, 835–45.

Christ, C. F. (1966) *Econometric Models and Methods*. New York: Wiley.

Christ, C. F. (1994) The Cowles Commission's contributions to econometrics at Chicago, 1939–1955. *Journal of Economic Literature,* **32**, 30–59.

Christiano, L. J., M. Eichenbaum, and C. Evans (1996) The effects of monetary policy shocks: evidence from the flow of funds. *The Review of Economics and Statistics,* **78**, 16–34.

Clement, M. P. (1995) Rationality and the role of judgement in macroeconomic forecasting, *Economic Journal*, **105**, 410–20.

Clement, M. P. and D. F. Hendry (1994) Towards a theory of economic forecasting, in C. P. Hargreaves (ed.), *Nonstationary Time Series Analysis and Cointegration*, Oxford University Press, pp. 9–52.

Cochrane, D. and G. Orcutt (1949) Application of least squares regression to relationships containing autocorrelated error terms, *Journal of American Statistical Association*, **44**, 32–61.

Cook, S. and D. F. Hendry (1993) The theory of reduction in econometrics, *Poznari Studies in the Philosophy of the Sciences and the Humanities*, **38**, 71–100.

Cooley, T. F. and M. Dwyer (1998) Business cycle analysis without much theory: A look at structural VARs. *Journal of Econometrics*, **83**, 57–88.

Cooley, T. F. and S. LeRoy (1981) Identification and estimation of money demand, *American Economic Review*, **71**, 825–44.

Cooley, T. F. and S. F. LeRoy (1985) Atheoretical macroeconometrics: a critique, *Journal of Monetary Economics* **16**, 283–308.

Cooley, T. F. and E. C. Prescott (1973) An adaptive regression model, *International Economic Review*, **14**, 364–71.

Cooley, T. F. and E. C. Prescott (1976) Estimation in the presence of stochastic parameter variation, *Econometrica*, **44**, 167–84.

Cooper, R. L. (1972) The predictive performance of quarterly econometric models of the United States, in B. G. Hickman (ed.), *Econometric Models of Cyclical Behavior* vol. 2, New York: Columbia University Press, pp. 813–926.

Cowles Commission (1947) *Five–Year Report 1942–46*, Chicago: University of Chicago.

Cox, D. R. (1961) Tests of separate families of hypotheses, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, Berkley: University of California Press, pp. 105–23.

Cox, D. R. (1962) Further results on tests of separate families of hypotheses, *Journal of the Royal Statistical Society Series B,* **24**, 406–24.

Cox, D. R. (1990) Role of models in statistical analysis, *Statistical Science*, **5**, 169–74.

Cramer, J. S. (1969) *Empirical Econometrics*, Amsterdam and London: North-Holland.

Cross, R. (1995) *The Natural Rate of Unemployment: Reflections on 25 Years of the Hypothesis*, Cambridge: University of Cambridge Press.

Davidson, J. E. H., D. F. Hendry, F. Srba, and S. Yeo (1978) Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom, *Economic Journal*, **88**, 661–92.

Dawson, A. (1981) Sargan wage equation—A theoretical and empirical reconstruction. *Applied Economics,* **13**, 351–63.

DeJong, D. N., B. F. Ingram, and C. H. Whiteman (1996) A Bayesian approach to calibration, *Journal of Business and Economic Statistics*, **14**, 1–9.

de Marchi, N. and C. L. Gilbert (eds) (1989) *History and Methodology of Econometrics*, Oxford: Clarendon Press.

Desai, M. (1975) The Phillips curve: A revisionist interpretation. *Economica,* **42**, 1–19.

Desai, M. (1984) Wage, prices and unemployment a quarter century after the Phillips curve, in D. F. Hendry and K. F. Wallis (eds), *Econometrics and Quantitative Economics*, Oxford: Basil Blackwell, pp. 253–73.

Dhrymes, P. J. (1970) *Econometrics: Statistical Foundations and Applications*, New York: Harper & Row.

Dhrymes, P. J., E. P. Howrey, S. H. Hymans, J. Kmenta, E. E. Leamer, R. E. Quandt, J. B. Ramsey, H. T. Shapiro, and V. Zarnowitz (1972) Criteria for evaluation of econometric models, *Annals of Economic and Social Measurement*, **1**, 291–324.

Dickey, D. A. and W. A. Fuller (1979) Distribution for the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association,* **74**, 427–31.

Dicks-Mireaux, L. A. and J. C. R. Dow (1959) The determinants of wage inflation: United Kingdom, 1946–56. *Journal of the Royal Statistical Society. Series A*, **122**, 145–84.

Diebold, F. (2003) The ET interview: Professor Robert F. Engle, *Econometric Theory*, **19**, 1159–93.

Doan, T., R. Litterman, and C. A. Sims (1984) Forecasting and conditional projection using realistic prior distributions, *Econometric Review,* **3**, 1–100.

Doornik, J. A. and D. F. Hendry (2007). Empirical econometric modelling: PcGive 12, vol. I, London: Timberlake Consultants Press.

Dow, J. C. R. and L. A. Dicks-Mireaux (1958) The excess demand for labour: A study of conditions in Great Britain, 1946–56. *Oxford Economic Papers,* **10**, 1–33.

Drèze, J. (1962) The Bayesian approach to simultaneous equations estimation. O.N.R. Research Memorandum 67, Northwestern University; published in D. Qin (ed.) (2013) *The Rise of Econometrics*, vol. 3, London: Routledge, pp. 3–62.

Drèze, J. (1968) Limited Information Estimation from a Bayesian Viewpoint. CORE discussion paper 6816, University of Louvain.

Drèze, J. (1972) Econometrics and decision theory. *Econornetrica,* **40**, 1–17.

Drèze, J. and J.-A. Morales (1976) Bayesian full information analysis of simultaneous equations. *Journal of the American Statistical Association*, **71**, 919–23.

Duesenberry, J. S., G. Fromm, L. R. Klein, and E. Kuh (1965) *The Brookings Quarterly Econometric Model of the United States*. Chicago: Rand McNally.

Duesenberry, J. S., G. Fromm, L. R. Klein, and E. Kuh (1969) *The Brookings Model: Some Further Results*. Chicago: Rand McNally & Company.

Dufour, J.-M. (1982) Recursive stability analysis of linear regression relationships: an exploratory methodology, *Journal of Econometrics*, **19**, 31–76.

Durbin, J. (1954) Errors in variables, *Review of the International Statistical Institute*, **22**, 23–32.

Durbin, J. (1960a) Estimation of parameters in time–series regression models, *Journal of The Royal Statistical Society Series B*, **22**, 139–53.

Durbin, J. (1960b) The fitting of time-series models, *Review of the International Statistical Institute*, **28**, 233–43.

Durbin, J. and G. S. Watson (1950) Testing for serial correlation in least squares regression I, *Biometrica*, **37**, 409–28.

Durbin, J. and G. S. Watson (1951) Testing for serial correlation in least squares regression II. *Biometrika*, **38**, 159–78.

Eichenbaum, M. (1985) Vector autoregressions for causal inference? Comment, *Carnegie-Rochester Conference Series on Public Policy*, **22**, 305–18.

Eichenbaum, M. (1995) Some comments on the role of econometrics in economic theory, *The Economic Journal*, **105**, 1609–21.

Einstein, A. (1922) *Sidelights on Relativity*, London: Methuen & Co.

Elliott, J. W. (1973) A direct comparison of short-run GNP forecasting models, *Journal of Business*, **46**, 33–60.

Engle, R. F. and Granger, C. W. J. (1987) Co-integration and error correction: representation, estimation and testing, *Econometrica,* **55**, 251–76.

Engle, R. F., D. F. Hendry and J.-F. Richard (1983) Exogeneity, *Econometrica,* **51**, 277–304.

Epstein, R. (1987) *A History of Econometrics*, Amsterdam: North-Holland.

Epstein, R. (1989) The fall of OLS in structural estimation. *Oxford Economic Papers*, **41**, 94–107.

Ericsson, N. R. and D. F. Hendry (2004) The ET Interview: Professor David F. Hendry, *Econometric Theory*, **20**, 745–806.

Evans, G. W. and S. Honkapohja (2005) An interview with Thomas J. Sargent. *Macroeconomic Dynamics,* **9**, 561–83.

Evans, M. K. and L. R. Klein (1968) The Wharton econometric forecasting model, *Studies in Quantitative Economics*, No. 2, University of Pennsylvania.

Fair, R. C. (1984) Evaluating the predictive accuracy of models, in K. J. Arrow and M. D. Intriligator (eds), *Handbook of Econometrics*, vol. 3, Amsterdam, North-Holland, pp. 1979–95.

Farrar, D. E. and R. R. Glauber (1967) Multicollinearity in regression analysis: the problem revisited, *Review of Economics and Statistics*, **49**, 92–107.

Faust, J. and C. H. Whiteman (1997) General-to-specific procedures for fitting a data-admissible, theory-inspired, congruent, parsimonious, encompassing weakly-exogenous, identified, structural model to the DGP: A translation and critique, *Carnegie-Rochester Conference Series on Public Policy,* **47**, 121–61.

Fisher, F. M. (1966) *The Identification Problenz in Econometrics*. New York: McGraw-Hill.

Fisher, F. M. and C. Kaysen (1962) *A Study in Econometrics: The Demand for Electricity in the United States*, Amsterdam: North-Holland.

Fisher, I. (1973) I discovered the Phillips curve. *Journal of Political Economy,* **81**, 496–502.

Fisher, W. D. (1962) Estimation in the linear decision model. *International Economic Review*, **3**, 1–29.

Flood, M. M. (1940) Recursive methods in business-cycle analysis, *Econometrica,* **8**, 333–53.

Florens, J.-P., M. Mouchart, and J.-M. Rolin (1990) *Elements of Bayesian Statistics*. New York: Marcel Dekker.

Fomby, T. B. and Hill, R. C. (1979) Multicollinearity and the minimax conditions for the Bock Stein-like estimator, *Econometrica*, **47**, 211–12.

Fox, K. A. (1956) Econometric models of the United States, *Journal of Political Economy,* **64**, 128–42.

Fox, K. A. (1958) *Econometric Analysis for Public Policy*. Ames: Iowa State College Press.

Fox, K. A. (1968) *Intermediate Economic Statistics*. New York: Wiley.

Friedman, M. (1940) Review of Tinbergen's *Business Cycles in the United States*. *American Economic Review,* **30**, 657–60.

Friedman, M. (1957) *A Theory of the Consumption Function*, Princeton: Princeton University Press.

Friedman, M. (1968) The role of monetary policy, *American Economic Reviews,* **58**, 1–17.

Frisch, R. (1933) Propagation problems and impulse problems in dynamic economics, in K. Koch (ed.), *Economic Essays in Honour of Gustav Cassel*, London: Allen and Unwin, pp. 171–205.

Frisch, R. (1934) *Statistical Confluence Analysis by Means of Complete Regression Systems*, Oslo, Universitets Økonomiske Institutt.

Frisch, R. (1938) Autonomy of Economic Relations: Statistical versus Theoretical Relations in Economic Macrodynamics, (Reproduced by University of Oslo in 1948 with Tinbergen's comments), *Memorandum*, Oslo; also in Hendry and Morgan (eds) (1995) *The Foundations of Econometric Analysis*, Cambridge: Cambridge University Press, pp. 407–24.

Frisch, R. (1970) Econometrics in the world of today, in W. A. Eltis, M. F G. Scott, and J. N. Wolfe (eds), *Induction, Growth and Trade: Essays in Honour of Sir Roy Harrod*, Oxford: Clarendon Press, pp. 153–66.

Frisch, R. and F. V. Waugh (1933) Partial time regression as compared with individual trends, *Econometrica*, **1**, 378–401.

Geweke, J. (1986) The superneutrality of money in the United States: An interpretation of the evidence. *Econometrica,* **54**, 1–21.

Geweke, J. (1988) Antithetic acceleration of Monte Carlo integration in Bayesian econometrics, *Journal of Econometrics,* **38**, 73–90.

Geweke, J. (1989) Bayesian inference in econometric models using Monte Carlo integration, *Econometrica*, **57**, 1317–40.

Geweke, J. and K. J. Singleton (1981) Latent variable models for time series: A frequency domain approach with an application to the permanent income hypothesis. *Journal of Econometrics,* **17**, 287–304.

Geweke, J., G. Koop, and H. van Dijk (eds) (2011) *The Oxford Handbook of Bayesian Econometrics*, Oxford: Oxford University Press.

Gilbert, C. L. (1976) The original Phillips curve estimates, *Economica,* **43**, 51–7.

Gilbert, C. L. (1986) Professor Hendry's econometric methodology, *Oxford Bulletin of Economics and Statistics*, **48**, 283–307.

Gilbert, C. L. (1988) The Development of Econometrics in Britain Since 1945, PhD thesis, Oxford University.

Gilbert, C. L. (1989) LSE and the British approach to time series econometrics, *Oxford Economic Papers*, **41**, 108–28.

Gilbert, C. L. (1991a) Stone, Richard, demand theory and the emergence of modern econometrics, *Economic Journal*, **101**, 288–302.

Gilbert, C. L. (1991b) Do economists test theories?—demand analysis and consumption analysis as tests of theories of economic methodology, in N. de Marchi and M. Blaug (eds), *Appraising Economic Theories: Studies in the Methodology of Research Programs*, 137–68, Cheltenham: Edward Elgar.

Gilbert, C. L. and D. Qin (2006) The first fifty years of modern econometrics, in K. Patterson and T. C. Mills (eds), *Palgrave Handbook of Econometrics*, vol. 1. Houndmills: Palgrave Macmillan, pp. 117–55.

Gilbert, C. L. and D. Qin (2007) Representation in econometrics: A historical perspective, in M. Boumans (ed.), *Measurement in Economics: A Handbook*. Amsterdam: Elsevier Inc., pp. 251–69.

Girshick, M. A. and T. Haavelmo (1947) Statistical analysis of the demand for food: Examples of simultaneous estimation of structural equations, *Econometrica*, **15**, 79–110.

Godfrey, L. G. (1978) Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables, *Econometrica*, **46**, 1303–10.

Goldberger, A. S. (1959) *Impact Multipliers and Dynamic Properties of the Klein–Goldberger Model*, Amsterdam: North-Holland Publishing Company.

Goldberger, A. S. (1964) *Econometric Theory*. New York: Wiley.

Goldfeld, S. M. and R.E. Quandt (1973) A Markov model for switching regressions, *Journal of Econometrics*, **1**, 3–16.

Goldstein, M. (1972) The trade-off between inflation and unemployment: a survey of the econometric evidence for selected countries. *International Monetary Fund Staff Working Papers,* **19**, 647–95.

Goodwin, T. H. (1995) Business-cycle analysis with a Markov-switching model, *Journal of Business & Economic Statistics*, **11**, 331–9.

Gordon, R. A. (1949) Business cycles in the interwar period: The 'quantitative-historical' approach, *The American Economic Review*, **39**, Papers and Proceedings of the Sixty-first Annual Meeting, 47–63.

Gordon, R. J. (1970) The Brookings model in action: A review article, *Journal of Political Economy,* **78**, 489–525.

Gordon, R. J. (1990) Comments: The Phillips curve now and then, in P. Diamond (ed.), *Growth, Productivity, Unemployment*. Cambridge, MA: MIT Press, pp. 207–17.

Gordon, R. J. (2011) The history of the Phillips Curve: Consensus and bifurcation, *Economica,* **77** (309), 10–50.

Gordon, R. J., S. King, and F. Modigliani, (1982) The output cost of disinflation in traditional and vector autoregressive models. *Brookings Papers on Economic Activity,* **1**, 205–44.

Granger, C. W. J. (1969) Investigating causal relations by econometric models and cross-spectral methods, *Econometrica,* **37**, 424–38.

Granger, C. W. J. (1981) Time series data and econometric model specification, *Journal of Econometrics,* **16**, 121–30.

Granger, C. W. J. (1983) Cointegrated variables and error correction models. *UCSD Discussion Paper,* 83–13a.

Granger, C. W. J. and M. Hatanaka (1964) *Spectral Analysis of Economic Time Series*, Princeton University Press, Princeton.

Granger, C. W. J. and O. Morgenstern (1963) Spectral Analysis of Stock Market Prices. *Kyklos,* **16**, 1–27.

Granger, C. W. J. and P. Newbold (1973) The time series approach to econometric model building, in *New Methods in Business Cycle Research: Proceedings from a Conference*, Federal Reserve Bank of Minneapolis, pp. 7–21.

Granger, C. W. J. and P. Newbold (1974a) Spurious regression in econometrics, *Journal of Econometrics*, **2**, 111–20.

Granger, C. W. J. and P. Newbold (1974b) Economic forecasting: the atheist's viewpoint, in G. A. Renton (ed.), *Modelling the Economy.* London: Heinemann, pp. 131–47.

Granger, C. W. J. and P. Newbold (1977a) The time series approach to econometric model building, in *New Methods in Business Cycle Research: Proceedings from a Conference*, Federal Reserve Bank of Minneapolis, pp. 7–22.

Granger, C. W. J. and P. Newbold (1977b) *Forecasting Economic Time Series*, New York: Academic Press.

Granger, C. W. J. and A. A. Weiss (1983) Time series analysis of error-correcting models, in K. Samuel, T. Amemiya, and L. A. Goodman (eds), *Studies in Econometrics, Time Series, and Multivariate Statistics*. New York: Academic Press, pp. 255–78.

Granger, C. W. J., M. L. King, and H. White (1995) Comments on testing economic theories and the use of model selection criteria, *Journal of Econometrics*, **67**, 173–87.

Gregory, A. W. and Smith, G. W. (1990) Calibration as estimation, *Econometric Reviews*, **9**, 57–89.

Gregory, A. W. and Smith, G. W. (1995) Business cycle theory and econometrics, *The Economic Journal*, **105**, 1597–608.

Griliches, Z. (1957) Specification bias in estimates of production functions, *Journal of Farm Economics*, **39**, 8–20.

Griliches, Z. (1967) Distributed lags: A survey, *Econometrica,* **35**, 16–49.

Griliches, Z. (1968) The Brookings model volume: A review article. *Review of Economics and Statistics*, **50**, 215–34.

Griliches, Z. and M. D. Intriligator, (1983) *Handbook of Econometrics*, vol. 1, Amsterdam: North-Holland.

Griliches, Z. and M. D. Intriligator (1986) *Handbook of Econometrics*, vol. 3, Amsterdam: North-Holland.

Haavelmo, T. (1938) The method of supplementary confluent relations, illustrated by a study of stock prices, *Econometrica*, **6**, 203–18.

Haavelmo, T. (1939) On the statistical testing of hypotheses in economic theory, published in D. Qin (ed.), (2013) *The Rise of Econometrics*, vol. 2, no. 25, London: Routledge.

Haavelmo, T. (1943) The statistical implications of a system of simultaneous equations, *Econometrica,* **11**, 1–12.

Haavelmo, T. (1944) The probability approach in econometrics, *Econometrica*, **12**, supplement; mimeograph (1941) at Harvard University.

Haavelmo, T. (1947) Methods of measuring the marginal propensity to consume, *Journal of the American Statistical Association*, **42**, 105–22.

Hahn, F. (1970) Some adjustment problems, *Econometrica*, **38**, 1–17.

Hamilton, J. D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, **57**, 357–84.

Hamouda, O. and R. Kowley (1988) *Expectation, Equilibrium and Dynamics*, Brighton: Wheatsheaf Books.

Hansen, B. E. (2005) Challenges for econometric model selection, *Econometric Theory*, **21**, 60–8.

Hansen, L. P. (2004) An interview with Christopher A. Sims. *Macroeconomic Dynamics,* **8**, 273–94.

Hansen, L. P. and Sargent, T. J. (1980) Formulating and estimating dynamic linear rational expectations models, *Journal of Economic Dynamics and Control*, **2**, 7–46.

Hansen, L. P. and T. J. Sargent (1991) Two difficulties in interpreting vector autoregressions, in R. Lucas and T. J. Sargent (eds), *Rational Expectations Econometrics*, Boulder: Westview Press, pp. 77–120.

Harding, D. and A. Pagan (2002) Dissecting the cycles: A methodological investigation, *Journal of Monetary Economics*, **49**, 365–81.

Harkema, R. (1971) Simultaneous Equations: A Bayesian Approach. PhD thesis, University of Rotterdam.

Harvey, A. C. (1985) Trends and cycles in macroeconomic time series, *Journal of Business and Economic Statistics*, **3**, 216–27.

Hausman, J. A. (1978) Specification tests in econometrics, *Econometrica*, **46**, 1251–71.

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. (2001) Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture, *Journal of Political Economy,* **109**, 673–748.

Heckman, J. J. (2008) Econometric causality, *International Statistical Review*, **76** (1), 1–27.

Hendry, D. F. (1971) Maximum likelihood estimation of systems of simultaneous regression equations with errors generated by a vector autoregressive process, *International Economic Review*, **12**, 257–72.

Hendry, D. F. (1974) Stochastic specification in an aggregate demand model of the United Kingdom, *Econometrica*, **42**, 559–78.

Hendry, D. F. (1975) The consequences of mis-specification of dynamic structure, autocorrelation, and simultaneity in a simple model with an application to the demand for imports, in G. A. Renton (ed.), *Modelling the Economy*, London: Social Science Research Council, pp. 286–320 (with discussion).

Hendry, D. F. (1976) The structure of simultaneous equations estimators, *Journal of Econometrics*, **4**, 51–88.

Hendry, D. F. (1979) Predictive failure and econometric modelling in macroeconomics: The transactions demand for money, in P. Ormerod (ed.), *Economic Modelling: Current Issues and Problems in Macroeconomic Modelling in the UK and the US*, London: Heinemann Education Books, pp. 217–42.

Hendry, D. F. (1980) Econometrics: Alchemy or science? *Economica*, **47**, 387–406.

Hendry, D. F. (1985) Monetary economic myth and econometric reality, *Oxford Review of Economy Policy*, **1**, 72–84.

Hendry, D. F. (1986a) Econometric modelling with cointegrated variables: An overview, *Oxford Bulletin of Economics and Statistics*, **48**, 201–12.

Hendry, D. F. (1986b) Using PC-GIVE in econometrics teaching. Oxford Bulletin of Economics and Statistics, **48**, 87–98.

Hendry, D. F. (1987) Econometric methodology: A personal perspective, in T. E. Bewley (ed.), *Advances in Econometrics: Fifth World Congress*, vol. 2, Cambridge: Cambridge University Press, pp. 29–48.

Hendry, D. F. (1995) *Dynamic Econometrics*, Oxford: Oxford University Press.

Hendry, D. F. (2002) Forecast failure, expectations formation and the Lucas critique, *Annales d'Economie et de Statistique*, ENSAE, issue 67–68, 21–40.

Hendry, D. F. (2003) J. Denis Sargan and the origins of LSE econometric methodology, *Econometric Theory*, **19**, 457–80.

Hendry, D. F. (2009) The methodology of empirical econometric modelling: Applied econometrics through the looking-glass, in K. Patterson and T. C. Mills (eds), *Palgrave Handbook of Econometrics*, vol. 2. Houndmills: Palgrave Macmillan, pp. 3–67.

Hendry, D. F. (2011) Empirical economic model discovery and theory evaluation, *Rationality, Markets and Morals*, **46** (2), 115–45.

Hendry, D. F. and G .J. Anderson, (1975) Testing dynamic specification in small simultaneous systems: An application to a model of building society behaviour in the United Kingdom. *Cowles Foundation Discussion Paper 398*; later published in M. D. Intriligator (ed.), *Frontiers in Quantitative Economics*, 3A (1977), Amsterdam: North-Holland, pp. 361–83.

Hendry, D. F. and N. R. Ericsson (1991a) An econometric analysis of U.K. money demand in *Monetary Trends in the United States and the United Kingdom* by Milton Friedman and Anna J. Schwartz, *American Economic Review*, **81**, 8–38.

Hendry, D. F. and N. R. Ericsson (1991b) Modeling the demand for narrow money in the United Kingdom and the United States, *European Economic Review*, **35**: 833–88.

Hendry, D. F. and H.-M. Krolzig (2003) New developments in automatic general-to-specific modelling, in B. P. Stigum (ed.), *Econometrics and the Philosophy of Economics*, Princeton: Princeton University Press, pp. 379–419.

Hendry, D. F. and G. E. Mizon, (1978) Serial correlation as a convenient simplification, not a nuisance: A comment on a study of the demand for money by the Bank of England. *Economic Journal*, **88**, 549–63.

Hendry, D. F. and G. E. Mizon (2000) Reformulating empirical macroeconometric modelling, *Oxford Review of Economic Policy*, **16**, 138–59.

Hendry, D. F. and M. S. Morgan (1989) A Re-analysis of confluence analysis, *Oxford Economic Papers*, **41**, 35–52.

Hendry, D. F. and M. S. Morgan (eds) (1995) *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press.

Hendry, D. F. and J.-F. Richard (1982) On the formulation of empirical models in dynamic econometrics, *Journal of Econometrics,* **20**, 3–33.

Hendry, D. F. and J.-F. Richard (1983) The econometric analysis of economic time series, *International Statistical Review*, **51**, 111–48.

Hendry, D. F. and J.-F. Richard (1989) Recent developments in the theory of encompassing, in B. Cornet and H. Tulkens (eds), *Contributions to Operations Research and Econometrics: The XX anniversary of CORE*, Cambridge, MA: MIT Press, pp. 393–440.

Hendry, D. F. and T. von Ungern-Sternberg (1981) Liquidity and inflation effects on consumers' expenditure, in A. S. Deaton (ed.), *Essays in the Theory and Measurement of Consumers' Behaviour*, Cambridge: Cambridge University Press, pp. 237–61.

Hendry, D. F. and K. F. Wallis (eds.) (1984) *Econometrics and Quantitative Economics*. Oxford: Basil Blackwell.

Hendry, D. F., A. R. Pagan and J. D. Sargan (1984) Dynamic specification, in Z. Griliches and M. D. Intriligator (eds), *Handbook of Econometrics*, vol. 2, Amsterdam: North-Holland, pp. 1023–100.

Hendry, D. F., E. E. Leamer, and D. J. Poirier (1990) A conversation on econometric methodology. *Econometric Theory,* **6**, 171–261.

Hess, G. D., C. S. Jones and R. D. Porter (1998) The predictive failure of the Baba, Hendry and Starr model of M1, *Journal of Economics and Business,* **50**, 477–507.

Hickman, B. G. (1972) *Econometric Models of Cyclical Behavior*, New York, Columbia University Press.

Hildreth, C. (1963) Bayesian statisticians and remote clients. *Econometrica,* **31**, 422–38.

Hildreth, C. (1986) *The Cowles Commission in Chicago 1939–1955*, Berlin: Springer-Verlag.

Hildreth, C. and J. P. Houck (1968) Some estimators for a linear model with random coefficients, *Journal of the American Statistical Association*, **63**, 584–95.

Hill, R. C. and L. C. Adkins (2001) Collinearity, in B. H. Baltagi (ed.), *A Companion to Theoretical Econometrics*, Oxford, Blackwell, 256–78.

Hirsch, J. E. (2005) An index to quantify an individual's scientific output, *Proceedings of the National Academy of Sciences,* **102** (46), 16569–72.

Hodrick, R. and E. C. Prescott (1981) Post-war U.S. business cycles: An empirical investigation, Working Paper, Carnegie-Mellon, University. Reprinted in *Journal of Money*, **29** (1), 1997.

Hoerl, A. E. and Kennard, R. W. (1970a) Ridge regression: biased estimation of nonorthogonal problems, *Technometrics*, **12**, 55–67.

Hoerl, A. E. and Kennard, R. W. (1970b) Ridge regression: applications to nonorthogonal problems, *Technometrics*, **12**, 69–82.

Holt, C., F. Modigliani, J. Muth, and H. Simon (1960) *Planning Production, Inventories and Work Force*, Englewood Cliffs: Prentice-Hall.

Hood, W. and T. Koopmans (eds) (1953) *Studies in Econometric Method*, New York: Cowles Commission Monograph 14.

Hoover, K. D. (2004) Lost causes, *Journal of the History of Economic Thought,* **26**, 149–64.

Hotelling, H. (1927) Differential equations subject to error, *Journal of the American Statistical Association*, **22**, 283–314.

Houthakker, H. S. and L. D. Taylor (1966) *Consumer Demand in the United States*, Harvard Economic Studies, Cambridge: Harvard University.

Howrey, E. P., L. R. Klein, and M. D. McCarthy (1974) Notes on testing the predictive performance of econometric models, *International Economic Review*, **15**, 366–83.

Howson, C. and P. Urbach (1989) *Scientific Reasoning: The Bayesian Approach*. La Salle, IL: Open Court.

Hsiao, C. (1983) Identification, in Z. Griliches and M. D. Intriligator (eds), *Handbook of Econometrics*, vol. 1, Amsterdam: North-Holland, pp. 223–83.

Humphrey, T. M. (1985) The early history of the Phillips curve, *Economic Review*, Federal Reserve Bank of Richmond, **71**, (Sept.), 17–24.

Hurwicz, L. (1944) Stochastic models of economic fluctuations, *Econometrica*, **12**, 114–24.

Hurwicz, L. (1962) On the structural form of interdependent systems, in E. Nagel, P. Suppes, and A. Tarski (eds), *Logic, Methodology and Philosophy of Science*, Proceedings of the 1960 International Congress, Stanford University Press, 232–39.

Huxley, A. (1959) A case of voluntary ignorance, in *Collected Essays*, New York, Harper.

Ilmakunnas, P. and H. Tsurumi (1985) Testing the Lucas hypothesis on output–inflation trade-offs. *Journal of Business and Economic Statistics,* **3**: 43–53.

Intriligator, M. D. (1983) Economic and econometric models, in Z. Griliches and M. D. Intriligator (eds) *Handbook of Econometrics*, vol. 1, Amsterdam: North-Holland, pp. 181–221.

Jacobs, J. (1998) *Econometric Business Cycle Research*. Boston/London: Kluwer Academic Publishers.

Jarque, C. M. and A. K. Bera (1980) Efficient tests for normality, homoscedasiticity and serial independence of regression residuals, *Economic Letters*, **6**, 255–9.

Johansen, S. (1988) Statistical analysis and cointegrating vectors, *Journal of Economic Dynamics and Control*, **12**, 231–54.

Johansen, S. (2006) Cointegration: an overview, in K. Patterson and T. C. Mills (eds), *Palgrave Handbook of Econometrics*. Houndmills: Palgrave Macmillan, pp. 540–77.

Johnston, J. (1963) *Econometric Methods*, New York: McGraw-Hill.

Jorgenson, D. (1963) Capital theory and investment behavior, *American Economic Review*, **53**, 247–59.

Judge, G. G. and M. E. Bock (1978) *The Statistical Implications of Pre-test and Stein-rule Estimators in Econometrics*, Amsterdam: North-Holland.

Judge, G. G., W. E. Griffiths, R. C. Hill, and T.-C. Lee (1980) *The Theory and Practice of Econometrics*, New York: John Wiley & Sons.

Kabaila, P. (1995) The effect of model selection on confidence regions and prediction regions, *Econometric Theory*, **11**, 537–49.

Kadane, J. B. and J. M. Dickey (1980) Bayesian decision theory and the simplification of models, in J. Kmenta and J. Ramsey, J. (eds), *Evaluation of Econometric Models*, New York: Academic Press, pp. 245–68.

Kalman, R. E. (1960) A new approach to linear filtering and prediction problems, *Transactions of the ASME—Journal of Basic Engineering*, **82**, 35–45.

Kaufman, G. M. (1967) Some Bayesian Moment Formulae. CORE discussion paper 6710, University of Louvain.

Kaufman, G. M. (1971) Comments. In M. D. Intriligator (ed.), *Frontiers of Quantitative Economics*, Amsterdam: North-Holland, pp. 208–10.

Kim, K. and A. R. Pagan (1995) The econometric analysis of calibrated models, in M. H. Pesaran and M. Wickens (eds), *Handbook of Applied Econometrics*, Oxford: Blackwell, 356–90.

King, R. G. and Plosser, C. I. (1984) Money, credit and prices in a real business cycle, *American Economic Review*, **74**, 363–80.

King, R. G. and M. W. Watson (1994) The post-war U.S. Phillips curve: a revisionist econometric history, *Carnegie-Rochester Conference Series on Public Policy,* **41**, 157–219.

Klein, L. R. (1947) The use of econometric models as a guide to economic policy, *Econometrica*, **15**, 111–51.

Klein, L. R. (1950) *Economic Fluctuations in the United States 1921–1941*, Cowles Commission Monograph 11, New York.

Klein, L. R. (1953) *A Textbook in Econometrics*, Illinois: Row, Peterson and Company.

Klein, L. R. (1960) The efficiency of estimation of econometric models, in R. W. Pfouts (ed.), *Essays in Economic and Econometrics: A Volume in Honor of Harold Hotelling*. Chapel Hill, NC: University of North Carolina Press, pp. 216–32.

Klein, L. R. (1967) Wage and price determination in macroeconometrics, in A. Phillips and O. E. Williamson (eds), *Prices: Issues in Theory, Practice, and Public Policy*. Philadelphia: University of Pennsylvania Press, pp. 82–100.

Klein, L. R. (1971) Whither econometrics, *Journal of the American Statistical Association*, **66**, 415–21.

Klein, L. R. and R J. Ball (1959) Some econometrics of the determination of absolute prices and wages, *The Economic Journal,* **69**, 465–82.

Klein, R. W. and S. J. Brown (1984) Model selection when there is 'minimal' prior information, *Econometrica*, **52**, 1291–312.

Klein, L. R. and Goldberger, A. S. (1955) *An Econometric Model of the United States 1929–1952*, Amsterdam: North-Holland.

Klein, P. A. and G. H. Moore (1985) *Monitoring Growth Cycles in Market-Oriented Countries: Developing and Using International Economic Indicators*, NBER Book Series in Business Cycles, Cambridge MA: Ballinger Publishing Company.

Klein, L. R. and M. Nakamura (1962) Singularity in the equation systems of econometrics: Some aspects of the problem of multicollinearity, *International Economic Review*, **3**, 274–99.

Klein, L. R., R. J. Ball, A. Hazlewood, and P. Vandome (1961) *An Econometric Model of the United Kingdom*, Oxford: Basil Blackwell.

Kloek, T. and H. K. van Dijk (1978) Bayesian estimates of equation system parameters: An application of integration by Monte Carlo, *Econometrica,* **46**, 1–19.

Kmenta, J. and Ramsey, J. (eds) *Evaluation of Econometric Models*, New York: Academic Press.

Koop, G. (1994) Recent progress in applied Bayesian econometrics, *Journal of Economic Survey,* **8**, 1–34.

Koop, G., D. J. Poirier and J. L. Tobias (2007) *Bayesian Econometric Methods*, Cambridge: Cambridge University Press.

Koopmans, T. C. (1937) *Linear Regression Analysis of Economic Time Series*, Haarlem: Netherlands Economic Institute.

Koopmans, T. C. (1947) Measurement without theory, *Review of Economics and Statistics*, **29**, 161–79.

Koopmans, T. C. (1949a) Reply to Rutledge Vining, *Review of Economics and Statistics*, **31**, 86–91.

Koopmans, T. C. (1949b) The econometric approach to business fluctuations, *The American Economic Review: Papers and Proceedings of the Sixty-first Annual Meeting*, **39**, 64–72.

Koopmans, T. C. (ed.) (1950) *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph 10, New York: Wiley.

Koopmans, T. C. (1957) *Three Essays on the State of Economic Science*, New York: McGraw-Hill.

Koopmans, T. C. and W. C. Hood (1953) The estimation of simultaneous linear economic relationships, in W. C. Hood and T. Koopmans (eds), *Studies In Econometric Method*, Cowles Commission Monograph 14, New Haven, CT: Yale University Press, pp. 112–99.

Kosobud, R. F. (1970) Forecasting accuracy and uses of an econometric model, *Applied Economics*, **2**, 253–63.

Koutsoyiannis, A. (1973) *Theory of Econometrics: An Introductory Exposition of Econometric Methods*, New York: Barnes & Noble Books.

Koyck, L. M. (1954) *Distributed Lags and Investment Analysis*, Amsterdam: North-Holland.

Krugman, P. (2009) How did economists get it so wrong? *The New York Times*, 2 September.

Kuhn, T. S. (1962) *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press.

Kuznets, S. (1946) *National Product Since 1869*, New York: NBER.

Kydland, F. and E. Prescott (1982) Time to build and aggregate fluctuations, *Econometrica,* **50**, 1345–70.

Kydland, F. and E. Prescott (1991) The econometrics of the general equilibrium approach to business cycles. *Scandinavian Journal of Economics,* **93**, 161–78.

Kydland, F. and E. Prescott (1996) The computational experiment: an econometric tool, *Journal of Economic Perspective,* **10**, 69–85.

Laidler, D. (1992) The cycle before new-classical economics, in M. T. Belongia and M. R. Garfinkel (eds), *The Business Cycle: Theories and Evidence: Proceedings of the Sixteenth Annual Economic Policy Conference of the Federal Reserve Bank of St Louis*, Boston: Kluwer, ch. 2.

Lakatos, I. (1977) *The Methodology of Scientific Research Programmes: Philosophical Papers*, vol. 1, ed. J. Worrall and G. Currie, Cambridge: Cambridge University Press.

Lancaster, T. (1968) Grouping estimators on heteroscedastic data, *Journal of the American Statistical Association*, **63**, 182–91.

Leamer, E. E. (1972a) Book review: An introduction to Bayesian inference in econometrics by Arnold Zellner (1971). *Journal of Economic Literature,* **10**, 1232–34.

Leamer, E. E. (1972b) A class of informative priors and distributed lag analysis, *Econometrica,* **40**, 1059–81.

Leamer, E. E. (1973) Multicollinearity: a Bayesian interpretation, *Review of Economics and Statistics*, **55**, 371–80.

Leamer, E. E. (1974) False models and post-data model construction. *Journal of the American Statistical Association,* **69**, 122–31.

Leamer, E. E. (1975) ‘Explaining your results’ as access-biased memory, *Journal of the American Statistical Association,* **70**, 88–93.

Leamer, E. E. (1978a) *Specification Searches: Ad Hoc Inference with Nonexperimental Data*, New York, John Wiley & Sons.

Leamer, E. E. (1978b) Regression selection strategies and revealed priors, *Journal of the American Statistical Association*, **73**, 580–7.

Leamer, E. E. (1983a) Let’s take the con out of econometrics, *The American Economic Review,* **73**, 31–43.

Leamer, E. E. (1983b) Model choice and specification analysis, in Z. Griliches and M. D. Intriligator (eds), *Handbook of Econometrics*, vol. 1, Amsterdam: North-Holland, pp. 285–330.

Leamer, E. E. (1985a) Sensitivity analyses would help, *American Economic Review*, **75**, 308–13.

Leamer, E. E. (1985b) Vector autoregressions for causal inference? *Carnegie-Rochester Conference Series on Public Policy,* **22**, 255–304.

Leamer, E. E. (1986) A Bayesian analysis of the determinants of inflation, in D. A. Belsley and K. Edwin (eds), *Model Reliability*, Cambridge, MA: MIT, pp. 62–89.

Leamer, E. E. (1991) A Bayesian perspective on inference from macroeconomic data, *Scandinavian Journal of Economics,* **93**, 225–48.

Leamer, E. E. and H. Leonard (1983) Reporting the fragility of regression estimates, *Review of Economics and Statistics,* **65**, 306–17.

Leeper, E. M. and C. A. Sims (1994) Towards a modern macroeconomic model usable for policy analysis, *NBER Working Paper Series,* 4761.

Leeson, R. (ed.) (2000) *A.W.H. Phillips: Collected Works in Contemporary Perspective*, Cambridge: Cambridge University Press.

Leijonhufvud, A. (1974) Life among the Econ, *Western Economic Journal*, **11**, 327–37.

Lempers, F. B. (1971) Posterior Probabilities of Alternative Linear Models. PhD thesis, University of Rotterdam.

Leontief, W. (1971) Theoretical assumptions and non-observed facts, *American Economic Review*, **61**, 1–7.

Leser, C. E. V. (1961) A simple method of trend construction, *Journal of the Royal Statistical Society,* Series B, **23**, 91–107.

Leser, C. E. V. (1966) *Econometric Techniques and Problems*. New York: Hafner Publishing Company.

Levenbach, H., J. P. Cleary and D. A. Fryk (1974) A comparison of ARIMA and econometric models for telephone demand, in *Proceedings of the American Statistical Association: Business and Economics Statistics Section*, 448–50.

Lipsey, R. G. (1960) The relationship between unemployment and the rate of change of money wage rates in the U.K., 1862–1957, *Economia*, **27**, 1–31.

Lipsey, R. G. (1978) The place of the Phillips curve in macroeconomic models, in A. Bergstrom, A. Catt, M. Peston, and B. Silverstone (eds), *Stability and Inflation*, Chichester: Wiley, pp. 49–75.

Litterman, R. B. (1979) Techniques of forecasting using vector autoregressions. *Research Department Working Papers of the Federal Reserve Bank of Minneapolis,* 115.

Litterman, R. B. (1986a) Forecasting with Bayesian vector autoregressions—five years of experience, *Journal of Business and Economic Statistics*, **4**, 25–38.

Litterman, R. B. (1986b) Specifying vector autoregressions for macroeconomic forecasting, in P. K. Goel and A. Zellner (eds), *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, Amsterdam: North-Holland, pp. 79–94.

Litterman, R. B. and L. Weiss (1985) Money, real interest rates, and output: A reinterpretation of postwar U.S. data, *Econometrica,* **53**, 129–56.

Liu, T.-C. (1955) A simple forecasting model for the U.S. economy, *International Monetary Fund Staff Papers*, **4**, 434–66.

Liu, T.-C. (1960) Underidentification, structural estimation and forecasting, *Econometrica*, **28**, 855–65.

Liu, T.-C. (1963) An exploratory quarterly econometric model of effective demand in the post-war U.S. economy, *Econometrica*, **31**, 301–48.

Long, J. B. and C. I. Plosser (1983) Real business cycles, *Journal of Political Economy,* **91**, 39–69.

Louçã, F. (2007) *The Years of High Econometrics: A Short History of the Generation that Reinvented Economics*, London: Routledge.

Lovell, M. C. (1983) Data mining, *Review of Economics and Statistics*, **65**, 1–12.

Lubrano, M., R. G. Pierse, and J.-F. Richard (1986) Stability of a U.K. money demand equation: A Bayesian approach to testing exogeneity, *Review of Economic Studies,* **53**, 603–34.

Lucas, R.E. (1972a) Expectations and the neutrality of money. *Journal of Economic Theory* **4**, 103–24.

Lucas, R. E. (1972b) Econometric testing of the natural rate hypothesis, in O. Eckstein (ed.), *The Econometrics of Price Determination Conference*, Washington, DC: Board of Governors of the Federal Reserve System, pp. 50–9.

Lucas, R. E. (1973) Some international evidence on output–inflation tradeoffs, *American Economic Review,* **63**, 326–34.

Lucas, R. E. (1976) Econometric policy evaluation: a critique, in K. Brunner and A. Meltzer (eds), *Stabilization of the Domestic and International Economy*, Amsterdam: North-Holland, pp. 7–29.

Lucas, R. E. and E. C. Prescott (1971) Investment under uncertainty, *Econometrica*, **39**, 659–81.

Lucas, R. E. and L. A. Rapping (1969a) Real wages, employment, and inflation, *Journal of Political Economy,* **77**, 721–54.

Lucas, R. E. and L. A. Rapping (1969b) Price expectations and the Phillips curve, *American Economic Review,* **59**: 342–50.

Lucas, R. E. and T. J. Sargent (1978) After Keynesian macroeconomics, in *After the Philips Curve: Persistence of High Inflation and High Unemployment*, Conference Series no. 19, Federal Reserve Bank of Boston, 44–72.

Lucas, R. E. and Sargent, T. J. (1979) After Keynesian macroeconomics, *Federal Reserve Bank of Minneapolis Quarterly Review,* **3** (Spring), 1–16.

McAleer, M. (1987) Specification tests for separate models: a survey, in M. L. King and D. E. A. Giles (eds), *Specification Analysis in the Linear Model*, London: Routledge & Kegan Paul, pp. 146–96.

McAleer, M., A. R. Pagan, and P. A. Volker (1985) What will take the con out of econometrics? *The American Economic Review,* **75**, 293–307.

McCloskey, D. N. (1983) The rhetoric of economics, *Journal of Economic Literature,* **21**, 481–517.

Machak, J. A., Spivey, W. A., and Wrobleski, W. J. (1985) A framework for time varying parameter regression modeling, *Journal of Business and Economic Statistics*, **3**, 104–11.

MacKinnon, J. G. (1983) Model specification tests against non-nested alternatives, *Econometric Review,* **2**, 85–110.

MacKinnon, J. G. (1992) Model specification tests and artificial regressions, *Journal of Economic Literature*, **30**, 102–46.

MacRoberts, M. H. and B. R. MacRoberts (1989) Problems of citation analysis: A critical review, *Journal of American Society for Information Science,* **40**, 342–9.

Maddala, G. S. (1977) *Econometrics*, Maidenhead: McGraw-Hill.

Maddock, R. (1984) Rational expectations macro theory: a Lakatosian case study in program adjustment, *History of Political Economy,* **16**, 291–309.

Magnus, J. R. and M. S. Morgan (1999) *Methodology and Tacit Knowledge: Two Experiments in Econometrics*, Chichester: Wiley.

Mäki, U. (ed.) (2002) *Fact and Fiction in Economics: Models, Realism and Social Construction*, Cambridge: Cambridge University Press.

Malinvaud, E. (1966) (French edition 1964), *Statistical Methods in Econometrics*, Amsterdam: North-Holland.

Manicas, P. T. (2006) *A Realist Philosophy of Social Science: Explanation and Understanding*, Cambridge: Cambridge University Press.

Mankiw, N. G. (2001) The inexorable and mysterious trade-off between inflation and unemployment, *The Economic Journal,* **111**, C45–61.

Mann, G. S., D. Mimno, and A. McCallum (2006) Bibliometric impact measures leveraging topic analysis, *Joint Conference on Digital Libraries (JCDL'06)*, June, 11–15.

Mann, H. B. and A. Wald (1943) On the statistical treatment of linear stochastic difference equations, *Econometrica,* **11**, 173–220.

Marget, A. W. (1929) Morgenstern on the methodology of economic forecasting, *Journal of Political Economy*, **37**, 132–9.

Marschak, J. (1946) Quantitative studies in economic behaviour (Foundations of rational economic policy), *Report to the Rockefeller Foundation*, New York: Rockefeller Archive Centre.

Marschak, J. (1950) Statistical inference in economics: An introduction, in T. Koopmans (ed.), *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph 10, New York: Wiley, pp. 1–50.

Marschak, J. (1953) Economic Measurements for Policy and Prediction, in W. C. Hood and T. Koopmans (eds), *Studies in Econometric Method*, Cowles Commission Monograph 14, New Haven: Yale University Press, pp. 1–26.

Marschak, J. (1954) Probability in the Social Sciences, Cowles Commission Papers, New Series 82.

Marshall, A. W. (1950) A test of Klein's Model III for changes of structure (abstract), *Econometrica*, **18**, 291.

Massy, W. F. (1965) Principal components regression in exploratory statistical research, *Journal of the American Statistical Association*, **60**, 234–56.

Miller, R. A. (1997) Estimating models of dynamic optimization with microeconomic data, in M. H. Pesaran and P. Schmidt (eds), *Handbook of Applied Econometrics, Vol. II: Microeconomics*, Oxford: Blackwell, pp. 246–99.

Miller, R. W. (1987) *Fact and Method: Explanation, Confirmation and Reality in the Natural and the Social Sciences*, Princeton: Princeton University Press.

Mills, T. C. (2011) Bradford Smith: An econometrician decades ahead of his time, *Oxford Bulletin of Economics and Statistics*, **73**, 276–85.

Mills T. C. and K. Patterson (2006) Editors' introduction, in T. C. Mills and K. Patterson (eds), *Palgrave Handbook of Econometrics*, vol. 1, Houndmills: Palgrave Macmillan, pp. xiii–xxv.

Mintz, I. (1969) *Dating Postwar Business Cycles: Methods and their Application to Western Germany, 1950–67*, NBER Book Series in Business Cycles, New York: Columbia University Press.

Mirowski, P. (1989) *More Heat Than Light: Economics as Social Physics, Physics as Nature's Economics*, Cambridge: Cambridge University Press.

Mitchell, W. C. and A. F. Burns (1938) Statistical indicators of cyclical revivals, NBER; reprinted in W. C. Mitchell and A. F. Burns 1961, *Business Cycle Indicators*, vol. 1, New York: NBER, pp. 162–83.

Mitra-Kahn, B. H. (2008) Debunking the myths of computable general equilibrium models, SCEPA Working Paper 01-2008.

Mitroff, I. (1972) The myth of objectivity or why science needs a new psychology of science, *Management Science*, **18**, B613–18.

Mitroff, I. (1974) *The Subjective Side of Science: A Philosophical Inquiry into the Psychology of the Apollo Moon Scientists*. Amsterdam: Elsevier.

Mizon, G. E. (1977a) Inferential procedures in nonlinear models: An application in a UK industrial cross section study of factor substitution and returns to scale, *Econometrica*, **45**, 1221–42.

Mizon, G. E. (1977b) Model selection procedure, in M. J. Artis and A. R. Nobay (eds), *Studies in Modern Economic Analysis*, Oxford: Basil Blackwell, ch. 4.

Mizon, G. E. (1984) The encompassing approach in econometrics, in D. F. Hendry and K. F. Wallis (eds), *Econometrics and Quantitative Economics*, Oxford: Basil Blackwell, pp. 135–72.

Mizon, G. E. (1995) Progressive modelling of macroeconomic time series: the LSE methodology, in K. D. Hoover (ed.), *Macroeconometrics: Developments, Tensions, and Prospects*, Boston: Kluwer Academic Publishers, pp. 107–70.

Mizon, G. E. and J.-F. Richard (1986) The encompassing principle and its application to testing non-nested hypotheses, *Econometrica*, **54**, 657–78.

Morales, J. A. (1971) *Bayesian Full Information Structural Analysis*. Berlin: Springer-Verlag.

Morgan, M. S. (1990) *The History of Econometric Ideas*, Cambridge: Cambridge University Press.

Morgan, M. S. (1991) The stamping out of process analysis in econometrics, in N. de Marchi and M. Blaug (eds), *Appraising Economic Theories: Studies in the Methodology of Research Programmes*, Cheltenham: Edward Elgar, pp. 237–72.

Morgenstern, O. (1928) *Wirtschaftsprognose: Eine untersuhung ihrer Voraussetzungen und Moglichkeiten*, Vienna: Verlag von Julius Springer.

Morgenstern, O. (1959) *International Financial Transactions and Business Cycles*, Princeton: Princeton University Press.

Morgenstern, O. (1961) A new look at economic time series analysis, in H. Hegeland (ed.), *Money, Growth, and Methodology and other essays in economics: in honor of Johan Akerman*, Lund: CWK Gleerup Publishers, pp. 261–72.

Muth, J. F. (1960) Optimal properties of exponentially weighted forecasts, *Journal of the American Statistical Association*, **55**, 299–306.

Muth, J. F. (1961) Rational expectations and the theory of price movements, *Econometrica,* **29**, 315–35.

Narasimham, G., V. F. Castellino, and N. D. Singpunvalla (1974) On the predictive performance of the BEA quarterly econometric model and a Box–Jenkins type ARIMA model, in *Proceedings of the American Statistical Association: Business and Economics Section*, 501–4.

Neftci, S. N. (1982) Optimal prediction of cyclical downturns, *Journal of Economic Dynamics and Control*, **4**, 225–41.

Neftci, S. N. (1984) Are economic time series asymmetric over the business cycle? *Journal of Political Economy*, **92**, 307–28.

Nelson, C. R. (1972) The prediction performance of the FRB-MIT-PENN model of the U.S. economy. *American Economic Review*, **5**, 902–17.

Nelson, C. R. and C. R. Plosser (1982) Trends and random walks in macroeconomic time series: some evidence and implications, *Journal of Monetary Economics*, **10**, 139–62.

Nerlove, M. (1958) Distributed lags and the estimation of long-run supply and demand elasticities: Theoretical considerations, *Journal of Farm Economics*, **40**, 301–11.

Nerlove, M. and W. Addison (1958) Statistical estimation of long-run elasticities of supply and demand, *Journal of Farm Economics*, **40**, 861–81.

Nicholls, D. F. and A. R. Pagan (1983) Heteroscedasticity in models with lagged dependent variables, *Econometrica*, **51**, 1233–42.

Nickell, S. J. (1985) Error correction, partial adjustment and all that: an expository note, *Oxford Bulletin of Economics and Statistics*, **47**, 119–30.

Nietzsche, F. (1873) On the uses and disadvantages of history for life, *Untimely Meditations*, translated by R. J. Hollingdale (1997), Cambridge: Cambridge University Press.

Olsen, W. and J. Morgan (2005) A critical epistemology of analytical statistics: Addressing the sceptical realist, *Journal for the Theory of Social Behaviour*, **35**, 255–84.

Orcutt, G. (1948) A study of the autoregressive nature of the time series used for Tinbergen's model of the economic system of the United States 1919–1932. *Journal of the Royal Statistical Society Series B* **10**, 1–45.

Orcutt, G. (1952) Toward partial redirection of econometrics, *Review of Economics and Statistics*, **34**, 195–200.

Orcutt, G. and Cochrane, D. (1949) A sampling study of the merits of autoregressive and reduced form transformations in regression analysis, *Journal of the American Statistical Association*, **44**, 356–72.

Oswald, A. J. 1985. The economic theory of trade unions: An introductory survey, *The Scandinavian Journal of Economics,* **87**, 160–93.

Pagan, A. (1987) Three econometric methodologies: A critical appraisal, *Journal of Economic Surveysi,* **1**, 3–24.

Pagan, A. (1995) Three econometric methodologies: An update, in L. Oxley, D. A. R. George, C. L. Roberts, and S. Sayer (eds), *Surveys in Econometrics*, Oxford: Blackwell, pp. 30–41.

Pencavel, J. (1985) Wages and employment under trade unionism: Microeconomic models and macroeconomic applications, *The Scandinavian Journal of Economics,* **87**, 197–225.

Perry, G. L. (1964) The determinants of wage rate changes, *Review of Economic Studies,* **31**, 287–308.

Perry, G. L. (1966) *Unemployment, Money Wage Rates, and Inflation*. Cambridge MA: MIT Press.

Pesaran, M. H. (1974) On the general problem of model selection, *Review of Economic Studies*, **41**, 153–71.

Pesaran, M. H. (1982) A critique of the proposed tests of the natural-rate rational expectations hypothesis, *Economic Journal,* **92**, 529–54.

Pesaran, M. H. (1987) *The Limits to Rational Expectations*. Oxford: Basil Blackwell.

Pesaran, M. H. (1988) The role of theory in applied econometrics, *The Economic Record*, **64**, 336–9.

Pesaran, M. H. and A. S. Deaton (1978) Testing non-nested nonlinear regression models, *Econometrica*, **46**, 677–94.

Phelps, E. S. (1968) Money–wage dynamics and labour market equilibrium, *Journal of Political Economy,* **76**: 678–711.

Phelps Brown, E. H. (1937) Report of the Oxford Meeting, September 25–29, 1936, *Econometrica*, **5**, 361–83.

Phelps Brown, E. H. (1972) The underdevelopment of economics, *Economic Journal*, **82**, 1–10.

Phillips, A. W. (1954) Stabilisation policy in a closed economy, *Economic Journal*, **64**, 290–323.

Phillips, A. W. (1958) The relationship between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957, *Economica,* **25**, 283–99.

Phillips, P. C. B. (1988) Reflections on econometric methodology, *Economic Record*, **64**, 344–59.

Phillips, P. C. B. (1991) Unit roots, *Cowles Foundation Discussion Papers* 998, Cowles Foundation for Research in Economics, Yale University.

Phillips, P. C. B. (1995a) Bayesian model selection and prediction with empirical applications, *Journal of Econometrics,* **69**, 289–331.

Phillips, P. C. B. (1995b) Automated forecasts of Asia-Pacific economic activity, *Asia Pacific Economic Review*, **1**, 92–102.

Phillips, P. C. B. (1996) Econometric model determination, *Econometrica*, **64**, 763–812.

Phillips, P. C. B. (1997) The ET interview: Professor Clive Granger, *Econometric Theory,* **13**, 253–303.

Phillips, P. C. B. (2003) Vision and influence in econometrics: John Denis Sargan, *Econometric Theory*, **19**, 495–511.

Phillips, P. C. B. and J. W. McFarland (1997) Forward exchange market unbiasedness: The case of the Australian dollar since 1984, *Journal of International Money and Finance*, **16**, 885–907.

Phillips, P. C. B. and W. Ploberger (1991) Time series modelling with a Bayesian frame of reference: Concepts, illustrations and asymptotics, Cowles Foundation Discussion paper 980.

Phillips, P. C. B. and W. Ploberger (1994) Posterior odds testing for a unit root with data-based model selection. *Econometric Theory*, **10**, 774–808.

Phillips, P. C. B. and W. Ploberger (1996) An asymptotic theory of Bayesian inference for time series, *Econometrica*, **64**, 381–412.

Phillips, P. C. B. and J. D. Sargan (1985) ET Interviews: Professor J. D. Sargan, *Econometric Theory*, **1**, 119–39.

Poirier, D. J. (1988) Frequentist and subjectivist perspectives on the problems of model building in economics, *Journal of Economic Perspective,* **2**, 121–44.

Pötscher, B. M. (1991) Effects of model selection on inference, *Econometric Theory*, **7**, 163–81.

Press, S. J. (1980) Bayesian computer programs, in A. Zellner (ed.), *Bayesian Analysis in Econometrics and Statistics*, Amsterdam: North-Holland, pp. 429–42.

Qin, D. (1989) Formalisation of identification theory, *Oxford Economic Papers*, **41**, 73–93.

Qin, D. (1993) *The Formation of Econometrics: A Historical Perspective*, Oxford: Clarendon Press.

Qin, D. (1994) Has Bayesian estimation principle ever used Bayes' Rule? Economics discussion paper 318, Queen Mary and Westfield College.

Qin, D. (1996) Bayesian econometrics: The first twenty years, *Econometric Theory,* **12**, 500–16.

Qin, D. (2011a) Rise of VAR modelling approach, *Journal of Economic Surveys*, **25**, 156–74.

Qin, D. (2011b) The Phillips curve from the perspective of the history of econometrics, *History of Political Economy*, **43**, 283–308.

Qin, D. (ed.) (2013) *The Rise of Econometrics*, London: Routledge.

Qin, D. and C. L. Gilbert (2001) The error term in the history of time series econometrics, *Econometric Theory,* **17**, 424–50.

Quah, D. T. (1995) Business cycle empirics: Calibration and estimation, *Economic Journal*, **105**, 1594–96.

Quandt, R. E. (1958) The estimation of the parameters of a linear regression system obeying two separate regimes, *Journal of the American Statistical Association*, **53**, 873–80.

Raj, B. and A. Ullah (1981) *Econometrics: A Varying Coefficients Approach*, London: Croom Helm Ltd.

Ramsey, J. B. (1969) Tests for specification errors in classical least-squares regression analysis, *Journal of the Royal Statistical Society B*, **31**, 350–71.

Reiss, P. C. and F. A. Wolak (2007) Structural econometric modelling: Rationales and examples from industrial organization, in J. J. Heckman and E. E. Leamer (eds), *Handbook of Econometrics*, vol. 6A, Amsterdam: North Holland, pp. 4277–415.

Reschenhofer, E. (1999) Improved estimation of the expected Kullback–Leibler discrepancy in case of misspecification, *Econometric Theory*, **15**, 377–87.

Richard, J.-F. (1980) Models with several regimes and changes in exogeneity, *Review of Economic Studies,* **47**, 1–20.

Richard, J.-F. (1995) Bayesian model selection and prediction with empirical applications discussion, *Journal of Econometrics,* **69**, 337–49.

Rissanen, J. (1987) Stochastic complexity, *Journal of the Royal Statistical Society,* B, **49**, 223–39 & 252–65.

Rissanen, J. (1989) *Stochastic Complexity in Statistical Inquiry*, Singapore: World Scientific Publishing Co.

Romer, C. D. (1994) Remeasuring business cycles, *Journal of Economic History*, **54**, 573–609.

Roose, K. D. (1952) The empirical status of business-cycle theory, *Journal of Political Economy*, **60**, 412–21.

Rothenberg, T. J. (1963) A Bayesian Analysis of Simultaneous Equations Systems. Econometric Institute report 6315, Netherlands School of Economics.

Rothenberg, T. J. (1971) The Bayesian approach and alternatives in econometrics-11, in M. D. Intriligator (ed.), *Frontiers of Quantitative Economics*, Amsterdam: North-Holland, pp. 194–207.

Rothenberg, T. J. (1973) Efficient estimation with a priori information, Cowles Foundation Monograph 23, New Haven, CT: Yale University Press.

Rothenberg, T. J. (1975) Bayesian analysis of simultaneous equations models. In S. E. Fienberg and A. Zellner (eds), *Studies in Bayesian Econometrics and Statistics*, Amsterdam: North-Holland, pp. 405–24.

Rowley, R. and O. Hamouda (1987) Troublesome probability and economics, *Journal of Post Keynesian Economics,* **10**, 44–64.

Ruben, H. (1950) Note on random coefficients, in T. C. Koopmans (ed.), *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph No. 10, pp. 419–21.

Salmon, M. (1982) Error correction mechanisms, *Economic Journal*, **92**, 615–29.

Samuelson, P. and R. Solow (1960) Problem of achieving and maintaining a stable price level: Analytical aspects of anti-inflation policy, *American Economic Review,* **50**, 177–204.

Samuelson, P. and R. Solow (1965) Our menu of policy choices, in A. M. Okun (ed.), *The Battle again Unemployment*, New York: WW Norton, pp. 71–76.

Santomero, A. M. and J. J. Seater (1978) The inflation–unemployment trade-off: A critique of the literature, *Journal of Economic Literature,* **16**, 499–544.

Sargan, D. (1958) The estimation of economic relationships using instrumental variables, *Econometrica*, **26**, 393–415.

Sargan, D. (1959) The estimation of relationships with autocorrelated residuals by the use of instrumental variables, *Journal of the Royal Statistical Society,* Series B, **21**, 91–105.

Sargan, D. (1961) The maximum likelihood estimation of economic relationships with autoregressive residuals, *Econometrica,* **29**, 414–26.

Sargan, D. (1964) Wages and prices in the United Kingdom: a study in econometric methodology, in R. E. Hart, G. Mills, and J. K. Whittaker (eds), *Econometric Analysis for National Economic Planning*, London: Butterworth, pp. 25–63.

Sargan, D. (2003) The development of econometrics at LSE in the last 30 years, *Econometric Theory*, **19**, 429–38.

Sargent, T. J. (1973a) Interest rates and prices in the long run, *Journal of Money, Credit, and Banking,* **5**, 385–450.

Sargent, T. J. (1973b) Rational expectations, the real rate of interest, and the natural rate of unemployment, *Brookings Papers on Economic Activity*, **43**, 429–80.

Sargent, T. J. (1976a) A classical macroeconometric model for the United States, *Journal of Political Economy*, **84**, 207–37.

Sargent, T. J. (1976b) The observational equivalence of natural and unnatural rate theories of macroeconomics, *Journal of Political Economy,* **84**, 631–40.

Sargent, T. J. (1977) The demand for money during hyperinflations under rational expectations, *International Economic Review,* **18**, 59–82.

Sargent, T. J. (1978a) Estimation of dynamic labor demand schedules under rational expectations, *Journal of Political Economy*, **86**, 1009–44.

Sargent, T. J. (1978b) Rational expectations, econometric exogeneity, and consumption, *Journal of Political Economy,* **86**, 673–700.

Sargent, T. J. (1979) Estimating vector autoregressions using methods not based on explicit economic theories, *Federal Reserve Bank of Minneapolis Quarterly Review,* **3** (3), 8–15.

Sargent, T. J. (1980) Rational expectations and the reconstruction of macroeconomics, *Federal Reserve Bank of Minneapolis Quarterly Review*, **4**, 15–9.

Sargent, T. J. (1981) Interpreting economic time series, *Journal of Political Economy,* **89**, 213–47.

Sargent, T. J. and C. A. Sims (1977) Business cycle modelling without pretending to have too much a priori economic theory, in *New Methods in Business Cycle Research: Proceedings from a Conference*, Federal Reserve Bank of Minneapolis, pp. 45–109.

Sargent, T. J. and N. Wallace (1973) Rational expectations and the dynamics of hyperinflation, *International Economic Review,* **14**, 328–50.

Sargent, T. J. and Wallace, N. (1975) 'Rational' expectations, the optimal monetary instrument, and the optimal money supply rule, *Journal of Political Economy*, **83**, 241–54.

Sawa, T. (1978) Information criteria for discriminating among alternative regression models, *Econometrica*, **46**, 1273–91.

Schachter, G. (1973) Some developments in economic science since 1965: Methods, ideas, approaches, *American Journal of Economics and Sociology*, **32**, 331–35.

Schotman, P. C. and H. K. Van Dijk (1991) On Bayesian routes to unit roots, *Journal of Applied Econometrics*, **6**, 387–401.

Schwarz, G. (1978) Estimating the dimension of a model, *Annals of Statistics* **6**, 461–64.

Sent, E.-M. (1997) Sargent versus Simon: bounded rationality unbound, *Cambridge Journal of Economics,* **21**, 323–38.

Sent, E.-M. (1998) *The Evolving Rationality of Rational Expectations: An Assessment of Thomas Sargent's Achievements*, Cambridge: Cambridge University Press.

Sent, E.-M. (2002) How (not) to influence people: the contrary tale of John F. Muth, *History of Political Economy,* **34**, 291–319.

Sheffrin, S. M. (1983) *Rational Expectations*, Cambridge: Cambridge University Press.

Shiller, R. J. (1973) A distributed lag estimator derived from smoothness priors, *Econometrica,* **41**, 775–88.

Silagadze, Z. K. (2009) Citation entropy and research impact estimation, *Acta Physica Polonica*, B41, 2325–33, also available at *Physics and Society* <arxiv.org/abs/0905.1039v1>, accessed 4 February 2013.

Sims, C. A. (1971) Discrete approximations to continuous time distributed lags in econometrics, *Econometrica,* **39**, 545–64.

Sims, C. A. (1972a) The role of approximate prior restrictions in distributed lag estimation, *Journal of the American Statistical Association,* **67**, 169–75.

Sims, C. A. (1972b) Money, income and causality, *American Economic Review,* **62**, 540–52.

Sims, C. A. (1977a) Exogeneity and causal orderings in macroeconomic models, in *New Methods in Business Cycle Research: Proceedings from a Conference*, Federal Reserve Bank of Minneapolis, pp. 23–43.

Sims, C. A. (1977b) Introduction, in *New Methods in Business Cycle Research: Proceedings from a Conference*, Federal Reserve Bank of Minneapolis, pp. 1–5.

Sims, C. A. (1978) Small econometric models of the U.S. and West Germany without prior restrictions, University of Minnesota, Department of Economics Discussion Paper, no. 78–105.

Sims, C. A. (1980a) Macroeconomics and reality, *Econometrica,* **48**, 1–48.

Sims, C. A. (1980b) Comparison of interwar and postwar business cycles: monetarism reconsidered, *American Economic Review,* **70**, 250–57.

Sims, C. A. (1982a) Policy analysis with econometric models, *Brookings Papers on Economic Activity,* **1**, 107–52.

Sims, C. A. (1982b) Scientific standards in econometric modeling, in M. Hazewinkel and A. H. G. Rinnooy Kan (eds) *Current Developments in the Interface: Economics, Econometrics Mathematics*, Holland: D Reidel Publishing Company, pp. 317–37.

Sims, C. A. (1983) Is there a monetary business cycle? *American Economic Review*, **73** (2), Papers and Proceedings of the Ninety-Fifth Annual Meeting of the American Economic Association (May, 1983), pp. 228–33.

Sims, C. A. (1986) Are forecasting models usable for policy analysis? *Quarterly Review of the Federal Reserve Bank of Minneapolis*, Winter, 2–16.

Sims, C. A. (1987) Making economics credible, in T. F. Bewley (ed.), *Advances in Econometrics: Fifth World Congress* vol. II, Cambridge: Cambridge University Press, pp. 49–60.

Sims, C. A. (1988) Bayesian skepticism on unit root econometrics, *Journal of Economic Dynamics and Control*, **12**, 463–74.

Sims, C. A. (1989) Models and their uses, *American Journal of Agricultural Economics,* **71**, 489–94.

Sims, C. A. (1991) Comments: 'Empirical analysis of macroeconomic time series: VAR and structural models', by Michael P. Clement and Grayham E. Mizon, *European Economic Review,* **35**, 922–32.

Sims, C. A. (1993) A nine-variable probabilistic macroeconomic forecasting model, in J. H. Stock and M. W. Watson (eds), *Business Cycles, Indicators, and Forecasting,* Chicago, University of Chicago Press, pp. 179–204.

Sims, C. A. (2000) Using a likelihood perspective to sharpen econometric discourse: Three examples, *Journal of Econometrics*, **95**, 443–62.

Sims, C. A. (2008) Inflation expectations, uncertainty, the Phillips curve, and monetary policy, contribution to *Federal Reserve Bank of Boston Conference*, 10–11 June.

Sims, C. A. and H. Uhlig (1991) Understanding unit rooters: A helicopter tour, *Econometrica*, **59**, 1591–99 (primary version in 1988, Discussion Paper of the Institute for Empirical Macroeconomics 4, Federal Reserve Bank of Minneapolis).

Sims, C. A. and T. Zha (1998) Bayesian methods for dynamic multivariate models, *International Economic Review*, **39**, 949–68.

Sims C. A., J. H. Stock and M. W. Watson (1990) Inference in linear time series models with some unit roots, *Econometrica,* **58**, 113–44.

Sin, C.-Y. and H. White (1996) Information criteria for selecting possibly misspecified parametric models, *Journal of Econometrics*, **71**, 207–25.

Slutsky, E. (1937) The summation of random causes as the source of cyclic processes, *Econometrica*, **5**, 105–46; (originally published in Russian in 1927).

Solow, R. M. (1957) Technical change and the aggregate production function. *Review of Economics and Statistics,* **39**, 312–20.

Spanos, A. (2006) Econometrics in retrospect and prospect, in K. Patterson and T. C. Mills (eds), *Palgrave Handbook of Econometrics*, Houndmills: Palgrave Macmillan, pp. 3–58.

Spanos, A. (2009) Philosophy of econometrics, Department of Economics, Virginia Tech.

Stigum, B. P. (ed.) (2003) *Econometrics and the Philosophy of Economics*, Princeton: Princeton University Press.

Stock, J. H. (1987) Measuring business-cycle time, *Journal of Political Economy*, **95**, 1240–61.

Stock, J. H. and Watson, M. W. (1988) Variable trends in economic time-series, *Journal of Economic Perspectives*, **2**, 147–74.

Stock, J. H. and Watson, M. W. (1989) New indexes of coincident and leading economic indicators, in O. Blanchard and S. Fischer (eds), *NBER Macroeconomic Annual*, Cambridge MA: MIT Press, pp. 351–94.

Stock, J. H. and Watson, M. W. (1993) A procedure for predicting recessions with leading indicators: Econometric issues and recent experience, in J. H. Stock and M. W. Watson (eds), *Business Cycles, Indicators and Forecasting*, Chicago: University of Chicago Press, pp. 95–156.

Stock, J. H. and Watson, M. W. (1999) A comparison of linear and nonlinear models for forecasting macroeconomic time series, in R. F. Engle and H. White (eds), *Cointegration, Causality and Forecasting*, Oxford: Oxford University Press, pp. 1–44.

Stone, R. H. (1954) *The Measurement of Consumers' Expenditure and Behaviour in the United Kingdom 1920–1938*, Cambridge: Cambridge University Press.

Suits, D. B. (1962) Forecasting and analysis with an econometric model, *American Economic Review*, **52**, 104–32.

Summers, L. H. (1991) The scientific illusion in empirical macroeconomics, *Scandinavian Journal of Economics,* **93**, 129–48.

Swann, G. M. P. (2006) *Putting Econometrics in Its Place: A New Direction in Applied Economics*, Cheltenham, Edward Elgar.

Theil, H. (1951) Estimates and their sampling variance of parameters of certain heteroscedastic distributions, *Review of International Statistical Institute*, **19**, 141–7.

Theil, H. (1953) Estimation and simultaneous correlation in complete equation systems, The Hague: Centraal Planbureau, (Mimeographed).

Theil, H. (1957) Specification errors and the estimation of economic relationships, *Review of International Statistical Institute*, **25**, 41–51.

Theil, H. (1961) *Economic Forecasts and Policy*, Amsterdam: North-Holland.

Theil, H. (1963) On the use of incomplete prior information in regression analysis, *Journal of the American Statistical Association,* **58**, 401–14.

Theil, H. (1971) *Principles of Econometrics*, Amsterdam: North-Holland Publishing Co.

Theil, H. and Boot, J. G. G. (1962) The final form of econometric equation systems, *Review of International Statistical Institute*, **30**, 136–52.

Theil, H. and A. S. Goldberger (1961) On pure and mixed statistical estimation in economics, *International Economic Review,* **2**, 65–78.

Thomas, J. J. (1989) The early econometric history of the consumption function, *Oxford Economic Papers*, **41**, 131–49.

Thornber, H. (1967) Finite sample Monte Carlo studies: An autoregressive illustration. *Journal of the American Statistical Association,* **62**, 801–18.

Tiao, G.C. and A. Zellner (1964a) Bayes' theorem and the use of prior knowledge in regression equations. *Biometrika* **51**, 219–30.

Tiao, G. C. and A. Zellner (1964b) On the Bayesian estimation of multivariate regression. *Journal of the Royal Statistical Societyi,* **26**, 277–85.

Tinbergen, J. (1935) Annual survey: suggestions on quantitative business cycle theory, *Econometrica*, **3**, 241–308.

Tinbergen, J. (1937) *An Econometric Approach to Business Cycle Problems*, Paris: Hermann & Cie.

Tinbergen, J. (1939) *Statistical Testing of Business-Cycle Theories*, Geneva: League of Nations.

Tinbergen, J. (1951) *Econometrics*, New York: the Blakiston Company.

Tinbergen, J. (1952) *On the Theory of Economic Policy*, Amsterdam: North-Holland.

Tintner, G. (1938) A note on economic aspects of the theory of errors in time series, *Quarterly Journal of Economics*, **53**, 141–9.

Tintner, G. (1940a) *The Variate Difference Method*, Cowles Commission Monograph No 5, Bloomington.

Tintner, G. (1940b) The analysis of economic time series, *Journal of American Statistical Association*, **35**, 93–100.

Tintner, G. (1952) *Econometrics*, New York: John Wiley & Sons, Inc.

Tobin, J. (1955) Multiple probit regression of dichotomous economic variables, *Cowles Foundation Discussion Paper,* No. 1.

Tobin, J. (1958) Estimation of relationships of limited dependent variables, *Econometrica*, **26**, 24–36.

Trivedi, P. K. (1984) Uncertain prior information and distributed lag analysis, in D. F. Hendry and K. F. Wallis (eds), *Econometrics and Quantitative Economics*, Oxford: Basil Blackwell, pp. 173–210.

Turner, D. S. (1990) The role of judgement in macroeconomic forecasting, *Journal of Forecasting,* **9**, 315–45.

Valavanis, S. (1959) *Econometrics*, New York: McGraw-Hill Book Company, Inc.

Van Dijk, H.K. (2003) On Bayesian structural inference in a simultaneous equation model, in B. P. Stigum (ed.) *Econometrics and Philosophy of Economics*, Princeton, NJ: Princeton University Press, pp. 642–82.

van Raan, A. F. J. (1998) In matters of quantitative studies of science the fault of theorists is offering too little and asking too much, *Scientometrics,* **43**, 129–39.

van Raan, A. F. J. (2004) Measuring science, in H. F. Moed, W Glänzel, and U. Schmoch (eds), *Handbook of Quantitative Science and Technology Research*, Boston: Kluwer Academic Publishers, pp. 19–50.

Vining, R. (1949) Koopmans on the choice of variables to be studied and of methods of measurement, A rejoinder, *Review of Economics and Statistics,* **31**, 77–86; 91–4.

Wagner, H. M. (1958) A Monte Carlo study of estimates of simultaneous linear structural equations, *Econometrica*, **26**, 117–33.

Wald, A. (1936) Calculation and elimination of seasonal fluctuations [Berechnung und ausschaltung von saisonschwankungen], Julius Springer, Vienna, Chapter 1; English translation in D. F. Hendry and M. S. Morgan (eds) (1995) *The Foundations of Econometric Analysis*, Cambridge: Cambridge University Press, pp. 175–9.

Wallace, T. D. (1977) Pretest estimation in regression: A survey, *American Journal of Agricultural Economics*, **59**, 431–43.

Wallis, K. F. (1969) Some recent developments in applied econometrics: Dynamic models and simultaneous equation systems, *Journal of Economic Literature*, **7**, 771–96.

Wallis, K. F. (1973) *Topics in Applied Econometrics*, London: Gray-Mills.

Wallis, K. F. (1980) Econometric implications of the rational expectations hypothesis, *Econometrica*, **48**, 49–73.

Watson, M. W. (1986) Univariate detrending methods with stochastic trends, *Journal of Monetary Economics*, **18**, 49–75.

Watson, M. W. (1993) Measures of fit for calibrated models, *Journal of Political Economy* **101**, 1101–41.

Watson, M. W. (1994) Vector autoregressions and cointegration, in R. F. Engle and D. L. McFadden (eds), *Handbook of Econometrics*, vol. 4, Amsterdam: Elvevier Science, pp. 2843–915.

Waugh, F. V. (1961) The place of Least Squares in econometrics, *Econometrica,* **29**: 386–96.

Wecker, W. E. (1979) Predicting the turning points of a time series, *Journal of Business*, **52**, 35–50.

White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica*, **48**, 817–38.

White, H. (1990) A consistent model selection procedure based on *m*-testing, in C. W. J. Granger (ed.), *Modelling Economic Series*, Oxford: Oxford University Press, pp. 369–83.

Whiteman, C. H. (1983) *Linear Rational Expectations Models: A User's Guide*, Minneapolis: University of Minnesota Press.

Wicksell, K. (1907) Krisernas gåta, *Statsøkonomisk Tidsskrift*, 255–86, Oslo.

Wold, H. O. A. (1938) *A Study in the Analysis of Stationary Time Series*, Uppsala: Almqvist & Wiksells.

Wold, H. O. A. (1954) Causality and econometrics, *Econometrica*, **22**, 162–77.

Wold, H. O. A. (1956) Causal inference from observational data: A review of ends and means, *Journal of Royal Statistical Society Series* A, **119**, 28–61.

Wold, H. O. A. (1960) A generalization of causal chain models, *Econometrica,* **28**, 443–63.

Wold, H. O. A. (ed.) (1964) *Econometric Model Building: Essays on the Causal Chain Approach*, Amsterdam: North-Holland Publishing Company.

Wold, H. O. A. (1967) Book review: C.W.J. Granger (in association with M. Hatanaka) Spectral analysis of economic time series. *The Annals of Mathematical Statistics*, **38**, 288–93.

Wold, H. O. A. and L. Juréen (1953) *Demand Analysis: A Study in Econometrics*, New York: Wiley and Sons.

Wonnacott, R. J. and T. H. Wonnacott (1970) *Econometrics*, New York: Wiley.

Working, E. J. (1954) *The Demand for Meat*, Chicago: Institute of Meat Packing.

Worswick, G. D. N. (1972) Is progress in economics science possible? *Economic Journal*, **82**, 73–86.

Wright, D. M. (1951) Comments on 'Reformulation of current business cycle theories as refutable hypotheses' by Tinbergen, in *Conference on Business Cycles*, New York: NBER, pp. 145–7.

Wulwick, N. L. (1987) The Phillips curve: Which? Whose? To do what? How? *Southern Economic Journal,* **54**: 834–57.

Wynn, R. F. and K. Holden (1974) *An Introduction to Applied Econometric Analysis*, New York: Wiley.

Yitzhaki, S. (1991) Comment on E.E. Leamer, 'A Bayesian perspective on inference from macroeconomic data', *Scandinavian Journal of Economics,* **93**, 259–62.

Young, W. and W. Darity, Jr. (2001) The early history of rational and implicit expectations, *History of Political Economy*, **33**, 773–813.

Yule, G. (1927) On a method of investigating periodicities in disturbed series, with special application to Wolfert's sun spot numbers, *Philosophical Transactions of the Royal Society (Series A)*, **226**, 267–98.

Zarnowitz, V. (1985) Recent work on business cycles in historical perspective: A review of theories and evidence, *Journal of Economic Literature*, **23**, 523–80.

Zarnowitz, V. (1992) *Business Cycles, Theory, Indicators, and Forecasting,* NBER Book Series in Business Cycles, Chicago: University of Chicago Press.

Zellner, A. (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *Journal of the American Statistical Association*, **57**, 348–68.

Zellner, A. (1965) Bayesian inference and simultaneous equation econometric models. Paper presented at the First World Congress of the Econometric Society, Rome.

Zellner, A. (1971a) *An Introduction to Bayesian Inference in Econometrics*, New York: Wiley.

Zellner, A. (1971b) The Bayesian approach and alternatives in econometrics-I, In M. D. Intriligator (ed.), *Frontiers of Quantitative Economics*, Amsterdam: North-Holland, pp. 178–93.

Zellner, A. (1978) Jeffrey–Bayes posterior odds ratio and the Akaike information criterion for discriminating between models, *Economic Letters*, **1**, 337–42.

Zellner, A. (1984) *Basic Issues in Econometrics*, Chicago: University of Chicago Press.

Zellner, A. (2008) Bayesian econometrics: Past, present and future. In S. Chib, W. Griffiths, G. Koop, and D. Terrell (eds), *Bayesian Econometrics,* vol 23 of the series: Advances in Econometrics, Bingley: Emerald Group Publishing Limited, pp. 11–60.

Zellner, A. and V. K. Chetty (1965) Prediction and decision problems in regression models from the Bayesian point of view, *Journal of the American Statistical Association,* **60**, 608–16.

Zellner, A. and M. S. Geisel (1970) Analysis of distributed lag models with applications to consumption function estimation, *Econometrica,* **38**, 856–88.

Zellner, A. and G. C. Tiao (1964) Bayesian analysis of the regression model with autocorrelated errors, *Journal of the American Statistical Association,* **59**, 763–78.

Zellner, A., L. Bauwens, and H. K. Van Dijk (1988) Bayesian specification analysis and estimation of simultaneous equation models using Monte Carlo methods, *Journal of Econometrics,* **38**, 39–72.

Zellner, A., C. Hong, and G. M. Gulati (1990) Turning points in economic time series, loss structures, and Bayesian forecasting, in S. Geisser, J. S. Hodges, S. J. Press, and A. Zellner (eds), *Bayesian and Likelihood Methods in Statistics and Econometrics: Essays in Honor of George A. Barnard*, Amsterdam: North-Holland, pp. 371–93.

This page intentionally left blank

Author index

Subject index