A Short History of Macro-econometric Modelling

David F. Hendry*
Economics Department and Institute for New Economic Thinking
at the Oxford Martin School, University of Oxford, UK

July 24, 2018

Abstract

The key stages in the development of empirical macro-econometric model building are briefly described. Essential steps included characterizing the economy as a system, collating aggregate macroeconomic time series on prices and quantities, isolating the many interacting concepts necessary for understanding how to empirically model economies, inventing viable methods of estimation and inference for dynamic systems, developing hardware and software calculating devices to make such approaches operational, then combining all of these to implement empirical macro-econometric modelling, forecasting and policy making. Its history is littered with both successes and failures, leading overall to substantive progress in understanding, but highlighting the roles of fashions in economic theory dominating empirical evidence and the pernicious impacts of location shifts causing forecast failures and entailing theory failures.

JEL classifications: B22, B23.
KEYWORDS: History of Macroeconomic Modelling; Macro-econometric models; National Accounts; Forecasting, Economic policy.

1 Introduction

A very brief summary of the history of macro-econometric modelling might read as follows. Early empirical economic modelling by Henry Moore (1911, 1914) built on the then recently developed method of regression analysis (Francis Galton, 1886), but was attacked both for a lack of identification and its irrelevance to economics. The first economic forecasting approaches employed the new ideas of business barometers and ABC curves as in Warren Persons (1919, 1924). These prospered with the burgeoning US economy in the period to the late 1920s, then died in the Great Depression. Empirical macroeconomic-system modelling began with the Keynesian revolution, was facilitated by the development of National Accounts and the econometrics tools developed by the Cowles Commission, burnished by success in predicting the effects of the Kennedy 1960's stimulus in the USA, but then was attacked on numerous fronts, especially after forecast failures during the stagflations following the 1970s Oil Crises. Dynamic stochastic general equilibrium models (DSGEs) began life at the time of those oil crises implementing rational expectations and real-business cycle ideas, and should have ended with forecast failures during the Great Recession, but are currently in denial about their problems.

*Financial support from the Institute for New Economic Thinking and Robertson Foundation is gratefully acknowledged, as are helpful comments from Roger Backhouse, Jennifer L. Castle, Jurgen A. Doornik, Pedro G. Duarte, Neil R. Ericsson, Kevin D. Hoover, Katarina Juselius, Andrew B. Martinez, John N.J. Muellbauer, Bent Nielsen and Richard A. Werner. Invited for the *Journal of Banking, Finance and Sustainable Development*. email: david.hendry@nuffield.ox.ac.uk

Nevertheless, digging deeper to understand the evolution of macro-econometric modelling reveals the substantial number of disparate ideas that had to be discerned then combined to create an approach to empirical system modelling that could characterize whole economies, en route inventing the essential ingredients for building macro-econometric models. Moreover, with hindsight, an explanation can be offered for why forecast failure occurs; why it often entails theory failures; what aspects of imperfect models matter when forecasting and what do not; and why the non-stationarity of economic data has to be confronted if useful empirical macroeconomic models are to be built.

The structure of the paper is as follows. After a summary background sketch in Section 2, we track the history in greater detail, seeking to integrate the many strands that led to the initial flourishing of macro-econometric modelling. Section 3 notes some of the early pioneers, especially those producing the first macroeconomic data and models thereof. Section 3.2 discusses some of the basic problems facing any empirical modelling in economics, and how they were gradually understood and solved. Section 4 considers developments during and just after World War II, when both econometric theory developments and national income data bases flourished. Section 5 moves onto the era of big macromodels, then Section 6 investigates the reasons why macro-econometric models intermittently fail. Section 7 briefly notes the continuing life of macro-econometric models in the commercial world. Section 8 discusses the evolving fads and fashions of macroeconomic model building and Section 9 concludes.

2 Summary background

The first step required conceptualizing an economy as an interacting system, implicit in Adam Smith (1776), but first formalised by Leon Walras (1874), and John Maynard Keynes (1936) (with the 'IS-LM' summary in Sir John Hicks, 1937). Proofs of the existence of a static general equilibrium under strong assumptions were provided by Kenneth Arrow and Gerard Debreu (1954), leading in due course to dynamic stochastic general equilibrium models. We refer to those who first characterized economies as systems, as *system conceptualizers* (for histories of economic thought, see Joseph Schumpeter, 1954, and Mark Blaug, 1997, *inter alia*).

A second essential was envisaging, defining and collecting aggregate time-series of macroeconomic data. An appropriate conceptualization of national accounts was first developed by Colin Clark (1932, 1937) and Simon Kuznets (1937, 1946). We refer to those collating and curating time-series of macroeconomic evidence as *aggregate-data creators* (for histories of National Income Accounts, see C.S. Carson, 1970, John Kendrick, 1970, Frits Bos, 2011, and Diane Coyle, 2001, and for price indices, see Erwin Diewert, 1988). However, these first two steps built on even earlier efforts that are the topic of section 3.

Third, to empirically model an economic system, even given a theoretical understanding and sufficient aggregate data, required unravelling a range of inter-related yet distinct concepts, the successes and failures of which are the topics of sections 3.1 and 3.2. Precursors included formalizing models of dynamic relations (see Udny Yule, 1927, and Eugene Slutsky, 1937), understanding the difference between partial, as against simple, correlations (see Ragnar Frisch, 1929), and the need for identification of equations in systems, first for supply and demand curves (see Elmer Working, 1927) and later for complete systems (see Tjalling Koopmans, 1949). We refer to those who isolated the key concepts necessary for understanding how to empirically model economic (and other) systems as *technical solvers* (for histories of economic model building, see e.g., Lawrence Boland, 2014, and of econometrics, see Mary Morgan, 1990, and Duo Qin, 1993, 2013).

Fourth, the resulting idealization of an economy as a dynamic, simultaneous system of stochastic relations in turn required inventing viable statistical tools for estimation and inference: see H.B. Mann and Abraham Wald (1943), Koopmans (1937, 1950a) and Trygve Haavelmo (1944). Much of this path-breaking research took place at the Cowles Foundation (see Edmond Malinvaud, 1988, and Carl Christ, 1994, for histories, and Qin, 2013, for a citation analysis of its enduring influence). We refer to those who invented methods of estimation and inference as *tool makers*, building on a rich tradition in theoretical statistics recorded in Stephen Stigler (1986) and Anders Hald (1998) (for a history of time-series analysis, see Judy Klein, 1997, and the histories of econometrics just noted).

Fifth, to implement these achievements in understanding required two more steps, the first of which was the creation of hardware and software calculating devices. Initially, these were combined in paper 'spreadsheets' for numerical methods of inversion etc., then came mechanical calculators operated by *computors*, usually female, and later electronic computers. William Gossett, aka Student (1908), and Udny Yule (1926) had to undertake their simulation studies manually, and Louis Bean (1929) suggested it took about 8 hours work to calculate a 4-variable regression on 30 data points (for histories of econometric computing, see Laurence Klein, 1987, David Hendry and Jurgen Doornik, 1999, and Charles Renfro, 2009). We refer to those who created the necessary implementation infrastructure as *tool providers*.

Sixth, the second of those two implementation steps was due to *empirical modellers* who combined theory, data, methods and models to create operational empirical systems, drawing on all the above developments. Early examples of estimated macro-econometric systems include Jan Tinbergen (1940, 1951), and Klein and Arthur Goldberger (1955) (for histories of empirical macro-modelling, see Ronald Bodkin, Klein and Kanta Marwah, 1991, Roger Backhouse, 1995, Kevin Hoover, 1995, and Hendry and Morgan, 1995).

Seventh, we have the users of the resulting empirical models, which included *forecasters and policy analysts* in commerce, finance and government (for a history of early economic forecasting, see William Friedman, 2014). Jacob Marschak (1953), Bill Phillips (1954), and Henri Theil (1961) were early contributions to policy analyses based on economic models.

Of course, these developments in empirical macroeconomic modelling did not occur *seriatim*, but overlap greatly and all draw on developments in cognate disciplines, especially mathematics, statistics and computing, and on the intellectual, social and political milieus of their times. A detailed chronology for every relevant conceptualization is too large a task for this paper, but we will consider all of these aspects, albeit briefly. As a *caveat*, it is often unclear precisely what was actually known at the time of each advance and by whom, so it is hard to avoid anachronisms.

3 Early pioneers: the first macroeconomic data and models, 1066–1900

The first recorded 'National Wealth and Income' in the Western World was the Domesday Book ordered by England's King William I around 1087–88. There were just 3 observations (two in 1066 based on data before and after the invasion, and the famous one in 1087), but even so, this was probably the first comprehensive macroeconomic database—and like many later ones, its aim was to facilitate higher taxation: see https://www.nationalarchives.gov.uk/domesday/, analyzed in John McDonald and G.D. Snooks (1986).

John Graunt (1662) was the next English data collector, creating aggregate life tables, followed by Sir Edmond Halley (of comet fame) calculating life annuities. Their friend Sir William Petty (1690) estimated the total numerical 'value' of England's physical and human assets. Gregory King (1696, 1697) then estimated both the population and wealth of England, and summarised its trade and wealth over 1600–1688. King also calculated the first known estimate of the relationship between the supply of corn and its price. Almost a century later, François Quesnay (1766) developed his *Tableau économique*, to map inter-relationships in the economy, perhaps the first attempt at a 'macroeconomic model', with an emphasis on the economy as a system, albeit mainly agriculture. Around the same time, William Playfair (1786) published 43 time-series plots ('time lines', building on earlier work by Joseph Priestley), and a bar chart, a form of graphical presentation he had invented, making his book one of the first to portray statistical graphs in an economic context. This was followed by Playfair (1801) which included one of the first published pie charts.

The next group of pioneers started statistical analyses of economic data that led to building empirical models. Stanley Jevons (1884) investigated movements in currency and finance. As Hendry and Morgan (1995) remark "in the original publication, some of the graphs had cyan shading—nearly a century prior to computer colour graphics and modern printing technology", yet our publisher was unable to reproduce those beautiful fold-out graphs. Louis Bachelier (1900) introduced the idea of a random walk for speculative prices, the first formal model of a non-stationary economic process. Moreover, contemporaneously Reginald Hooker (1901) "explicitly considers non-stationarity due to both stochastic trends and regime shifts as well as deterministic trends; cross serial correlations and lead-lag determination; and issues of model selection when there are multiple correlated causes so that the empirical model has to be discovered from the data" (our quote): perhaps a little anachronistic, but far ahead of his time. Also, Backhouse (1995) discusses late 19th Century writers who saw problems facing economic systems arising from violations of Say's Law, a theme that will prove important after Keynes (1936).

3.1 Early 20th Century developments and debates

Single-variable empirical studies by Marcel Lenoir (1913), R.A. Lehfeldt (1914), Henry Moore (1914, 1925) and Henry Schultz (1928) all estimated demand curves, contemporaneously with Wesley Mitchell (1913) writing on business cycles, setting the scene for the next stage. The results of Moore and Schultz were criticised for lacking both identification and an adequate economic theory basis by Philip Wright (1915, 1929), albeit that Wright sought to be constructive. This was the first of many debates on the value of empirical modelling, 'identification' and the role therein of economic theory, including an intervention by Ragnar Frisch (1933a) (see Hendry and Morgan, 1995).

Tinbergen (1930) produced one of the first empirically estimated econometric systems, although it was not a macromodel. The work was carefully undertaken, and included a new estimator, later known as indirect least squares, designed to take into account the system nature of his study. Tinbergen also developed the first comprehensive national model for the Netherlands in 1936: see Geert Dhaene and Anton Barten (1989), and Barten (1991). The empirical business-cycle system Tinbergen (1940) built for the League of Nations was heavily criticised by Frisch (1938), Keynes (1939) and Milton Friedman (1940) *inter alia*, leading to another famous debate involving Marschak and Oscar Lange (1940) on the defence, discussed respectively by e.g., Hendry (1980) and Neil Ericsson, Hendry and Stedman Hood (2016), as well as Hendry and Morgan (1995). Two key issues for the critics were the idea that empirical evidence might dominate theory, earlier firmly denounced by Lionel Robbins (1932), and the likely lack of 'autonomy' in empirical relations, meaning that changes to the economy would alter estimated equations. Both themes were to continue into the 21st Century.

Indeed, the immediate post-World War II period saw a debate on 'measurement without theory', triggered by Koopmans (1947), but rebuffed by Rutledge Vining (1949a, b) (again see Hendry and Morgan, 1995). The more recent debates about retaining DSGEs despite their inconsistency with empirical evidence, are the latest manifestations of maintaining abstract theory in preference to data coherence: see Adrian Pagan (2003) as compared to Katarina Juselius and Massimo Franchi (2007) and Hoover, Søren Johansen, and Juselius (2008).

3.2 Solving some of the essential problems

Perhaps the greatest difficulty facing early empirical modellers was that many distinct concepts had to be distinguished among a huge range of unknown technical problems. The interaction of problems made it hard to even ascertain what needed to be discovered before it could be solved. The first achievement was isolating then understanding the many interacting issues that determined the properties of observational data generated by dynamic systems, including many of the concepts noted in the introduction, namely: identification in systems, first for supply and demand curves (Elmer Working, 1927); partial, as against simple, correlations (Frisch, 1929); collinearity and confluence in relations (Frisch, 1934); measurement errors versus equation errors (Koopmans, 1937); simultaneity (Haavelmo, 1943); and 'structures' as against 'derived forms' of systems (Tinbergen, 1930, and Koopmans, 1950a).

Another major success was discovering the key features that affected the properties of economic times-series data. These especially included dynamics and autocorrelation (Yule, 1927, and Eugen Slutsky, 1937), with important insights later on their relationship in Denis Sargan (1964), who showed how to discriminate between these two sources, leading to his 'COMFAC' procedure in Sargan (1980), first described in Hendry and Grayham Mizon (1978). In distinct but closely related formulations, Bill Phillips (1954) and Sargan (1964) introduced equilibrium-correction formulations (discussed in Hendry and Mizon, 2000, and Hendry, 2003, respectively). Next came an understanding of nonsense regressions and unit-root non-stationarity (Yule, 1926, and Bradford Smith, 1926). The latter nested models in levels and differences, but somehow this great step forward was missed (see Terence Mills, 2011). Later analyses by Peter Phillips (1986) clarified the problem, and were closely followed by the introduction of cointegration by Rob Engle and Sir Clive Granger (1987) and Johansen (1988), with an extensive analysis in Juselius (2006), closing the circle by their link back to equilibrium correction. Third, location shift non-stationarity was slowly unravelled (Smith, 1929, with the precursor of Hooker, 1901). Location shifts are abrupt changes in the level of a (non-trending) time series, such as the oil-crises jumps in oil prices. Since the 1980s there have been massive advances in understanding and modelling unit-root non-stationarity, less in handling location shifts, although see Jennifer Castle, Doornik, Hendry, and Felix Pretis (2015).

Difficult technical problems also needed to be solved concerning how to estimate macroeconomic systems, with important contributions on multivariate relationships and joint distributions (Haavelmo, 1944); statistical estimation of (stationary) dynamic models by H.B. Mann and Abraham Wald (1943), including estimating systems with autocorrelated errors in Sargan (1959); and the properties of least squares versus instrumental variables (Olav Reiersøl, 1945, and R.C. Geary, 1949, discussed in John Aldrich, 1993). In turn, that last step depended on a coherent concept of an exogenous variable (Koopmans, 1950b, followed up by Engle, Hendry and Jean-François Richard,1983), although ambiguous concepts such as strict exogeneity remain in use. The famous Cowles monographs Koopmans (1950a) and William Hood and Koopmans (1953) essentially assumed that the models to be estimated were known *a priori*, and correctly represented the actual data generating process (DGP). While a necessary simplification to initially develop statistical analyses, the need for model selection as against estimating known perfectly-specified models was addressed by Ted Anderson (1962, 1971)), who demonstrated the substantial advantages of commencing selection from the most general specification under consideration and sequentially simplifying that as against the opposite of simple-to-general. Sargan (1980) had also emphasised general-to-specific (*Gets*), an issue taken up extensively since (see Julia Campos, Ericsson, and Hendry, 2005, for an overview and reprints of many of the key papers). Consequently, methods for comparing models were also required, initiated by Sir David Cox (1961, 1962) on testing non-nested hypotheses, adapted for econometrics by Hashem Pesaran (1974) and developed into encompassing by Mizon and Richard (1986) (see Christophe Bontemps and Mizon, 2010, for a recent overview). Doornik (2008) shows the important role of encompassing in automatic *Gets* modelling. These topics have all seen an explosion in research contributions since the early advances.

A fourth set of issues that generated considerably more heat than light was the role of economic theory in macro-econometrics, already discussed, with an additional emphasis on autonomy (Frisch, 1938, and Bob Lucas, 1976, reviewed in Aldrich, 1989), and the counter that there was 'little empirically-relevant economic theory' (Haavelmo, 1958). All too little progress has occurred on these two issues.

3.3 A long history of forecast failure

All forms of forecasting have had a chequered history, and most historical methods have been discredited, with the very names of those methods becoming tarnished: see Hendry (2001a), who provides more than 25 synonyms in English for 'looking into the future'. Ancient Egyptians tried to foretell harvests from the level reached by the Nile in the flood season, though plagues of locusts could thwart their best efforts. The 'oracles of Delphi' and Nostradamus were early examples of ambiguous forecasters, an approach that many later forecasters probably wished they had followed.

Sir William Petty had discerned what he viewed as a seven-year business cycle, suggesting a possible basis for systematic economic forecasts, although historically, cycles "vary greatly in duration and intensity": see Victor Zarnowitz (2004).

A forecasting industry developed in the USA between 1910–1930, represented by Roger Babson (1909) and Persons (1919, 1924) *inter alia*, but much of it was wiped out by the Great Depression, which it failed to foresee but then thought would end speedily, as recounted by Friedman (2014). Despite that debacle prompting Alfred Cowles to help fund burgeoning efforts for a more scientific approach, namely embryonic econometrics, systematic forecast failure would continue to plague macroeconomic models for the next 85 years. Mike Clements and Hendry (1998) explain the problem as due to unanticipated location shifts hitting equilibrium-correction models, which then persist in trying to return to the previous equilibrium rather than adjusting to the new one. Almost all econometric models are equilibrium correction, including regressions, scalar and vector autoregressions, DSGEs, cointegrated systems, and e.g., autoregressive conditional heteroscedastic error processes—ARCH: see Engle, 1982—generalized or not, so this is a pervasive issue. As well as shifts within equations, other proximate causes of forecast failure include omitting a relevant explanatory variable that shifts, or is $\mathsf{I}(1)$ and starts drifting sufficiently to induce a location shift. We consider this problem facing equilibrium-correction models in more detail in Section 6.

4 Developments during World War II and immediate post-war

During World War II there were major developments in the provision of macroeconomic data. UK National income accounts (NIAs) were created during the War by James Meade and Richard Stone–perhaps following from the calculations in Keynes (1920) of the impossibility of Germany paying the Reparations imposed in the Treaty of Versailles, as well as Keynes's desire to know what resources the UK had available to fight the War. In addition to measuring the 'physical' aspects of national income, Keynes also wanted, but did not get, flow-of-funds accounts: see Stone (1978).

To be useful, NIAs needed systematic compilation of data estimates over time, which were initially often undertaken by private institutions (e.g., NBER and Brookings in the USA), and only later by governments. Originally just calculated annually, there was a gradual development of higher-frequency aggregate data, first quarterly then monthly. In addition, the formulation of social accounting frameworks (see e.g., Stone, 1985), national balance sheets and comprehensive Government accounts allowed a better merging of stocks and flows. In July 2018, the UK Office of National Statistics began releasing rolling monthly and 3-monthly estimates of GDP using output gross value added, so developments continue.

The concept and then construction of appropriate price indices was essential to these steps. Following some earlier attempts, index numbers to represent aggregate prices of bundles of goods relative to a baseline were proposed by Étienne Laspeyres (1871) and Hermann Paasche (1875), using somewhat different weighted averages of price relatives. Despite proposals for improved alternative measures as in Irving Fisher (1921), François Divisia (1926) and the discrete approximation by Leo Törnqvist (1936), and Diewert (1976, 1978) *inter alia*, both Paasche and Laspeyres indices remain in use, usually as chained indices, rather than with intermittent changes to the base period. Nevertheless, the UK still calculates its

Retail Price Index (RPI) which suffers from the 'Carli bias' due to its lack of transitivity, such that when prices return to an earlier level the chained price index does not and stays somewhat higher. Despite its problems, RPI remains in use in the policy arena (e.g., UK student loan repayments). In practice, to maintain consistency between nominal and real aggregates within macroeconomic models, implicit deflators are used, based on the ratio of the nominal aggregate to the total of the real components.

Both aggregates and their implicit deflators are bound to be measured with errors, both in relation to the concept they seek to emulate (say GDP) and even compared to an error-free measure of what they actually correspond to after final revisions. Hendry (1995) considers the impacts on measures of inflation from revisions to price index components, often leading to a lack of cointegration between the original and revised measures. For example, the three aggregate price indices in Hendry (2001b) were only weakly cointegrated, and no pairs cointegrated with a unit coefficient. Nevertheless, James Duffy and Hendry (2017) show that the impact on cointegration relationships of even integrated measurement errors can be 'compensated' by strong trends and/or large location shifts—one of two valuable benefits of such features—whereas cointegration can be hidden by integrated measurement errors when there are no shifts or trends. The other benefit of location shifts is reducing collinearities between highly correlated variables.

Price indices not only sustain relationships between nominal and real variables, they also 'hide' changes over time in index compositions and weights, so models with indices can look constant even when all the components are changing greatly: see Hendry (1996). Aggregation is often seen as a drawback, but this benefit is important, as is that following from the major variance reductions of log transforms on linear aggregates: see Hendry (1995). Laws of large numbers and central limit theorems seem to be valued except when applied to modelling aggregates.

During World War II, there were also a number of important breakthroughs in econometric theory, methods and models: see in particular Haavelmo (1944) and Koopmans (1950a). Although many parts of the puzzle were still (unknowingly) missing, and others were known but ignored or greatly simplified, the increasing availability of time-series of aggregate data plus viable econometric methods and developing computer power set the scene for empirical macro-econometric systems. One of the first 'Keynesian' new generation US models was Klein (1950), followed by the Klein–Goldberger US model in Goldberger (1959). Around the same time, Tinbergen (1951), Klein, James Ball, Arthur Hazlewood and Peter Vandome (1961) then Ball and Terry Burns (1968) all built econometric models of the UK: see Hoover (2003) for more details.

At the same time, another positive development, though not econometric, was MONIAC (Monetary National Income Analogue Computer) built by Bill Phillips in 1949, converting 'swords into plowshares' from parts of scrapped warplanes. This hydraulic 'Phillips machine' illustrated how economies functioned by flows of water between tanks for consumers' expenditure, investment etc., allowing changes in taxes to be simulated. Although trade flows could also be studied by connecting machines, a key feature was that the system was closed (other than accidental water leaks–perhaps a warning about the potential damage from location shifts!), so reduced investment had consequences elsewhere in the economy. Alan Bollard (2016) provides a gripping biography of Phillips life, with a description of MONIAC and its construction, a working version of which still resides in the Reserve Bank of New Zealand.

5 The era of big macro-econometric systems

The era of big macro-econometric models really began after the success in predicting the 'balanced-budget' effects of the Kennedy stimulus in the early 1960s. James Duesenberry, Gary Fromm, Klein and Edwin Kuh (1965, 1969) developed the very large Brookings model and Otto Eckstein, Edward Green and Allen Sinai helped build the Data Resources Inc. (DRI) model.

Then Bert Hickman (1972) guided the FRB-MIT-PENN econometric model, which morphed into the FRB/US model: see Flint Brayton, Andrew Levin, Ralph Tryon, and John Williams (1979). However, the forecast accuracy of the FRB-MIT-PENN model was immediately questioned by e.g., Charles Nelson (1972), who showed that simple 'naive' devices could outperform. At the time, the main reference for the theory of economic forecasting was a chapter Haavelmo (1944) had added after his initial draft. This set out explicitly that a necessary condition for successful forecasting was that the distribution of future outcomes was the same as the in-sample distribution. Essentially, this is also the forecasting framework in Klein (1971). In wide-sense non-stationary processes like economic aggregates, a constant distribution is unlikely to hold for the levels of variables, an issue we return to in section 6. Nevertheless, global macroeconomic models were developed along the same lines as the FRB-MIT-PENN model, beginning with project LINK: see Jean Waelbroeck (1976).

Over the same period, many UK macro-economic models were built, especially by the National Institute for Economic and Social Research (NIESR) (see J.A. Bispham, 1975), the London Business School (LBS) (see Ball, D.B. Boatwright, Burns, P.W.M. Lobban, and G.W. Miller, 1975), and at H.M. Treasury (1980), which reported on a long-running project. The large number of UK models, and their regular revisions, led to the creation of the ESRC Macroeconomic Modelling Bureau directed by Ken Wallis, which produced a series of evaluation volumes: see the 4 reviews in Wallis, Martin Andrews, David Bell, Paul Fisher and John Whitley (1984), through to Wallis, Fisher, Andrew Longbottom, David Turner and Whitley (1987).

Systematic forecast failures after the 1970's Oil Crises and the resulting stagflation, with the 'breakdown' of the 'Phillips curve', originally due to Bill Phillips (1958), led to greatly increased criticism of both the models and their Keynesian theoretical basis. However, Castle and Hendry (2014) show that the real-wage has had a remarkably constant, but non-linear, relationship to inflation, productivity and unemployment for 150 years, so part of the 'breakdown' is due to the simplicity of the asserted statistical relation just between price inflation and unemployment.[1] Nevertheless, the criticisms spawned several alternative approaches, including a revival of the autonomy debate in the form of Lucas's (1976) critique, and an emphasis on 'rational expectations' (after John Muth, 1961), rather than previous 'adaptive' mechanisms. What Lucas failed to realise was his critique applied equally forcibly to theory models based on inter-temporal optimization. Location shifts lead to forecast failure unless properly accounted for, but as they are generally unpredictable, they are almost invariably inconsistent with 'optimizing' economic theory. Consequently, 'rational expectations' becoming systematically biased, and remain so until agents learn the new locations of the relevant distributions. Indeed some derivations of the so-called 'New-Keynesian Phillips Curve' (NKPC) use slight-of-hand to 'prove' that future expectations are unbiased by not dating the expectations operator, then imposing their formulations on data without handling location shifts over their sample. Castle, Doornik, Hendry, and Ragnar Nymoen (2014) show that including indicators for such shifts obliterates the apparent significance of future variables.

'Monetarism', with associated macro-econometric models like those at LBS, harking back to Irving Fisher (1925), also flourished for a while. Friedman and Anna Schwartz (1982) argued strongly for the role of money based on a relatively 'constant velocity'. However, this was rejected by Hendry and Ericsson (1991) for the United Kingdom, and by Ericsson, Hendry, and Hood (2016) for the United States, who found that Friedman had essentially doubled the measured US money stock in 1864 to sustain his claims. As the graph in https://voxeu.org/article/milton-friedman-and-data-adjustment reveals, without his 'adjustment', velocity falls from more than 5 in 1880 to almost 1 in 1930. Although it seems to have been widely assumed that 'money' was difficult to measure, and no single measure had a constant relationship with its determinants, Alvaro Escribano (2004) showed that a broad money demand equation first estimated up to 1970 remained constant to the end of the century. Also Josh Ryan-Collins, Richard

[1] https://voxeu.org/article/real-wage-productivity-nexus provides a less technical discussion, and http://www.timberlake.co.uk/macroeconometrics.html provides an empirically based macro-econometrics textbook.

Werner and Castle (2016) emphasise the important distinction between money and credit, and show the latter is an important missing ingredient in understanding the Financial Crisis.

However, a revival of the ideas in Frisch (1933b) led to real business-cycle models which regarded money as irrelevant (see e.g., Finn Kydland and Robert Prescott, 1990, 1991), and finally to DSGEs, nonetheless oddly beloved by Central Banks (see e.g., Frank Smets and Raf Wouters, 2003). The underlying drive was for 'micro-foundations', seeking to embed macroeconomics in theories of 'individual inter-temporal optimization' with rational expectations: see Pedro Duarte (2009) for greater detail. It is far from obvious *what* 'micro-foundations' to use, although a static all-knowing 'representative agent' has little claim to plausibility. At the opposite end of the spectrum, 'atheoretical' vector autoregressions (VARs) were advocated by Tom Sargent and Chris Sims (1977). This built on the earlier work by Herman Wold (1938, 1949), especially his decomposition theorem for stationary processes, and Maurice Quenouille (1957) on dynamic systems, mediated by the notion of 'causality' in Granger (1969).

However, the Financial Crisis and Great Recession led to systematic forecast failures by most econometric systems including DSGEs, with the abandonment of e.g., the Bank of England's quarterly econometric model (BEQEM, Richard Harrison, Kalin Nikolov, Meghan Quinn, Gareth Ramsay, Andrew Scott, and Ryland Thomas, 2005), replaced by COMPASS (Stephen Burgess, Emilio Fernandez-Corugedo, Charlotta Groth, Harrison, Francesca Monti, Konstantinos Theodoridis, and Matt Waldron, 2013): see the revealing review by Nicholas Fawcett, Riccardo Masolo, Lena Koerber and Waldron https://bankunderground.co.uk/2015/11/20/how-did-the-banks-forecasts-perform-before-during-and-after-the-crisis/ As shown below, the failure of COMPASS was all too predictable,[2] although the events that precipitated the failures were not necessarily forecastable. Omitting real economy–financial sector linkages was an important mistake, exposed by the financial crisis. Moreover, Marjorie Flavin (1981), John Muellbauer (1983) and John Campbell and Greg Mankiw (1989) had already shown that Euler-equation formulations for the key macro-relationship of the consumption function were dominated by more extensive explanations, consistent with James Davidson and Hendry (1981) rejecting the specification of Robert Hall (1978). Thus, in many ways, history is repeating itself...

6 Why macro-econometric models fail

All macro-econometric models suffer from many flaws, but which matter? Here is a partial list of potential problems:

(A) most systems are incomplete specifications with important omitted variables;

(B) most systems involve simplistic dynamics;

(C) and impose incorrect, evolving and overly abstract economic theory;

(D) using restrictive models of expectations;

(E) estimated from mis-measured data;

(F) assuming incorrect exogeneity conditions;

(G) selecting models by naive methods, with all too little evaluation;

(H) usually assuming stationarity, especially constant parameters;

(I) misunderstanding 'identification' as just uniqueness, rather than as a correspondence to reality;

[2] See e.g., https://voxeu.org/article/why-standard-macro-models-fail-crises for a non-technical discussion.

(J) and using linear approximations.

If economic data were stationary, barring gross over-fitting, estimated models would on average forecast about as well as they fitted in-sample, as shown by Preston Miller (1978) and Hendry (1979). Consequently, most of the mistakes in (A)–(G) cannot cause systematic forecast failure for stationary data. However, all policy interventions, technological changes, and intermittent crises etc., ensure that economies are not stationary. Now many of the above flaws matter greatly, albeit depending on the type of non-stationarity: policies can go awry, forecasts fail systematically, and models suffer ex post parameter non-constancy. Surprisingly, however, many changes in parameter values need not cause forecast failure, as the next subsection illustrates, although most will lead to policy analysis failure if the unknowingly altered parameters play a role therein.

Conversely, policy interventions usually involve location shifts, directly creating non-stationarity in contradiction to (H), so unless the subsystem of all relevant relationships is invariant to that shift, a version of the Frisch–Lucas critique will apply and the outcome will not be as anticipated from the scenario analysis. Policy mistakes are avoidable under super exogeneity of the policy variable or invariance of the equations of interest to the intervention. Fortunately, such properties are testable in advance as illustrated by Castle, Hendry, and Andrew Martinez (2017). On (I), Hendry (1995) discusses the three distinct attributes of 'identification', namely uniqueness (which can be achieved artificially), interpretation ('have you identified the demand function?'), and the crucial correspondence to reality, and also (J).

However, a more fundamental misunderstanding is the notion that the theory specification usually determines the model outcome. Certainly the theory *restricts* the outcome to its domain, but the resulting estimates are in fact determined by the underlying DGP. Consequently, the theory-model form is a reduction of that DGP to the set of theory variables, further constrained by the theory's parameter restrictions. Although every set of variables has a local DGP (LDGP) determined by the reductions implemented in eliminating all other relevant variables with their imposed functional forms, the resulting LDGP may be a poor approximation to the actual DGP. Moreover, a further mis-specification can arise by a theory model restricting the relationships between the LDGP variables so ending with a poor representation of even the LDGP. Thus, models derive from exclusions, even if theories try to specify by inclusions that indirectly entail what is excluded, but not transparently to the investigator. As all models arise from reductions of the DGP to the LDGP of their variables, that LDGP represents the maximum knowledge that can be obtained for the given specification. Consequently, the LDGP has to be the target for model selection and any theory compared to, and evaluated against, that. Further, for an estimated LDGP not to be a poor representation of its 'parent' DGP, it should be congruent with the empirical evidence. Section 7 describes an approach that allows the theory model to be retained unaltered when it captures all the relevant information, but enables a better model to be discovered otherwise.

6.1 Autoregressive distributed-lag DGP illustration

The following analysis draws on Hendry and Mizon (2012). Consider a first-order scalar autoregressive DGP:

$$y_t = \mu + \rho y_{t-1} + \gamma z_t + \epsilon_t \tag{1}$$

with a strongly exogenous variable $z_t \sim \mathsf{IN}[\kappa, \sigma_\nu^2]$, denoting an independent Normal distribution with mean κ and variance σ_ν^2, where $\epsilon_t \sim \mathsf{IN}[0, \sigma_\epsilon^2]$ and $|\rho| < 1$. When all the DGP parameters $(\mu, \rho, \gamma, \kappa)$ are known, the forecasts from each $y_{T+h}, h = 1, \ldots, H$ for y_{T+h+1} with known future values of z_t are given by:

$$\widehat{y}_{T+h+1|T+h} = \mu + \rho y_{T+h} + \gamma z_{T+h+1} \tag{2}$$

producing an unbiased forecast:

$$\mathsf{E}\left[\left(y_{T+h+1} - \widehat{y}_{T+h+1|T+h}\right) | y_{T+h}, z_{T+h+1}\right] = 0$$

with the smallest possible forecast-error variance, denoted σ_f^2:

$$\mathsf{V}\left[\left(y_{T+h+1} - \widehat{y}_{T+h+1|T+h}\right)\right] = \sigma_f^2 = \sigma_\epsilon^2.$$

The analysis generalises to $\mathbf{y}_t$ and $\mathbf{z}_t$ being vectors related in an open VAR, which is an equilibrium-correction formulation often used to represent a macro-econometric system.

Estimates of the parameters in (1) will add to the forecast-error variance. In particular, $\widehat{\rho}$ also adds the bias component $(\rho - \mathsf{E}[\widehat{\rho}])y_{T+h}$, where its variance contribution is $\mathsf{V}[\widehat{\rho}]y_{T+h}^2$. Omitting z_t from (2) also adds a bias component and further increases the forecast-error variance, as do shifts in the DGP parameters. But which mistakes actually cause systematic forecast failure? We first demonstrate many mistakes that do not really matter for forecasting, before turning to mistakes that most certainly do cause systematic forecast failure, and how some forecasting devices can help mitigate that problem.

6.1.1 Some mistakes that do not really matter for forecasting

The baseline parameter values used for generating a sample of artificial data from (2) are:
$\rho = 0.8$, $\mu = 0$, $\gamma = 1$, $\kappa = 0$, $\sigma_\epsilon^2 = 1$, $\sigma_\nu^2 = 1$, $T = 40$, with a forecast horizon of ten 1-step ahead forecasts for known future $\{z_{T+h}\}$. All forecast intervals shown by bars are based on $\pm 2\widehat{\sigma}_f$.

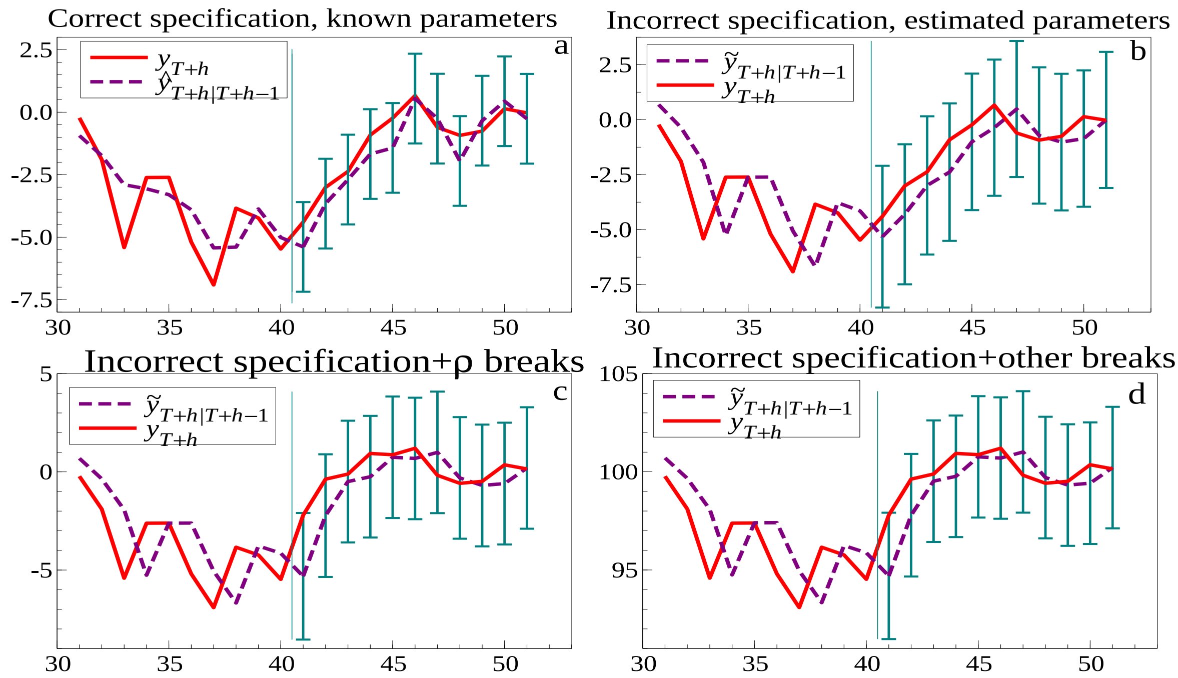

Figure 1: Forecasts of y_{T+h} in four different settings.

Figure 1, panel *a*, illustrates the forecasts from (2) made with **known parameters and known** z_{T+h}. Then panel *b* shows the effects of mis-specification (A) by omitting $\{z_t\}$ from (2) and forecasting using coefficients estimated up to T by:

$$\widetilde{y}_{T+h|T+h-1} = \widehat{\mu} + \widehat{\rho} y_{T+h-1} \quad (3)$$

Next in panel *c*, we add the further complication that ρ unknowingly shifts from $\rho = 0.8$ to $\rho^* = 0.4$ at $T = 41$, then returns to $\rho = 0.8$ at $T = 46$, but we are still forecasting from the now mis-specified model

(3). Finally, in panel *d*, we are still forecasting from (3), but the DGP has changed even more by $\mu = 10$ also changing to $\mu^* = 50$ (i.e., $40\sigma_\epsilon$) at $T = 41$, when $\kappa = 10$, again reverting to $\mu = 10$ at $T = 46$ so the DGP becomes:

$$y_{T+h} = \mu^* + \rho^* y_{T+h-1} + \gamma z_{T+h} + \epsilon_{T+h} \tag{4}$$

As can be seen from Figure 1, there is little difference between the various forecasts and the corresponding values of y_{T+h} other than changes of scale, and no sign of forecast failure from (3) even over the new regime in panel *d*, despite massive changes in μ as well as changes in ρ.

In fact, panels *c* and *d* are identical apart from the scale, as the non-zero values of μ and κ shift the mean of y_t to 100 rather than zero. These outcomes are predictable from the forecast-error taxonomy in Clements and Hendry (1998): **zero-mean effects**, including mis-specifications, mis-estimations and breaks, are non-problems for forecasting. The huge shift of $40\sigma_\epsilon$ in $\mu = 10$ to $\mu^* = 50$ when $\kappa = 10$ is actually a zero-mean change, as can be seen by writing (4) before the shifts as:

$$y_t = 20 + 0.8y_{t-1} + (z_t - 10) + \epsilon_t \tag{5}$$

so that:

$$y_t - 100 = 0.8(y_{t-1} - 100) + (z_t - 10) + \epsilon_t \tag{6}$$

which over the period of the shifts becomes:

$$y_{T+h} = 50 + 0.4y_{T+h-1} + z_{T+h} + \epsilon_{T+h} \tag{7}$$

so that:

$$y_{T+h} - 100 = 0.4(y_{T+h-1} - 100) + (z_{T+h} - 10) + \epsilon_{T+h} \tag{8}$$

so remains mean zero and hence is isomorphic to panel *c* other than the data units being around 100 rather than 0. But location shifts do matter, as we will see after an explanation taken from Hendry and Mizon (2014) of the three main concepts of unpredictability.

6.2 Unpredictability comes in three varieties

A random variable X is unpredictable with respect to some information $\mathcal{I}$ if knowing that information does not change our knowledge about X.

Intrinsic unpredictability in a known distribution arises from chance distribution sampling, 'random errors', etc., although it may matter which draw occurs (e.g., it is not a good strategy to bet on Red at Roulette but get Black). The unpredictability is intrinsic to the random variable, and is unaffected by our behaviour or any additional knowledge: $\epsilon_t \sim \mathsf{IN}[0, \sigma_\epsilon^2]$ is an example of an intrinsically unpredictable random variable.

Instance unpredictability could be called a known unknown, exemplified by an outlier arising from a known 'fat-tailed' distribution, but at an unanticipated time and with an unpredictable sign and magnitude: see Nassim Taleb (2009).

Extrinsic unpredictability corresponds to the over-used 'unknown unknown' deriving from unanticipated shifts of distributions, which can occur in unknown numbers, signs, magnitudes and timings. The most pernicious form of extrinsic unpredictability is that due to unanticipated location shifts which are changes from the previous 'level' of X at unexpected times by unknown amounts. As Figure 2 illustrates, location shifts make what is in fact the new 'ordinary' situation that will persist seem unusual relative to past outcomes. Such shifts have occurred on many occasions historically, and have many potential causes. They are pernicious because they not only wreck econometric modelling if not handled, and induce systematic forecast failure, but also invalidate the very mathematical basis of intertemporal optimization and the assumed result that conditional expectations are minimum mean-square error unbiased predictors: see Hendry and Mizon (2014). As a consequence it becomes irrational to hold 'rational expectations' when such shifts occur, and the law of iterated expectations fails. Specifically, when unanticipated location shifts occur, today's conditional expectation of events tomorrow can be biased, as Figure 2 illustrates, and can be dominated by other predictors: Hendry (2018) and Hendry and John Muellbauer (2018) provide further discussion of this fundamental problem and its implications.

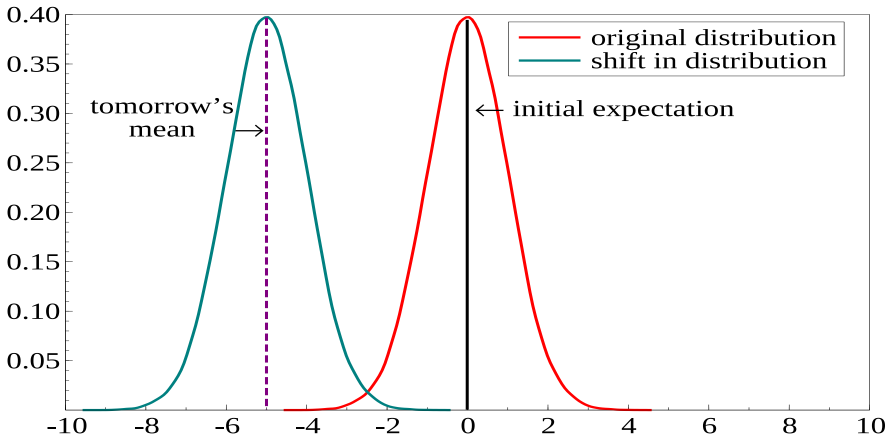

Figure 2: Today's expectation can be a poor estimate of tomorrow's outcome.

Not only do location shifts invalidate key ingredients of modern macroeconomic theory, they also wreck economic agents' ability to plan inter-temporally, so the assumptions in DSGE models about how such agents behave are also wrong, as agents must use 'error-correction mechanisms' after the occurrence of shifts to avoid being wrong systematically. Such shifts are absent from most macroeconomic theories, and consequently, the corresponding models lack adjustment mechanisms, exacerbating the invalid derivations in many theory-based models like COMPASS.

6.3 The key problem for forecasting is an unanticipated location shift

In Figure 3*a*, the forecasting model **is** the DGP in (2), so is correctly specified in-sample with known parameters and no estimation uncertainty, again with known future z_{T+h}, forecasting for the same breaks in ρ as in Figure 1, panel *a*, but now with $\mu = 10$ rather than $\mu = 0$ and still $\kappa = 0$. What could go wrong?

Just catastrophic systematic forecast failure—despite using the known in-sample DGP and known future z_{T+h}. The second through 6th forecasts are all above the **previous outcomes** as emphasized by the upward sloping lines joining them, and the interval forecasts are all well outside the next outcomes till $T = 47$ when the parameters have reverted back to their original values. Thus, the most naive possible forecast, namely extrapolating the previous value, $\widetilde{y}_{T+h|T+h-1} = y_{T+h-1}$, would have outperformed the **in-sample DGP** in forecasting: remember Charles Nelson's result? The key insight is that other than $\mu \neq 0$, this scenario is **identical** to that in Figure 1*a* which showed no forecast failure: when $\mu \neq 0$, changes in ρ cause location shifts, as the mean of y_{T+h} drops from 50 at $h = 0$ to about 16.67 at $h = 5$, then returns to 50 towards the end of the forecast horizon. Because (2) is an equilibrium-correction mechanism, the failure would have continued had the changed parameters persisted.

Panel *b* uses that same DGP and data, but now the forecasting device is the first-difference of the estimated mis-specified autoregression (3):

$$\widetilde{y}_{T+h|T+h-1} = y_{T+h-1} + \widehat{\rho}\Delta y_{T+h-1} \tag{9}$$

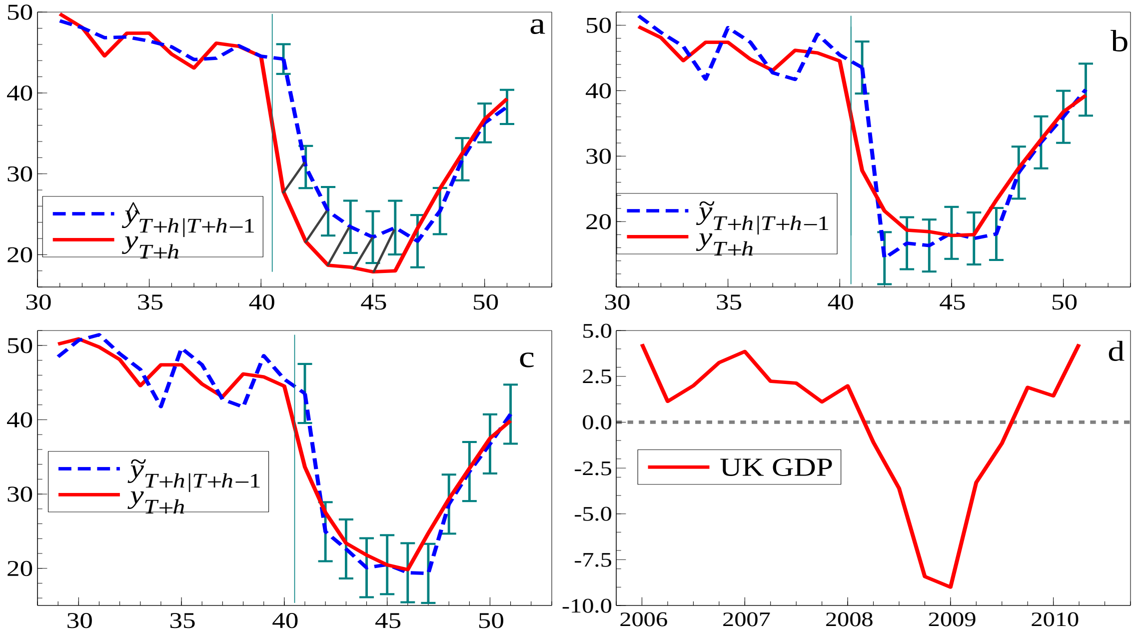

Figure 3: Panel a: forecasts of y_{T+h} from the in-sample DGP; panel b: forecasts from a robust device for the same shifts; panel c: forecasts from a robust device for different shifts; panel d: annualized % change in UK GDP over the Great Recession as a comparison.

shown in Hendry (2006) to be robust after location shifts, extended by Castle, Clements, and Hendry (2015) to a class of devices with varying robustness after location shifts. Panel b shows that despite (9) being a completely wrong model for (2), incorrectly omitting the intercept and z_{T+h}, but including the 'irrelevant' y_{T+h-1} and using the wrong estimate for ρ^*, nevertheless its forecasts are far better than those from the in-sample DGP. The initial mis-forecast is inevitable for all devices that fail to anticipate the location shift, then the device over-shoots as the parameters switch back, but thereafter performs well.

Panel c still uses the DGP (1) to generate data, but with a different constellation of parameter values and changes therein, namely $\mu = 5$; $\rho = 0.8$; $\gamma = 1$; $\kappa = 5$, changed to $\mu^* = 2.5$; $\rho^* = 0.6$; $\gamma^* = 0.86$; $\kappa = 5$, so almost every parameter is altered, again forecasting from (9). Panel c shows that forecasts from (9) are even better despite more DGP parameters shifting, and also illustrates that closely similar location shifts can be created by completely different parameter values and changes therein. Thus, it is possible to create essentially the same $\mathcal{V}$ shape by changing many combinations of parameters in artificial data, similar to annualized % change in UK GDP over the Great Recession shown in Panel d. Neither economists nor economic agents could tell what DGP 'parameters' had shifted till long afterwards as learning new parameters takes time. Hence, even 'rational agents' could not form conditional expectations from past distributions of non-stationary processes subject to location shifts.

Thus, Figure 3a explains systematic forecast failures of macroeconomic models after location shifts, whereas panels b and c explain why the so-called 'naive' devices used by Nelson (1972) could outperform, as they were unknowingly robust after shifts, like (9), as are some other forecasting devices. Together these results emphasize the lack of connection between the 'goodness' of a model, here the actual in-sample DGP with known parameters and known future values of the strongly exogenous variable, and the resulting forecasts: in-sample known DGPs can be beaten in forecasting by estimated naive devices when location shifts occur. Obviously, that possibility does not entail such models are useful in any other context, such as understanding or policy analysis.

The misunderstandings about the causes of forecast failure are legion, and inevitably used by those with political agendas to further their world view, even if the observed failure is unrelated to the quality of the theoretical basis and its empirical implementation. The 'classic' example thereof is the failure of Apollo 13 to arrive on the Moon at the forecast time and date, following the explosion on board of an oxygen cylinder. Neither the underlying theory of Newtonian dynamics based on the inverse square law, nor NASA's forecasting algorithms are impugned by that debacle: the key lesson was to manufacture more robust oxygen cylinders. Thus, abandoning 'Keynesian' macromodels after the Oil crises location shifts was not justified from their forecast failure, nor one might add is that of giving up on DSGEs from their awful performance over the Financial Crisis and Great Recession: what matters is that the latter are among the least 'structural' representations of economic behaviour possible as their very mathematical bases of inter-temporal optimization and conditional expectations fail with every distributional shift, and so cannot represent how intelligent economic agents would behave.

7 Commercialization of macro-econometric models

Although many governmental agencies and central banks have moved to use DSGEs as their main empirical model form, the commercial world has continued to use less restrictive macroeconomic models with more data congruence. Such models were developed to provide forecasts and quantitative economic analyses outside governments. Examples include Wharton Econometric Forecasting Associates and DRI now both part of IHS Markit (see https://ihsmarkit.com/products/economics-analysis-forecasting.html), and Ray Fair (see https://fairmodel.econ.yale.edu/mmm2.htm) in the USA, Oxford Economics (previously Oxford Economic Forecasting: see https://www.oxfordeconomics.com/about-us), Cambridge Econometrics (see https://www.camecon.com/) and National Institute of Economic and Social Research (NIESR: see https://www.niesr.ac.uk/forecasts-models in the UK, and Statistics Norway in Oslo (see https://www.ssb.no/en/forskning/makrookonomi/konjunkturanalyser-og-prognoser) *inter alia*. That companies are willing to pay for such services suggests that forecasts from non-DSGE models remain value for money, but also contradicts the assumption about how the agents in DSGEs form expectations about the future.

8 Fads and fashions in macro-econometric model building

New modes of empirical macro-econometric modelling arise from intellectual developments, then fail from the cold hard reality of intransigent non-stationary data, with heated debates in between. Four phases are discernible historically:
[1] early empirical demand models, which were heavily criticised on 'economic theory' grounds;
[2] first attempts at economic forecasting, which died in the Great Depression;
[3] Keynesian macroeconomic models that were rejected after the oil crises stagflation;
[4] DSGEs, which failed in the Great Recession, but are not yet discarded despite their theoretical and empirical failures.

Macro-econometric model-building fashions sink following forecast failure after major unanticipated location shifts, a situation that is likely to continue until such shifts can be predicted. However, rejection (retention) decisions based solely on forecast failure (success) are often *non-sequiturs*, as Apollo 13 dramatically demonstrated, although protagonists usually also claim theoretical flaws (advantages) in the approaches they wish to discard (support).

But fashion is fickle. Not all central banks switched to making DSGE models their main analytic tool: for example, the Federal Reserve retained its large FRB-US model of the US economy during the fad for New Keynesian DSGEs (NK-DSGEs). Other central banks, including those of Canada, the Netherlands and Australia, have developed non-DSGE models, and at the ECB, non-DSGE models of the five major European economies are under development. The paper by Tom Cusbert and Elizabeth Kendall (2018) introducing MARTIN the new model at the Reserve Bank of Australia note that a "weakness of DSGE models is that they often do not fit the data as well as other models, and the causal mechanisms do not always correspond to how economists and policymakers think the economy really works... The key strength of full-system econometric models like MARTIN is that they are flexible enough to incorporate the causal mechanisms that policymakers believe are important and fit the observable relationships in the data reasonably well." The Pagan diagram (see Adrian Pagan, 2003) apparent trade-off between 'theory consistency' and 'empirical coherence' implicitly assumes a specific theory with which a model must be consistent: given other theories, empirical models matching the data evidence can retain consistency (see e.g., Simon Wren-Lewis, 2018).

However, most macroeconomic models developed so far do not adequately capture the links between the real economy and finance, usually forcing wealth and credit effects into a single net worth measure and ignoring shifts in credit supply, as discussed in Hendry and Muellbauer (2018). In the review by the *Journal of Economic Perspectives* of the state of macroeconomics 10 years after the global financial crisis, Greg Kaplan and Giovanni Violante (2018) criticise the micro-foundations of the NK-DSGE approach. In the same issue, Atif Mian and Amir Sufi (2018) argue that the "credit-driven household demand channel" is crucial for explaining not only the global crisis but economic cycles in many countries in the last 40 years: "expansions in credit supply, operating primarily through household demand, have been an important driver of business cycles". The implications for macro-econometric systems include modelling aggregate data on the household sector's joint consumption and portfolio decisions and measuring shifts in credit conditions. Hendry and Muellbauer (2018) summarise the approach, and Valerie Chauvin and Muellbauer (2018) develop a six-equation template applied to French data on the household sector.

Despite ill-based views like Larry Summers (1991) on the role of empirical evidence in macroeconomics, an 'apolitical' way ahead for macro-econometric model building is by nesting theory-driven and data-driven approaches using automatic model selection (see e.g., Doornik, 2009, available in Doornik and Hendry, 2018). That approach is described in Hendry (2018), building on Hendry and Doornik (2014) and Hendry and Johansen (2015).[3] The theory-model formulation is embedded in a much more general model comprising all alternative explanatory variables plus indicators for outliers and shifts. The explanatory variables are orthogonalised relative to the theory-inspired variables that will be retained without selection, while selecting other candidates at tight significance levels by automatic *Gets* to see if any of the rival hypotheses are relevant. If none matter, the theory-parameter estimates have precisely the same distribution as would be obtained by directly fitting the theory model to data, so there is no cost, with the huge benefit that the model has been stringently evaluated. Conversely, if an excess number of the alternative candidate variables are retained despite a tight significance level, the investigator learns that the theory model is at best incomplete and at worst misleading. A progressive research strategy taking account of empirical evidence is possible even in non-stationary economies.

[3] https://voxeu.org/article/improved-approach-empirical-modelling-0 and https://voxeu.org/article/data-mining-more-variables-observations provide non-technical discussions.

9 Conclusion

The section title refers to the conclusions of this paper, not to the history of macro-econometric modelling. We distinguished seven groups of contributors to the development of empirical macroeconomic models all of whom still have important roles:
(1) *system conceptualizers*, now operating at a global level (see Stephane Dees, Filippo di Mauro, Hashem Pesaran, and Vanessa Smith, 2007, and Ericsson and Erica Reisman, 2012) and trying to comprehend a rapidly evolving world economy subject to many unanticipated shifts and potentially very serious problems from climate change (see e.g., Spencer Weart, 2010);
(2) *aggregate-data creators*, struggling to invent measurement structures for digital economies (see Diane Coyle, 2018;
(3) technical solvers have not played a large role recently, but there are probably important new concepts still to be unravelled;
(4) *tool makers* have proliferated since econometrics was founded as a formal discipline in 1932, developing methods of estimation, inference and model selection for many data types and model forms;
(5) *computors* have mainly morphed into software developers, providing crucial infrastructure for empirical modelling, yet many articles published in major journals still do not record what software was used;
(6) *empirical modellers* now abound across all observational data disciplines, giving different weights to the roles of theory, data, methods and models in their approaches and findings;
(7) *forecasters and policy analysts* seem to appear on television nightly, with *ex post* explanations as to why some events happened that were not foreseen that morning, all too rarely drawing on academic expertise.

This paper offers an incomplete summary of a complicated history with many successes and failures, but undoubted increases in knowledge and understanding overall. Despite such advances, our understanding remains seriously incomplete in all seven roles, and given the non-stationarity of the global economy and environment, may always be incomplete even as greater knowledge accrues. Nevertheless, an explanation was offered for what does and does not cause systematic forecast failure and its implications for theory failure. Specifically, forecast failure results from unpredictable location shifts which invalidate the mathematics of inter-temporal optimization, so the law of iterated expectations fails and 'rational expectations' as usually construed become irrational, being systematically biased and inefficient. Conversely, a model selection approach was described that tackles non-stationarity and combines theory and evidence on an equal footing *ex ante* to implement a progressive research strategy of empirical model discovery and theory evaluation.

References

Aldrich, J. (1989). Autonomy. *Oxford Economic Papers 41*, 15–34.

Aldrich, J. (1993). Reiersøl, Geary and the idea of instrumental variables. *The Economic and Social Review 24*, 247–273.

Anderson, T. W. (1962). The choice of the degree of a polynomial regression as a multiple-decision problem. *Annals of Mathematical Statistics 33*, 255–265.

Anderson, T. W. (1971). *The Statistical Analysis of Time Series.* New York: John Wiley & Sons.

Arrow, K. J. and G. Debreu (1954). Existence of an equilibrium for a competitive economy. *Econometrica 22*, 265–290.

Babson, R. W. (1909). Business barometers used in the accumulation of money. Babson reports inc., Babson College, Wellesley, Mass.

Bachelier, L. (1900). Théorie de la spéculation. *Annales Scientifiques de lÉcole Normale Supérieure 3*, 21–86.

Backhouse, R. (1995). *Interpreting Macroeconomics. Explorations in the History of Macroeconomic Thought.* London: Routledge.

Ball, R. J., D. B. Boatwright, T. Burns, P. W. M. Lobban, and G. W. Miller (1975). The London Business School quarterly econometric model of the U.K. economy. See Renton (1975), pp. 3–37.

Ball, R. J. and T. Burns (1968). An econometric approach to short run analysis of the U.K. economy, 1955–66. *Operational Research Quarterly 19*, 225–256.

Barten, A. P. (1991). The history of Dutch macroeconometric modelling 1936–1986. See Bodkin, Klein, and Marwah (1991), pp. 153–194.

Bean, L. H. (1929). A simplified method of graphic curvilinear correlation. *Journal of the American Statistical Association 24*, 386–397.

Bispham, J. A. (1975). The NIESR model and its behaviour. See Renton (1975), Chapter Appendix to ch. 3.

Blaug, M. (1997). *Economic Theory in Retrospect*. Cambridge: Cambridge University Press.

Bodkin, R. G., L. R. Klein, and K. Marwah (Eds.) (1991). *A History of Macroeconometric Model-building*. Aldershot: Edward Elgar Publishing.

Boland, L. (2014). *Model Building in Economics*. Cambridge: Cambridge University Press.

Bollard, A. (2016). *A Few Hares to Chase*. Oxford: Oxford University Press.

Bontemps, C. and G. E. Mizon (2008). Encompassing: Concepts and implementation. *Oxford Bulletin of Economics and Statistics 70*, 721–750.

Bos, F. (2011). Three centuries of macro-economic statistics. *SSRN Electronic journal*, DOI: 10.2139/ssrn.1971604.

Brayton, F., A. Levin, R. Lyon, and J. C. Williams (1997). The evolution of macro models at the Federal Reserve Board. *Carnegie-Rochester Conference Series on Public Policy 47*, 43–81.

Burgess, S., E. Fernandez-Corugedo, C. Groth, R. Harrison, F. Monti, K. Theodoridis, and M. Waldron (2013). The Bank of England's forecasting platform: COMPASS, MAPS, EASE and the suite of models. Working paper no. 471 and appendices, Bank of England, London.

Campbell, J. Y. and N. G. Mankiw (1990). Permanent income, current income, and consumption. *Journal of Business and Economic Statistics 8*, 265–279.

Campos, J., N. R. Ericsson, and D. F. Hendry (Eds.) (2005). *Readings on General-to-Specific Modeling*. Cheltenham: Edward Elgar.

Carson, C. S. (1970). The history of the United States national income and product accounts: The development of an analytical tool. Technical report, Bureau of Economic Analysis, US Department of Commerce.

Castle, J. L., M. P. Clements, and D. F. Hendry (2015). Robust approaches to forecasting. *International Journal of Forecasting 31*, 99–112.

Castle, J. L., J. A. Doornik, D. F. Hendry, and R. Nymoen (2014). Mis-specification testing: Non-invariance of expectations models of inflation. *Econometric Reviews 33*, 553–574.

Castle, J. L., J. A. Doornik, D. F. Hendry, and F. Pretis (2015). Detecting location shifts during model selection by step-indicator saturation. *Econometrics 3(2)*, 240–264.

Castle, J. L. and D. F. Hendry (2014). Semi-automatic non-linear model selection. In N. Haldrup, M. Meitz, and P. Saikkonen (Eds.), *Essays in Nonlinear Time Series Econometrics*, pp. 163–197. Oxford: Oxford University Press.

Castle, J. L., D. F. Hendry, and A. B. Martinez (2017). Evaluating forecasts, narratives and policy using a test of invariance. *Econometrics 5(39)*, DOI:10.3390/econometrics5030039.

Chauvin, V. and J. N. J. Muellbauer (2018). Consumption, household portfolios and the housing market in France. *Economie et Statistique/Economics and Statistics*, forthcoming.

Christ, C. F. (1994). The Cowles Commission's contributions to econometrics at Chicago, 1939-1955. *Journal of Economic Literature 32*, 30–59.

Clark, C. (1932). *The National Income 1924-31*. London: MacMillan.

Clark, C. (1937). *National Income and Outlay*. London: MacMillan.

Clements, M. P. and D. F. Hendry (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.

Cox, D. R. (1961). Tests of separate families of hypotheses. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1, Berkeley, pp. 105–123. University of California Press.

Cox, D. R. (1962). Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society B, 24*, 406–424.

Coyle, D. (2001). *GDP: A brief but Affectionate History*. Princeton, NJ: Princeton University Press.

Coyle, D. (2018). Challenges in measuring the modern economy. Plenary presentation at 2018 International Symposium on Forecasting, Boulder, Colorado.

Cusbert, T. and E. Kendall (2018). Meet MARTIN, the RBA's new macroeconomic model. *Australian Reserve Bank Bulletin March*, 31–44.

Davidson, J. E. H. and D. F. Hendry (1981). Interpreting econometric evidence: The behaviour of consumers' expenditure in the UK. *European Economic Review 16*, 177–192. Reprinted in Campos, J., Ericsson, N.R. and Hendry, D.F. (eds.), *General to Specific Modelling*. Edward Elgar, 2005.

Dees, S., F. di Mauro, M. H. Pesaran, , and L. V. Smith (2007). Exploring the international linkages in the EURO area: A global VAR analysis. *Journal of Applied Econometrics 22*, 1–38.

Dhaene, G. and A. P. Barten (1989). When it all began: The 1936 Tinbergen model revisited. *Economic Modelling 6*, 203–219.

Diewert, W. E. (1976). Exact and superlative index numbers. *Journal of Econometrics 4*, 115–145.

Diewert, W. E. (1978). Superlative index numbers and consistency in aggregation. *Econometrica 46*, 883–900.

Diewert, W. E. (1988). *The Early History of Price Index Research*. Cambridge, MA: NBER.

Divisia, F. (1926). L'indice monétaire et la théorie de la monnaie. *Revue d'Économie Politique 40*, 49–81.

Doornik, J. A. (2008). Encompassing and automatic model selection. *Oxford Bulletin of Economics and Statistics 70*, 915–925.

Doornik, J. A. (2009). Autometrics. In J. L. Castle and N. Shephard (Eds.), *The Methodology and Practice of Econometrics*, pp. 88–121. Oxford: Oxford University Press.

Doornik, J. A. and D. F. Hendry (2018). *Empirical Econometric Modelling using PcGive: Volume I.* (8th ed.). London: Timberlake Consultants Press.

Duarte, P. (2009). Not going away? Microfoundations in the making of a new consensus in macroeconomics. Mimeo, Department of Economics (FEA-USP), University of São Paulo.

Duesenberry, J. S., G. Fromm, L. R. Klein, and E. Kuh (Eds.) (1969). *The Brookings Model: Some Further Results*. Amsterdam: North-Holland.

Duesenberry, J. S., L. R. Klein, G. Fromm, and E. Kuh (Eds.) (1965). *Brookings Quarterly Econometric Model of the United States*. Amsterdam: North-Holland Publishing Company.

Duffy, J. A. and D. F. Hendry (2017). The impact of near-integrated measurement errors on modelling long-run macroeconomic time series. *Econometric Reviews 36*, 568–587.

Eckstein, O., E. W. Green, and A. Sinai (1974). The Data Resources model: Uses, structure, and analysis of the U.S. economy. *International Economic Review 15*, 595–615.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflation. *Econometrica 50*, 987–1007.

Engle, R. F. and C. W. J. Granger (1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica 55*, 251–276.

Engle, R. F., D. F. Hendry, and J.-F. Richard (1983). Exogeneity. *Econometrica 51*, 277–304. Reprinted in Ericsson, N. R. and Irons, J. S. (eds.) *Testing Exogeneity*, Oxford: Oxford University Press, 1994; and in Campos, Ericsson and Hendry (2005) *op. cit.*

Ericsson, N. R., D. F. Hendry, and S. B. Hood (2016). Milton Friedman as a Statistician and Econometrician. In R. A. Cord and J. D. Hammond (Eds.), *Milton Friedman: Contributions to Economics and Public Policy*, pp. 91–142. Oxford: Oxford University Press.

Ericsson, N. R. and E. L. Reisman (2012). Evaluating a global vector autoregression for forecasting. *International Advances in Economic Research 18*, 247–258.

Escribano, A. (2004). Nonlinear error correction: The case of money demand in the UK (1878–2000). *Macroeconomic Dynamics 8*, 76–116.

Fisher, I. (1921). The best form of index number. *American Statistical Association Quarterly 17*, 533–537.

Fisher, I. (1925). Our unstable dollar and the so-called business cycle. *Journal of the American Statistical Association 20*, 181–198.

Flavin, M. A. (1981). The adjustment of consumption to changing expectations about future income. *Journal of Political Economy 89*, 974–1009.

Friedman, M. (1940). Book review of *Business Cycles in the United States of America, 1919–1932* by J. Tinbergen, Vol. II. *American Economic Review 30*, 657–660.

Friedman, M. and A. J. Schwartz (1982). *Monetary Trends in the United States and the United Kingdom: Their Relation to Income, Prices, and Interest Rates, 1867–1975*. Chicago: University of Chicago Press.

Friedman, W. A. (2014). *Fortune Tellers: The Story of America's First Economic Forecasters*. Princeton: Princeton University Press.

Frisch, R. (1929). Correlation and scatter in statistical variables. *Nordic Statistical Journal 8*, 36–102.

Frisch, R. (1933a). *Pitfalls in the Statistical Construction of Demand and Supply Curves*. Neue Folge, Heft 5, Leipzig: Veröffentlichungen der Frankfurter Gesellschaft für Konjunkturforschung.

Frisch, R. (1933b). Propagation problems and impulse problems in dynamic economics. In *Economic Essays in Honour of Gustav Cassel*, pp. 171–205. London: Allen and Unwin. Reprinted as p155-85 in *A.E.A. Readings in Business Cycle Research*, Homewood, Ill.

Frisch, R. (1934). *Statistical Confluence Analysis by means of Complete Regression Systems*. Oslo: University Institute of Economics.

Frisch, R. (1938). Statistical versus theoretical relations in economic macrodynamics. Mimeograph dated 17 July 1938, League of Nations Memorandum. Reproduced by University of Oslo in 1948 with Tinbergen's comments. Contained in Memorandum 'Autonomy of Economic Relations', 6 November 1948, Oslo, Universitets Økonomiske Institutt. Reprinted in Hendry and Morgan (1995), *op. cit.*

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland 15*, 246–263.

Geary, R. C. (1949). Determination of linear relations between systematic parts of variables with errors of observations, the variances of which are unknown. *Econometrica 17*, 30–58.

Goldberger, A. S. (1959). *Impact Multipliers and Dynamic Properties of the Klein-Goldberger Model*. Amsterdam: North Holland.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica 37*, 424–438.

Graunt, J. (1662). *Natural and Political Observations Made upon the Bills of Mortality*. London.

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*, 1–118. Supplement.

Haavelmo, T. (1958). The role of the econometrician in the advancement of economic theory. *Econometrica 26*, 351–357.

Hald, A. (1998). *A History of Mathematical Statistics from 1750 to 1930*. New York: Wiley.

Hall, R. E. (1978). Stochastic implications of the life cycle-permanent income hypothesis: Evidence. *Journal of Political Economy 86*, 971–987.

Harrison, R., K. Nikolov, M. Quinn, G. Ramsay, A. Scott, and R. Thomas (2005). The Bank of England quarterly model. Research paper, 244pp, Bank of England, London.

Hendry, D. F. (1979). The behaviour of inconsistent instrumental variables estimators in dynamic systems with autocorrelated errors. *Journal of Econometrics 9*, 295–314.

Hendry, D. F. (1980). Econometrics: Alchemy or science? *Economica 47*, 387–406. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.

Hendry, D. F. (1996). On the constancy of time-series econometric equations. *Economic and Social Review 27*, 401–422.

Hendry, D. F. (2001a). How economists forecast. In D. F. Hendry and N. R. Ericsson (Eds.), *Understanding Economic Forecasts*, pp. 15–41. Cambridge, Mass.: MIT Press.

Hendry, D. F. (2001b). Modelling UK inflation, 1875–1991. *Journal of Applied Econometrics 16*, 255–275.

Hendry, D. F. (2003). J. Denis Sargan and the origins of LSE econometric methodology. *Econometric Theory 19*, 457–480.

Hendry, D. F. (2006). Robustifying forecasts from equilibrium-correction models. *Journal of Econometrics 135*, 399–426.

Hendry, D. F. (2018). Deciding between alternative approaches in macroeconomics. *International Journal of Forecasting 34*, 119–135, with 'Response to the Discussants', 142–146.

Hendry, D. F. and J. A. Doornik (1999). The impact of computational tools on time-series econometrics. In T. Coppock (Ed.), *Information Technology and Scholarship*, pp. 257–269. Oxford: Oxford University Press.

Hendry, D. F. and J. A. Doornik (2014). *Empirical Model Discovery and Theory Evaluation*. Cambridge, Mass.: MIT Press.

Hendry, D. F. and N. R. Ericsson (1991). An econometric analysis of UK 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 S. Johansen (2015). Model discovery and Trygve Haavelmo's legacy. *Econometric Theory 31*, 93–114.

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*, 549–563. Reprinted in Campos, Ericsson and Hendry (2005) *op. cit.*

Hendry, D. F. and G. E. Mizon (2000). The influence of A. W. Phillips on econometrics. In R. Leeson (Ed.), *A. W. H. Phillips: Collected Works in Contemporary Perspective*, pp. 353–364. Cambridge: Cambridge University Press.

Hendry, D. F. and G. E. Mizon (2012). Open-model forecast-error taxonomies. In X. Chen and N. R. Swanson (Eds.), *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*, pp. 219–240. New York: Springer.

Hendry, D. F. and G. E. Mizon (2014). Unpredictability in economic analysis, econometric modeling and forecasting. *Journal of Econometrics 182*, 186–195.

Hendry, D. F. and M. S. Morgan (Eds.) (1995). *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press.

Hendry, D. F. and J. N. J. Muellbauer (2018). The future of macroeconomics: Macro theory and models at the Bank of England. *Oxford Review of Economic Policy 34*, 287–328.

Hickman, B. G. (Ed.) (1972). *Econometric Models of Cyclical Behaviour*. Number 36 in National Bureau of Economic Research Studies in Income and Wealth. New York: Columbia University Press.

Hicks, J. R. (1937). Mr Keynes and the 'Classics': A suggested interpretation. *Econometrica 5*, 147–159.

Hood, W. C. and T. C. Koopmans (Eds.) (1953). *Studies in Econometric Method*. Number 14 in Cowles Commission Monograph. New York: John Wiley & Sons.

Hooker, R. H. (1901). Correlation of the marriage rate with trade. *Journal of the Royal Statistical Society 64*, 485–492. Reprinted in Hendry and Morgan (1995), *op. cit.*

Hoover, K. D. (Ed.) (1995). *Macroeconometrics: Developments, Tensions and Prospects*. Dordrecht: Kluwer Academic Press.

Hoover, K. D. (2003). A history of postwar monetary economics and macroeconomics. In W. Samuels, J. Biddle, and J. Davis (Eds.), *A Companion to the History of Economic Thought*, pp. 411–427. Oxford: Blackwell.

Hoover, K. D., S. Johansen, and K. Juselius (2008). Allowing the data to speak freely: The macroeconometrics of the cointegrated vector autoregression. *American Economic Review: Papers and Proceedings 98*, 251–255.

Jevons, W. S. (1884). *Investigations in Currency and Finance*. London: Macmillan.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control 12*, 231–254. Reprinted in R.F. Engle and C.W.J. Granger (eds), *Long-Run Economic Relationships*, Oxford: Oxford University Press, 1991, 131–52.

Juselius, K. (2006). *The Cointegrated VAR Model: Methodology and Applications*. Oxford: Oxford University Press.

Juselius, K. and M. Franchi (2007). Taking a DSGE model to the data meaningfully. *Economics-The Open-Access, Open-Assessment E-Journal 2007-4*.

Kaplan, G. and G. Violante (2018). Microeconomic heterogeneity and macroeconomic shocks. *Journal of Economic Perspectives 32*, 167–194.

Kendrick, J. W. (1970). The historical development of national-income accounts. *History of Political Economy 2*, 284–315.

Keynes, J. M. (1920). *The Economic Consequences of the Peace*. London: Macmillan.

Keynes, J. M. (1936). *The General Theory of Employment, Interest and Money*. London: Macmillan.

Keynes, J. M. (1939). Professor Tinbergen's method. *Economic Journal 44*, 558–568. Reprinted in Hendry and Morgan (1995), *op. cit.*

King, G. (1696). *Natural and Political Observations and Conclusions upon the State and Condition of England*. London.

King, G. (1697). *Of the Naval Trade of England, 1688, and the National Profit then Arising thereby*. London.

Klein, J. L. (1997). *Statistical Visions in Time*. Cambridge: Cambridge University Press.

Klein, L. R. (1950). *Economic Fluctuations in the United States, 1921–41*. Number 11 in Cowles Commission Monograph. New York: John Wiley.

Klein, L. R. (1971). *An Essay on the Theory of Economic Prediction*. Chicago: Markham Publishing Company.

Klein, L. R. (1987). A history of computation in econometrics. Mimeo, University of Pennsylvania.

Klein, L. R., R. J. Ball, A. Hazlewood, and P. Vandome (1961). *An Econometric Model of the UK*. Oxford: Oxford University Press.

Klein, L. R. and A. S. Goldberger (1955). *An Econometric Model of the United States, 1929-1952*. Amsterdam: North Holland Publishing Co.

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*, 161–179.

Koopmans, T. C. (1949). Identification problems in economic model construction. *Econometrica 17*, 125–144. Reprinted with minor revisions in Hood, W. C. and Koopmans, T. C. (eds.) (1953), *Studies in Econometric Method*. Cowles Commission Monograph 14, New York: John Wiley & Sons.

Koopmans, T. C. (Ed.) (1950a). *Statistical Inference in Dynamic Economic Models*. Number 10 in Cowles Commission Monograph. New York: John Wiley & Sons.

Koopmans, T. C. (1950b). When is an equation system complete for statistical purposes? See Koopmans (1950a), Chapter 17.

Kuznets, S. (1937). *National Income and Capital Formation, 1919–1935*. New York: National Bureau of Economic Research.

Kuznets, S. (1946). *National Income. A Summary of Findings*. New York: National Bureau of Economic Research.

Kydland, F. E. and E. C. Prescott (1990). Business cycles: Real facts and a monetary myth. *Federal Reserve Bank of Minneapolis, Quarterly Review 14*, 3–18.

Kydland, F. E. and E. C. Prescott (1991). The econometrics of the general equilibrium approach to business cycles. *Scandinavian Journal of Economics 93*, 161–178.

Laspeyres, E. (1871). Die berechnung einer mittleren warenpreissteigerung. *Jahrbücher für Nationalökonomie und Statistik*.

Lehfeldt, R. A. (1914). The elasticity of the demand for wheat. *Economic Journal 24*, 212–217.

Lenoir, M. (1913). *Etudes sur la Formation et le Mouvement des Prix*. Paris.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. In K. Brunner and A. Meltzer (Eds.), *The Phillips Curve and Labor Markets*, Volume 1 of *Carnegie-Rochester Conferences on Public Policy*, pp. 19–46. Amsterdam: North-Holland Publishing Company.

Malinvaud, E. (1988). Econometric methodology at the cowles commission: Rise and maturity. *Econometric Theory 4*, 187–209.

Mann, H. and A. Wald (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica 11*, 173–220.

Marschak, J. (1953). Economic measurements for policy and prediction. See Hood and Koopmans (1953).

Marschak, J. and O. Lange (1940). Mr. Keynes on the statistical verification of business cycle theories. See Hendry and Morgan (1995).

McDonald, J. and G. D. Snooks (1986). *Domesday Economy: A New Approach to Anglo-Norman History*. Oxford: Oxford University Press.
Domesday Book on https://www.nationalarchives.gov.uk/domesday/

Mian, A. and A. Sufi (2018). Finance and business cycles: The credit-driven household demand channel. *Journal of Economic Perspectives 32*, 31–58.

Miller, P. J. (1978). Forecasting with econometric methods: A comment. *Journal of Business 51*, 579–586.

Mills, T. C. (2011). Bradford Smith: An econometrician decades ahead of his time. *Oxford Bulletin of Economics and Statistics 73*, 276–285.

Mitchell, W. C. (1913). *Business Cycles*, Volume 13. Berkeley: California University Memoirs.

Mizon, G. E. and J.-F. Richard (1986). The encompassing principle and its application to non-nested hypothesis tests. *Econometrica 54*, 657–678.

Moore, H. L. (1911). *Laws of Wages: An Essay in Statistical Economics*. New York: MacMillan.

Moore, H. L. (1914). *Economic Cycles–Their Law and Cause*. New York: MacMillan.

Moore, H. L. (1925). A moving equilibrium of demand and supply. *Quarterly Journal of Economics 39*, 359–371.

Morgan, M. S. (1990). *The History of Econometric Ideas*. Cambridge: Cambridge University Press.

Muellbauer, J. N. J. (1983). Surprises in the consumption function. *Economic Journal Conference Supplement*, 34–49.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica 29*, 315–335.

Nelson, C. R. (1972). The prediction performance of the FRB-MIT-PENN model of the US economy. *American Economic Review 62*, 902–917. Reprinted in T.C. Mills (ed.), *Economic Forecasting*. Edward Elgar, 1999.

Paasche, H. (1875). *Die Geldentwertung zu Halle in den letzten Decennien dieses Jahrhunderts*. Halle: Plötz.

Pagan, A. R. (2003). Report on modelling and forecasting at the Bank of England. *Bank of England Quarterly Bulletin*, Spring.
http://www.bankofengland.co.uk/archive/Documents/historicpubs/qb/2003/qb030106.pdf

Persons, W. M. (1919). Indices of business conditions. *Review of Economic Statistics 1*, 5–107.

Persons, W. M. (1924). *The Problem of Business Forecasting*. Number 6 in Pollak Foundation for Economic Research Publications. London: Pitman.

Pesaran, M. H. (1974). On the general problem of model selection. *Review of Economic Studies 41*, 153–171. Reprinted in Campos, Ericsson and Hendry (2005) *op. cit.*

Petty, W. (1690). *Political Arithmetick*. London.

Phillips, A. W. H. (1954). Stabilization policy in a closed economy. *Economic Journal 64*, 290–333. Reprinted in R. Leeson (ed.) (2000) *op. cit.*

Phillips, A. W. H. (1958). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. *Economica 25*, 283–299. Reprinted as pp243–260 in R. Leeson (ed.) (2000) *op. cit.*

Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics 33*, 311–340.

Playfair, W. (1786). *The Commercial and Political Atlas: Representing, by Means of Stained Copper-Plate Charts, the Progress of the Commerce, Revenues, Expenditure and Debts of England during the Whole of the Eighteenth Century*. London.

Playfair, W. (1801). *Statistical Breviary; Shewing, on a Principle Entirely New, the Resources of Every State and Kingdom in Europe*. London, Wallis.

Qin, D. (1993). *The Formation of Econometrics: A Historical Perspective*. Oxford: Clarendon Press.

Qin, D. (2013). *A History of Econometrics: The Reformation from the 1970s*. Oxford: Clarendon Press.

Quenouille, M. H. (1957). *The Analysis of Multiple Time Series*. London: Griffin.

Quesnay, F. (1766). *Tableau Economique: Analyse de la Formule Arithmétique du Tableau Économique de la Distribution des Dépenses Annuelles d'une Nation Agricole*. Leyden: Dupont de Nemours.

Reiersøl, O. (1945). Confluence analysis by means of instrumental sets of variables. *Arkiv for Matematik Astronomi och Fysik 32*, 1–19.

Renfro, C. G. (2009). *The Practice of Econometric Theory: An Examination of the Characteristics of Econometric Computing*. London: Springer.

Renton, G. A. (Ed.) (1975). *Modelling the Economy*. London: Heinemann Educational Books.

Robbins, L. (1932). *An Essay on the Nature and Significance of Economic Science*. London: Macmillan.

Ryan-Collins, J., R. A. Werner, and J. Castle (2016). A half-century diversion of monetary policy? an empirical horse-race to identify the uk variable most likely to deliver the desired nominal gdp growth rate. *Journal of International Financial Markets, Institutions and Money 43*, 158 – 176.

Sargan, J. D. (1959). The estimation of relationships with autocorrelated residuals by the use of instrumental variables. *Journal of the Royal Statistical Society B, 21*, 91–105. Reprinted as pp. 87–104 in Sargan J. D. (1988), *Contributions to Econometrics*, Vol. 1, Cambridge: Cambridge University Press.

Sargan, J. D. (1964). Wages and prices in the United Kingdom: A study in econometric methodology (with discussion). In P. E. Hart, G. Mills, and J. K. Whitaker (Eds.), *Econometric Analysis for National Economic Planning*, Volume 16 of *Colston Papers*, pp. 25–63. London: Butterworth Co. Reprinted as pp. 275–314 in Hendry D. F. and Wallis K. F. (eds.) (1984). *Econometrics and Quantitative Economics*. Oxford: Basil Blackwell, and as pp. 124–169 in Sargan J. D. (1988), *op. cit.*

Sargan, J. D. (1980). Some tests of dynamic specification for a single equation. *Econometrica 48*, 879–897. Reprinted as pp. 191–212 in Sargan J. D. (1988), *op. cit.*

Sargent, T. J. and C. A. Sims (1977). Business cycle modelling without pretending to have too much apriori economic theory. In C. A. Sims (Ed.), *New Methods in Business Cycle Research*. Minneapolis: Federal Reserve Bank of Minneapolis.

Schultz, H. (1928). *The Theory and Measurement of Demand*. University of Chicago Press.

Schumpeter, J. (1954). *History of Economic Analysis*. New York: Oxford University Press.

Slutsky, E. (1937). The summation of random causes as the source of cyclic processes. *Econometrica 5*, 105–146. (Translation from the Russian version of 1927.).

Smets, F. and R. Wouters (2003). An estimated stochastic dynamic general equilibrium model of the Euro Area. *Journal of the European Economic Association 1*, 1123–1175.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: W. Strahan & T. Cadell.

Smith, B. B. (1926). Combining the advantages of first-difference and deviation-from-trend methods of correlating time series. *Journal of the American Statistical Association 21*, 55–59.

Smith, B. B. (1929). Judging the forecast for 1929. *Journal of the American Statistical Association 24*, 94–98.

Stigler, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge Mass.: Harvard University Press.

Stone, J. R. N. (1978). Keynes, political arithmetic and econometrics. *Proceedings of the British Academy 64*, 55–92.

Stone, J. R. N. (1985). The disaggregation of the household sector in the national accounts. In G. Pyatt and J. I. Round (Eds.), *Social Accounting Matrices: A Basis for Planning*, pp. 145–185. Washington D.C.: The World Bank.

Student (1908). On the probable error of the mean. *Biometrika 6*, 1–25.

Summers, L. H. (1991). The scientific illusion in empirical macroeconomics. *Scandinavian Journal of Economics 93*, 129–148.

Taleb, N. N. (2009). Errors, robustness, and the fourth quadrant. *International Journal of Forecasting 25*, 744–759.

Theil, H. (1961). *Economic Forecasts and Policy,* (2nd ed.). Amsterdam: North-Holland Publishing Company.

Tinbergen, J. (1930). Determination and interpretation of supply curves: An example [Bestimmung und Deutung von Angebotskurven: ein Beispiel]. *Zeitschrift fur Nationalökonomie 1*, 669–679. Reprinted in Hendry and Morgan (1995), *op. cit.*

Tinbergen, J. (1940). *Statistical Testing of Business-Cycle Theories*. Geneva: League of Nations. Vol. II: Business Cycles in the United States of America, 1919–1932.

Tinbergen, J. (1951). *Business Cycles in the United Kingdom 1870–1914*. Amsterdam: North-Holland.

Törnqvist, L. (1936). The Bank of Finland's consumption price index. *Bank of Finland Monthly Bulletin 10*, 1–8.

Treasury, H. (1980). *Macroeconomic Model Technical Manual*. London: H.M. Treasury.

Vining, R. (1949a). Methodological issues in quantitative economics. *Review of Economics and Statistics 31*, 77–86.

Vining, R. (1949b). A rejoinder. *Review of Economics and Statistics 31*, 91–94.

Waelbroeck, J. K. (Ed.) (1976). *The Models of Project LINK*. Amsterdam: North-Holland Publishing Company.

Wallis, K. F., M. J. Andrews, D. N. F. Bell, P. G. Fisher, and J. D. Whitley (1984). *Models of the UK Economy, A Review by the ESRC Macroeconomic Modelling Bureau*. Oxford: Oxford University Press.

Wallis, K. F., P. G. Fisher, J. A. Longbottom, D. S. Turner, and J. D. Whitley (1987). *Models of the UK Economy: A Fourth Review by the ESRC Macroeconomic Modelling Bureau*. Oxford: Oxford University Press.

Walras, L. (1874). *Elements of Pure Economics, translated 1954*. London: Allen and Unwin.

Weart, S. (2010). The discovery of global warming. *http://www.aip.org/history/climate/co2.htm*.

Wold, H. O. A. (1938). *A Study in The Analysis of Stationary Time Series*. Stockholm: Almqvist and Wicksell.

Wold, H. O. A. (1949). Statistical estimation of economic relationships. *Econometrica 17*, 1–21. Supplement.

Working, E. J. (1927). What do statistical demand curves show? *Quarterly Journal of Economics 41*, 212–235.

Wren-Lewis, S. (2018). Ending the microfoundations hegemony. *Oxford Review of Economic Policy 34*, 55–69.

Wright, P. G. (1915). Review of Moore, 'Economic Cycles' (1915). *Quarterly Journal of Economics 29*, 631–641. Reprinted in Hendry and Morgan (1995), *op. cit.*

Wright, P. G. (1929). Review of H. Schultz: 'Statistical Laws of Demand and Supply'. *Journal of the American Statistical Association 24*, 207–215.

Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time series (with discussion). *Journal of the Royal Statistical Society 89*, 1–64. Reprinted in Hendry and Morgan (1995), *op. cit.*

Yule, G. U. (1927). On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society, A 226*, 267–298. Reprinted in Hendry and Morgan (1995), *op. cit.*

Zarnowitz, V. (2004). An important subject in need of much new research. *Journal of Business Cycle Measurement and Analysis 1*, 1–7.