# 1

# The Methodology of Empirical Econometric Modeling: Applied Econometrics Through the Looking-Glass

*David F. Hendry*

**Abstract**

This chapter considers the methodology of empirical econometric modeling. The historical background is reviewed from before the Cowles Foundation to the rise of economic theory-based econometrics and the decline of data concerns. A theory for "Applied Econometrics" suggests reinterpreting the role of economic theory given that the intrinsic non-stationarity of economic data vitiates analyses of incomplete specifications based on *ceteris paribus*. Instead, the many steps from the data-generation process (DGP) through the local DGP (LDGP) and general unrestricted model to a specific representation allow an evaluation of the main extant approaches. The potential pitfalls confronting empirical research include inadequate theory, data inaccuracy, hidden dependencies, invalid conditioning, inappropriate functional form, non-identification, parameter non-constancy, dependent, heteroskedastic errors, wrong expectations formation, misestimation and incorrect model selection. Recent automatic methods help resolve many of these difficulties. Suggestions on the teaching of "Applied Econometrics" are followed by revisiting and updating the "experiment in applied econometrics" and by automatic modeling of a four-dimensional vector autoregression (VAR) with 25 lags for the numbers of bankruptcies and patents, industrial output per capita and real equity prices over 1757–1989.

## 1.1 Introduction

> "Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!" (Quote from the Red Queen in *Through the Looking-Glass and What Alice Found There*, Lewis Carroll, Macmillan & Co., 1899, [henceforth cited as "Lewis Carroll, 1899"])

Most econometricians feel a bit like Alice did at having to run fast even to stand still. Handbooks are an attempt to alleviate the problem that our discipline moves forward rapidly, and *infoglut* can overwhelm, albeit that one has to run even faster for a short period to also find time to read and digest their contents. That will require some sprinting here, given that the contents of this *Handbook of Econometrics* provide up-to-date coverage of a vast range of material: time series, cross-sections, panels, and spatial; methodology and philosophy; estimation – parametric and nonparametric – testing, modeling, forecasting and policy; macro, micro, finance, growth and development; and computing – although I do not see *teaching*. Such general headings cross-categorize "Applied Econometrics" by types of data and their problems on the one hand – time series, cross-sections, panels, high frequency (see, e.g., Barndorff-Nielsen and Shephard, 2007), limited dependent variables (see, e.g., Heckman, 1976), or count data (excellently surveyed by Cameron and Trivedi, 1998), etc. – and by activities on the other (modeling, theory calibration, theory testing, policy analysis, forecasting, etc.). The editors considered that I had written on sufficiently many of these topics during my career to "overview" the volume, without also noting how markedly all of them had changed over that time. The main aim of an introductory chapter is often to overview the contents of the volume, but that is manifestly impossible for the *Handbook of Econometrics* given its wide and deep coverage. In any case, since the *Handbook* is itself an attempt to overview Applied Econometrics, such an introduction would be redundant.

Thus, my focus on empirical econometric modeling concerns only one of the activities, but I will also try to present an interpretation of what "Applied Econometrics" is; what those who apply econometrics may be trying to achieve, and how they are doing so; what the key problems confronting such applications are; and how we might hope to resolve at least some of them. Obviously, each aspect is conditional on the previous one: those aiming to calibrate a theory model on a claimed set of "stylized facts" are aiming for very different objectives from those doing data modeling, so how they do so, and what their problems are, naturally differ greatly. This chapter will neither offer a comprehensive coverage, nor will it be an uncontroversial survey. En route, I will consider why "Applied Econometrics" does not have the highest credibility within economics, and why its results are often attacked, as in Summers (1991) among many others (see Juselius, 1993, for a reply). Evidence from the contents of textbooks revealing the marginal role of "Applied Econometrics" and "economic statistics" within the discipline has been provided recently by Qin (2008) and Atkinson (2008) respectively. Since two aspects of our profession with even lower status than "Applied Econometrics" are data (measurement, collection and preparation), and teaching, I will try and address these as well, as they are clearly crucial to sustaining and advancing a viable "Applied Econometrics" community. Economic forecasting and policy are not addressed explicitly, being uses of empirical models, and because the criteria for building and selecting such models differ considerably from those applicable to "modeling for understanding" (see, e.g., Hendry and Mizon, 2000; and for complete volumes on forecasting, see Clements and Hendry, 2002a, 2005; Elliott, Granger and Timmermann, 2006).

Economists have long been concerned with the status of estimated empirical models. How a model is formulated, estimated, selected and evaluated all affect that status, as do data quality and the relation of the empirical model to the initial subject-matter theory. All aspects have been challenged, with many views still extant. And even how to judge that status is itself debated. But current challenges are different from past ones – partly because some of the latter have been successfully rebutted. All empirical approaches face serious problems, yet the story is one of enormous progress across uncharted terrain with many mountains climbed – but many more to surmount. I will recount some of that story, describe roughly where we are presently located, and peer dimly into the future. Why "*Applied Econometrics Through the Looking-Glass*"? Lewis Carroll was the pseudonym for Charles Dodgson, a mathematician who embodied many insights in the book which is cited throughout the present chapter: a Looking-Glass is a mirror, and applied findings in economics can only reflect the underlying reality, so obtaining a robust and reliable reflection should guide its endeavors.

Following the brief section 1.2 on the meaning of the topic, section 1.3 summarizes some of the history of our fallible discipline. Then section 1.4 proposes a "theory of Applied Econometrics" which highlights some of the problems empirical modeling confronts in a non-stationary environment, where non-stationary is used throughout in the "wide sense" to denote any changes in the distributions of the random variables modeled by economists. Section 1.5 discusses a recent tool for automatic modeling, Autometrics, based on the last decade of research into model selection (see Doornik, 2007a; Hendry and Krolzig, 2005; Hendry, with Doornik and Nielsen, 2007). Section 1.6 comments on teaching Applied Econometrics, and section 1.7 revisits the experiment in applied econometrics conducted by Magnus and Morgan (1999). Section 1.8 then looks at automatic modeling of a four-variable dynamic system related to industrial output since 1700, with 25 lags in its initial formulation and many outliers over more than 250 years. Section 1.9 concludes. Throughout, I draw heavily on a number of my previous papers. Despite there being almost 200 citations to other scholars, I am conscious that documentation is bound to be incomplete, and apologize for omitting many contributions.

## 1.2 What is "Applied Econometrics"?

> "When I use a word," Humpty Dumpty said in rather a scornful tone, "it means just what I choose it to mean – neither more nor less." (Lewis Carroll, 1899)

At the superficial level, "Applied Econometrics" is "any application of econometrics," as distinct from theoretical econometrics. If it were not for the imperialist tendencies of econometricians, that would suffice, but econometrics has been applied in space science, climatology, political science, sociology, epidemiology, marketing, *inter alia*, not to mention the claim in *How the Laws of Physics Lie* (see Cartwright, 1983) that econometrics is the key methodology for all of science ... Sorry to disappoint the eager reader, but I will not be covering even a wide range of the economic applications, never mind that plethora of outside studies.

Some applied econometricians would include any applications involving analyses of "real economic data" by econometric methods, making "Applied Econometrics" synonymous with empirical econometrics. However, such a view leads to demarcation difficulties from applied economics on the one hand and applied statistics on the other. Defining "econometrics," as in Frisch (1933), to comprise only studies involving the unification of economic theory, economic statistics (data), and mathematics (statistical methods) helps in demarcation, but limits its scope and inadvertently excludes (say) developing econometric theory itself, or just improving data measurement and collection.

Outsiders might have thought that "Applied Econometrics" was just the application of econometrics to data, but that is definitely not so; virtually no journal editor would publish such a piece. Rather, the notion of mutual penetration dominates – but as a one-way street. Economic theory comes first, almost mandatorially. Perhaps this just arises from a false view of science, namely that theory precedes evidence, even though, apart from a few famous occasions, science rarely proceeds by imposing a preconceived theory on evidence, and evidence regularly shapes and stimulates theory. Yet the latter is rarely the case in applied econometrics – and to see how we arrived at such a state, we need to consider the contingent history of our discipline.

## 1.3 Historical background

> "The time has come," the Walrus said, "To talk of many things: Of shoes – and ships – and sealing-wax – Of cabbages – and kings – And why the sea is boiling hot – And whether pigs have wings." (Lewis Carroll, 1899)

The histories of statistics and econometrics are now reasonably well documented: on the former, see, e.g., the books by Stigler (1986, 1999) and Hald (1990, 1998); and for the latter, see Epstein (1987), Morgan (1990), Qin (1993), Klein (1997), and Le Gall (2007); also see Christ (1994), Spanos (2006), Farebrother (2006) and Gilbert and Qin (2006); and for reprints of key papers, see Darnell (1994) and Hendry and Morgan (1995), with related material in Caldwell (1993), Hamouda and Rowley (1997), Mills (1999) and Campos, Ericsson and Hendry (2005). These books provide overall bibliographic perspective.

### 1.3.1 Pre-Cowles

> The shop seemed to be full of all manner of curious things – but the oddest part of it all was, that whenever she looked hard at any shelf, to make out exactly what it had on it, that particular shelf was always quite empty: though the others round it were crowded as full as they could hold. (Lewis Carroll, 1899)

An aspect of that history which is still somewhat under-emphasized, despite being stressed by Hendry and Morgan (1995), is the role that empirical studies have played as a driver of new econometric concepts, theories and methods, standing in some contrast to its direct impact on economics. Certainly, early attempts were replete with what we would now view as blunders – William Stanley Jevons' sunspot, and Henry Moore's Venus, theories of business cycles are regularly trotted out as examples of how silly econometricians can be, yet Jevons (1875) and Moore (1923) respectively need to be contrasted with other careful and insightful empirical analyses in Jevons' research, edited by Foxwell (1884), and in Moore (1911). On the former, see the appraisal in Peart (2001); and on the latter, e.g., Stigler (1962) comments: "Moore's standard of craftsmanship is high: the basic data are fully reported and the work was carefully done." Also, note the cited comment from Alfred Marshall to Moore in 1912 that "the 'ceteris paribus' clause – though formally adequate seems to me impracticable,"a point that will recur below. The "upward sloping demand curve" for pig-iron in Moore (1914) is perhaps the most notorious misinterpretation, but in fact led to many later insights – in particular, reactions to it helped unravel the whole complicated and intertwined issues of simultaneity, identification, exogeneity, and partial effects (see Wright, 1915, 1929; Working, 1927; Tinbergen, 1930, *inter alia*).

Equally importantly, many "strange" empirical correlations had been found that stimulated the unraveling of both spurious, and later nonsense, regressions in works such as Yule (1897), Hooker (1901), and especially the famous explanation in Yule (1926), leading first to a distinction between short-run and long-term relationships, then unit roots, and eventually cointegration, as in Granger (1981), and dozens of later contributions surveyed in Hendry and Juselius (2000, 2001). Despite obvious progress – Stigler (1962) begins his article about Moore by "If one seeks distinctive traits of modern economics, traits which are not shared to any important degree with the Marshallian or earlier periods, he will find only one: the development of statistical estimation of economic relationships" – trouble lay ahead.

The attack by Robbins (1932) on the empirical studies of Schultz (1928) – portrayed as the feckless Dr. Blank studying the demand for herring (rather than sugar) – was the first of several critiques which sought to deny any substantive role for econometrics in economics.[1] Tinbergen's attempts to build empirical models of investment activity brought down the wrath of John Maynard Keynes (see Tinbergen, 1939, 1940; Keynes, 1939, 1940), who insisted that the economist had to be "in the saddle" with the econometrician as the "patient ass," and sarcastically demanded that Tinbergen satisfy:

> an experiment on his part. It will be remembered that the seventy translators of the Septuagint were shut up in seventy separate rooms with the Hebrew text and brought out with them, when they emerged, seventy identical translations. Would the same miracle be vouchsafed if seventy multiple correlators were shut up with the same statistical material? And anyhow, I suppose, if each had a different economist perched on his *a priori*, that would make a difference to the outcome.

We will return in section 1.4 both to that issue, which may well now be possible, and to Keynes' general claims – one might like to ponder whether 70 economic theorists asked to tackle the same puzzle would derive precisely the same model? As ever, other more constructive outcomes followed from that debate, especially the memorandum by Frisch (1938), and it certainly did not discourage Haavelmo (1944).

### 1.3.2 War and post-war

> "I'll tell you all my ideas about Looking-glass House. First, there's the room you can see through the glass – that's just the same as our drawing-room, only the things go the other way." (Quote from Alice in Lewis Carroll, 1899)

Despite Koopmans (1937) being a key precursor to the establishment of modern econometrics in Haavelmo (1944), the attack by Koopmans (1947) on Burns and

Mitchell (1946) allowed the assertion of "measurement without theory" to become capable of dismissing empirical work without further serious consideration. The vigorous reply by Vining (1949a, 1949b) still merits reading. With a few honorable exceptions (such as Atkinson, 2005), even the use of the word "measurement" as a title for economics' papers seems to have decreased since (other than in "measurement errors"). Tress (1959) offers a near contemporary analysis of the acrimony between economics and econometrics at that time, and a possible reconciliation.

Keynes (1939) had asserted that a long list of "preconditions" had to be satisfied to validate empirical inferences, implicitly arguing that empirical econometrics must fail unless everything was known in advance (see, e.g., Hendry, 1980). But if it was impossible to empirically uncover things not already known theoretically, then no science could have progressed; rather, scholasticism would still rule. There are several flaws in Keynes' claims, of which three are the most important.

First, "partial knowledge" can be valuable, and can be learnt from evidence, with or without prior theories, albeit being subject to revision later. Our understanding of gravity remains incomplete, but has advanced greatly since Aristotle's early view of objects' natural places (smoke rises, stones fall, as natural to go to heaven or the centre of the earth, respectively), through Ptolemy's epicycle theory of planetary motions, Descartes' vortex theory, and Newtonian inverse-square laws, which Adam Smith (1795) presciently noted was just a model and not the "truth" (as most of his contemporaries assumed). Einstein's relativity theory is still not the "final answer." Retrospectively, Aristotle's theory did not go beyond "explaining" the phenomena themselves, whereas Newtonian theory, while closely based on Kepler's laws of motion of planetary bodies, explained many additional aspects, so was a clear advance in knowledge, even if later it too was found to be incomplete, and at relativistic speeds, incorrect. Moreover, despite neither Aristotle's nor Newton's theories being "true," both were at least consistent with the observed facts of their time. The relevant empirical regularities persisted through many theories, which provided better explanations, often with unanticipated predictions of new phenomena – genuine "mutual penetration." Thus, progress is the key to science, not one-off forging of true laws that hold forever.

Second, if there are invariant features of reality – as in physics and chemistry – then empirical research can discover them without prior knowledge, as happened historically in many branches of science. Conversely, if nothing is invariant (an extreme of Heraclitus of Ephesus supposed view that "reality is change"), neither economic theories nor econometric models would be of any practical value (see Hendry, 1995b). Following Bachelier (1900), equity prices have long been viewed as close to random walks, which may be thought to entail the absence of any invariants, but if correct – as suggested by some modern theories and empirical tests of efficient markets – is actually a powerful invariant, as contrasted with a data generating process whose structure alters every period.

Third, empirical econometrics could still "advance" by rejecting economic theories. This would at least allow economists to focus on theories that were not yet rejected, if any, and improve those that faced discordant evidence. However, progress might be somewhat inefficient when new theories are easily generated as variants of previous incarnations.

Similar comments apply to Koopmans' claim that "without resort to theory, in the sense indicated, conclusions relevant to the guidance of economic policies cannot be drawn." Such an argument is not sustainable in general. Originally, aspirin was a folk remedy for hangovers, derived from brewing willow-tree bark – of which acetylsalicylic acid, aspirin's active ingredient, is a natural constituent – without any theory as to how or why that policy intervention worked (see Weissmann, 1991). A less well known example is the use in folk medicine of fungal-based products, some of which contain natural antibiotics such as penicillin: over 3,000 years ago, the Chinese had used moldy soybean curd for treating skin infections, again with no theory on which to base that policy. Theories can be invaluable, and can enhance the credibility of proposed policies, but they are not essential, especially when they are incorrect.

### 1.3.3 The rise of economic theory-based econometrics

> "If you'll tell me what language 'fiddle-de-dee' is, I'll tell you the French for it!" she exclaimed triumphantly. (Quote from Alice to the Red Queen in Lewis Carroll, 1899)

Another critique of empirical modeling follows from the joint dependence of economic events, namely the resulting issues of endogeneity and identification. It seems widely believed that identification restrictions must be given *a priori* by economic theory, especially in simultaneous systems, yet that belief also does not have a substantive basis, as shown in section 1.4.7 on identification.

Together, the cumulative critiques just noted led to an almost monolithic approach to empirical econometric research: first postulate an individualistic, intertemporal optimization theory; next derive a model therefrom; third, find data with the same names as the theory variables; then select a recipe from the econometrics cookbook that appropriately blends the model and the data, or if necessary, develop another estimator; finally report the newly forged empirical law. Thus, we have a partial answer to the issue posed in section 1.2: the contingent history of econometrics suggested that the only viable route for applied research in economics, where all current-dated variables are potentially endogenous, was to provide the quantitative cloth for a completely pre-specified theoretical formulation derived from general economic principles. But that approach too is problematic and not without its critics. Economic theory has progressed dramatically over the past century – imagine being forced to impose 1900's economic theory today as the basis for empirical research. If you recoil in horror at that idea, then you have understood why much past Applied Econometrics research is forgotten: discard the economic theory that it quantified and you discard the empirical evidence. Instead of progress, we find fashions, cycles and "schools" in research. The problem is not that early pioneers ignored economic theory, but that the available theory was seriously incomplete – as it still is today. Indeed, the Cowles Commission research was essentially predicated on the belief that the relevant economic theory already existed, so complicated issues of model choice could be avoided by imposing valid restrictions derived from correct economic theories: on discovering that such theory was not available, many turned to help develop it (see, e.g., Qin, 2008; Bjerkholt, 2007 (Bjerkholt, 2005, is a useful precursor). Koopmans, Hurwicz and Arrow all made major contributions to economic theory, and to quote Bjerkholt (2007): "Haavelmo stated later on various occasions that economic theory needed further development for the econometric methods to become fully applicable" (also see Moene and Rødseth, 1991). Indeed, to quote Haavelmo (1989) himself:

> The basis of econometrics, the economic theories that we had been led to believe in by our forefathers, were perhaps not good enough. It is quite obvious that if the theories we build to simulate actual economic life are not sufficiently realistic, that is, if the data we get to work on in practice are not produced the way that economic theories suggest, then it is rather meaningless to confront actual observations with relations that describe something else.

He reiterated that view in his presidential address to the Econometric Society (published as Haavelmo, 1958):

> What I believe to be true, however, is this: The training in the technical skills of econometrics can represent a powerful tool for imaginative speculation upon the basic phenomena of economic life; and, furthermore, it would be fruitful to bring the requirements of an econometric "shape" of the models to bear upon the formulation of fundamental economic hypotheses from the very beginning.

Once model choice cannot be avoided, methodology becomes a salient issue, and it would seem every conceivable methodology has at least one advocate. Pagan (1987) considered what he viewed as the three main econometric methodologies, relating mine to Leamer (1978) and Sims (1980), yet the ubiquitous "theory-based" approach was not mentioned, albeit that there are really many variants thereof.

### 1.3.4 The decline of data concerns

> "You're travelling the wrong way." (Train guard to Alice in Lewis Carroll, 1899)

At about the same time that *a priori* theory-based econometrics became dominant, data measurement and quality issues were also relegated as a central component of empirical publications. Early on, data series were often published in their entirety, with careful caveats about accuracy, but later, at best, were recorded in appendices, or not at all. For example, Clark (1932) allowed the first famous estimate of the size of the Keynesian "multiplier." Although computerized databases have recently started to compensate for the absence of the printed record, electronic data are sometimes revised in situ, making it difficult for later investigators to duplicate previously-published findings. In a detailed study of a number of cases, Atkinson

(2008) emphasizes that the outcomes reported would change substantively if data had been more carefully evaluated prior to the econometric analysis. To quote:

> my concern (is) with the status, within economics, of economic statistics. By "economic statistics," I mean the study of how we create, use and assess economic data – what one might call "data appreciation." ... It is true that economists are using empirical data to an unprecedented extent, and applying tools of great sophistication. Economics is a much more data-driven subject than it was in the past. But, I shall argue, economists have too often come to take data for granted, without critical examination of their strengths and weaknesses.

With that caveat about data firmly in mind, let us turn to methodology: measurement is reconsidered in section 1.4.3, and an illustration of some effects of substantial revisions in section 1.7.1.

The route ahead views all models as arising from reductions of whatever process generated the data, which is a combination of the economic outcome and the measurement system. We discuss these reductions in relation to their impact on the parameters that actually governed the economic decisions of the relevant agents. Most reductions occur implicitly, as investigators usually approach modeling from the opposite perspective, namely what to include in their analysis, although its success or failure will depend on whether the sub-set of variables considered allows a model to capture the salient and constant characteristics of the data-generating process (DGP). What to include and how to include it certainly depends on the economics behind the analysis; but what is found depends on the unknown data-generating process and the losses of information from the reductions that were necessary to derive the postulated model.

## 1.4 A theory of Applied Econometrics

> "Why, sometimes I've believed as many as six impossible things before breakfast." (Quote from the White Queen in Lewis Carroll, 1899)

If only it were just six! To believe that he or she has ascertained the "truth," an applied econometrician would have to believe at least the following dozen impossible (composite) assumptions:

1. a correct, complete, and immutable underlying economic theory derivation
2. a correct, comprehensive choice of all relevant variables, including all dynamic specifications
3. exact data measurements on every variable
4. the absence of any hidden dependencies, including collinearity and simultaneity
5. the validity and relevance of all conditioning variables (including instruments)
6. the precise functional forms for every variable
7. that all parameters of interest are identified in the resulting model specification

8. that all entities treated as parameters are constant over time, and invariant to all potentially omitted variables and regime changes
9. the errors have "independent," homoskedastic, distributions
10. all expectations formulations are correct, or agents' expectations are accurately measured
11. the choice of estimator is appropriate at relevant sample sizes
12. a valid and non-distortionary method of model selection is used.

If all of these assumptions had to be perfectly correct to produce useful empirical evidence, there would be no hope of ever doing so. In Hendry (1987), I suggested the four "golden prescriptions" of econometrics, abbreviated here as:

(i) think brilliantly: if you think of the right answer before modeling, then the empirical results will be optimal and, of course, confirm your brilliance;
(ii) be infinitely creative: if you do not think of the correct model before commencing, the next best is to think of it as you proceed;
(iii) be outstandingly lucky: if you do not think of the "true model" before starting nor discover it en route, then luckily stumbling over it before completing the study is the final sufficient condition. This may be the most practical of these suggestions. Failing this last prescription:
(iv) stick to doing theory.

Lest the reader thinks the list of a dozen requirements above is overly dramatic, or even new, Hendry and Morgan (1995) record:

> In the thesis as a whole, Koopmans (1937) assembles together and confronts most of the major issues in econometrics, which we have translated into current terminology as:
> 1. the joint occurrence of errors-in-variables and errors-in-equations
> 2. the need for a complete set of determining variables to leave an innovation error
> 3. a reductionist approach of proceeding from general to simple
> 4. the distinctions between the activities of specification, estimation and distribution, as spelt out by R.A. Fisher
> 5. the non-experimental nature of economic data
> 6. the need to condition on systematic components with independently varying error terms
> 7. the choice of functional form, using linearity for convenience
> 8. the formulation of the parameters of interest
> 9. the need to test underlying assumptions
> 10. the importance of incorporating all relevant information
> 11. the avoidance of unnecessary assumptions
> 12. the need for the general model to be estimable
> 13. the need for the model specification to be robust.

We now consider each of the twelve assumptions in turn, devoting the separate section 1.5 to the last, namely model selection.

### 1.4.1 Economic theory

> "It seems very pretty," she said when she had finished it, "but it's rather hard to understand." (Alice after reading the Jabberwocky poem in Lewis Carroll, 1899)

Economic theory has created many major ideas that have in turn changed the world, from the "invisible hand" in (Smith, 1759, p. 350), understanding the gains from trade and the problems with mercantilism, through the effects of tariffs and taxes, to modern insights into issues such as welfare economics, option pricing, auctions, contracts, principal–agent and game theories, trust and moral hazard, asymmetric information, institutions, and all their attendant impacts on market functioning and industrial, and even political, organization. In doing so, economics has evolved dramatically over time, and will undoubtedly continue doing so, hence at no instant can it be claimed to be correct, complete and immutable. For example, most theories take preferences as a given – sometimes even as "deep parameters" – but there are many endogenous determinants, from learning, adaptation, and advertising among others (see, e.g., von Weizsacker, 2005), with psychological, behavioral, and neuro-economics bidding fair to play key roles in the future (see, *inter alia*, Fehr and Falk, 2002; Fehr, Fischbacher, Kosfeld, 2005; Camerer, 2007).

Theories need to be distinguished in terms of their "levels', where low-level theories are well established and widely accepted (e.g., the optical theory behind the design of telescopes and the interpretation of their evidence), whereas high-level theories usually assume the validity of many lower levels, but are subject to doubt (as in theories of the accelerating expansion of the universe as due to "dark energy"). Facts are items of empirical information which depend only on low-level theories and measurements, and can be reliably replicated. Since all empirical evidence is theory laden to some degree, albeit often just from very low-level theories, "measurement without theory" is trivially impossible, and must relate to the lack of use of high level theories – the appropriate blend of theory and empirical evidence affects research efficiency, not necessarily the validity of any resulting findings (see, e.g., Gilbert and Qin, 2007). Many low-level statements are correct, complete and immutable, such as $1 + 1 = 2$, and although essential to arithmetic, cannot "explain" economic behavior. Conversely, testing theories just by their predictions is problematic, as false assumptions can entail correct conclusions: assume $1 = 2$, then $2 = 1$, so adding both sides, $3 = 3$, which is valid and presumably thereby establishes that $1 = 2$ (also see Ericsson and Hendry, 1999).

General theories that do "explain" the *Gestalt* of empirical evidence are a boon, but are not essential. Similarly, although experimentation can be helpful, it is far from the only source of evidence: observational science is not an oxymoron. The progressivity of science – its cumulation of findings that cohere, are consolidated in theoretical explanations, and suggest where next to investigate – is its most salient attribute. There is simply no case that we understand no more than (say) Aristotle, or Kepler, etc.: lights work, computers run, planes fly. Moreover, it is possible to "predict" with considerable accuracy what changes to chips will speed up calculations, and what putative aircraft will not fly, which are inferences beyond any local set of experiments and evidence, are not purely inductive, and can be generalized, though doubtless with limits. Science seeks progress, whether by new experiments, new instruments or observations, new theories or refutations of extant ones. We now know that ulcers are caused by helicobacter pylori bacteria – not by stress – so cheap, painless antibiotics can cure ulcers, replacing life-threatening operations or expensive drugs. The path that led to that idea is irrelevant to its validity, and could be serendipity, careful testing, or a theory prediction, whereas stringent evaluation and replicability are crucial. In turn, such advances can lead to radical new thinking, albeit that initially they often face considerable opposition – sometimes even derision: scientists are humans, and rarely change their views until evidence overwhelms. Even then, theories are usually not rejected by evidence, but rather are replaced when "better" theories develop that explain more, especially if they account for previous anomalies.

Statistical analyses become essential in observational sciences, such as astronomy and economics, where "field" experiments are almost impossible to control. Then theory and modeling difficulties both explode and certainty declines, especially when behavioral change is possible: despite rendering previous analyses less than fully relevant to new settings, progress remains the key. It is widely recognized that special factors may intrude on a theory-based model (e.g., changes in credit rationing, nationalization, deregulation, price controls, wars, etc.), but less recognized that such special factors can dominate when accounting for data variability. Morgan (1990), Spanos (1995), Hendry (1995b) and Hendry and Mizon (2000) discuss some of the problems involved in testing theories using observational data.

Economists have not formally justified the principle of deriving empirical models from theory – most seem to assume it is obvious – so a substantial proportion of empirical econometric evidence is "high level" in that its credibility depends on the prior credibility of the theoretical model from which it was derived. Given any conjecture, we can usually test its empirical validity, thereby sustaining a destructive approach (see, e.g., Popper, 1963, on conjectures and refutations), although issues of inference from small and heterogeneous data samples complicate the analysis. If a theory implementation is simply discarded when it is rejected, the process fails to incorporate learning from the evidence. Conversely, if it is not discarded, some or all of the empirical model, the measurements and the theory must be revised, although there is no unique or even structured way of doing so. It is a *non sequitur* to assume that the particular alternative considered is true when the null is rejected. A progressive research approach of successively encompassing congruent (see section 1.4.2.4) models consolidated by empirically-relevant theories offers one possibility.

Alternative approaches to "macroeconomic theory" abound in the literature: Samuelson (1947) initiated a tradition of models based on constrained optimization, implemented by Hall (1978) as Euler equations; Kydland and Prescott (1990,

1991) formulate real-business cycle theories with rational expectations, leading to dynamic stochastic general equilibrium (DSGE) models as in Smets and Wouters (2003), whereas Hildenbrand (1994, 1999) emphasizes heterogeneity of endowments; Stiglitz (2003) stresses that asymmetric information can induce Keynesian effects, and Aghion *et al.* (2002) argue that agents only have imperfect-knowledge expectations. Moreover, many aspects of economic theory models can be chosen freely, such as the units of time and forms of utility functions: indeed, Stigum (1990) views theories as characterizing "toy agents in toy economies." However, when data are non-stationary, few transformations will be able to characterize the evidence in a constant relationship. For example, linear relationships between variables, which often arise in Euler equations, seem unlikely to be good descriptions in growing economies (see Ermini and Hendry, 2008, and Spanos, Hendry and Reade, 2008, for tests of log versus linear dependent variables in $\mathsf{I}(1)$ processes).

The absence from many economic theories of some of the main sources of data variability occurs across most research areas in economics, and although it differs in form, is probably part of the reason for the rash of "puzzles" (i.e., anomalous or even contradictory evidence) so beloved of the present generation of journal editors. In microeconomics, low $\mathsf{R}^2$ values reveal that much of the variability is not accounted for by the postulated models. That outcome is usually ascribed to individual heterogeneity and idiosyncrasies, which can indeed generate high levels of unexplained variability, but there has to be some doubt that all the major factors have been included. In panel data studies, much observed data variation is attributed to "individual effects," which are removed by (e.g.) differencing or deviations from individual means. However, if the evidence that most micro-variability is due to individual heterogeneity is correct, then "representative" agent theories cannot be the best basis for macro-behavior, although aggregation could sustain some approaches (see, e.g., Granger, 1987; Blundell and Stoker, 2005), but not others (Granger, 1980). Finally, cross-country studies rarely account for key institutional differences between the constituent economies, and often use averages of data over historical epochs where considerable changes occurred between periods (see, e.g., Sala-i-Martin, 1997, and the criticisms in Hoover and Perez, 2004; Hendry and Krolzig, 2004).

#### *1.4.1.1 Non-stationarity and* ceteris paribus

> "You don't know how to manage Looking-glass cakes," the Unicorn remarked. "Hand it round first, and cut it afterwards." (Lewis Carroll, 1899)

A time series process is non-stationary if its moments or distributional form change over time. Two important forms of non-stationarity are unit roots and structural breaks, both of which lead to permanent changes. The former induce stochastic trends, which can be eliminated by differencing, or cointegration can also remove unit roots and retain linear combinations of levels of the variables (however, unit roots and cointegration are only invariant under linear transformations of variables). There is a vast literature, and recent surveys include Hendry and Juselius

(2000, 2001) and Johansen (2006). Structural breaks matter most when they induce location shifts in the processes under analysis, but those can also be removed in part by differencing or co-breaking (see Hendry and Massmann, 2007). Forecast failure – defined as a significant deterioration in forecast performance relative to its anticipated outcome, usually based on historical performance – is all too common, and is almost certainly due to structural breaks (see, e.g., Clements and Hendry, 1998, 1999, 2002b).

Because much observed data variability is due to factors absent from economic theories, a serious gap exists between macroeconomic theory models and applied econometric findings (see Spanos, 1989; Juselius, 1993; Hendry, 1995b; Nymoen, 2002). All economic theories rely on implicit *ceteris paribus* clauses, as "controls in thought experiments," although in a general equilibrium system in which everything depends on everything else, *ceteris paribus* is suspect. In empirical modeling, *ceteris paribus* cannot apply under non-stationarity even if the relevant variables are strongly exogenous, since "other things" will not be "equal." Cartwright (2002) describes *ceteris paribus* as roughly equivalent to "if nothing interferes then ... some regularity is observed." In non-stationary processes, nothing will interfere only if all other factors are irrelevant, not because they will not change. Many sources of wide-sense non-stationarity impinge on economic data, including technical progress, R&D, new legislation, institutional changes, regime shifts, financial innovation, shifting demography, evolving social and political mores, as well as conflicts and other major catastrophes, inducing both evolution and structural breaks, all of which change the distributional properties of data.

Two resolutions are possible to wide-sense non-stationarity. First, a "minor influence" theorem could show on theoretical or evidential grounds that all omitted factors can be neglected, either because changes in them are of a smaller order of importance than included effects, or because they are orthogonal to all the effects that matter (see Hendry, 2005, and compare Boumans, 2005, who refers to *ceteris neglectis* and *ceteris absentibus*). Neither condition is plausible unless at least all the major influences have been included. Doing so brings us anyway to the second solution, namely including all potentially relevant variables at the outset, embedding theory models in more general systems that also allow for all the empirically-known influences, as well as the many historical contingencies that have occurred. Thus, institutional knowledge and economic history become essential ingredients in Applied Econometrics. Far from diminishing the importance of economic reasoning as a basis for empirical econometrics, including all non-stationarities seems the only way to reveal the underlying economic behavior uncontaminated by excluded changes. Of course, theory models of the likely behavioral reactions of economic agents to major changes would also help. As macro-data are the aggregates of the economic microcosm, these problems must afflict all empirical econometric studies, and are not merely a problem for time series analysts. Since the need to model all non-stationarities if empirical results are to be useful is important for both economics and econometrics, the next section considers its prevalence.

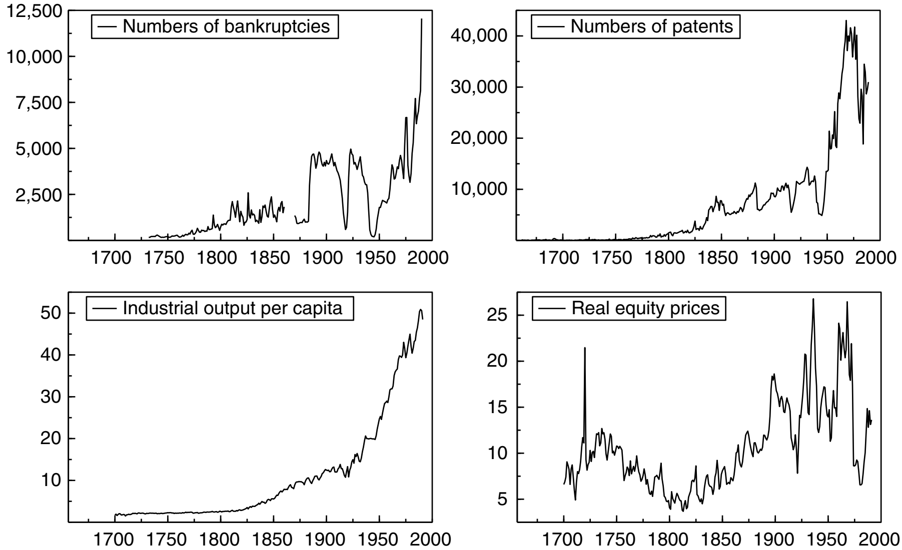


*Figure 1.1* Historical time series for the UK

### *1.4.1.2 Long-run change*

> "I see you're admiring my little box," the Knight said in a friendly tone. "It's my own invention – to keep clothes and sandwiches in. You see I carry it upside-down, so that the rain can't get in."
>
> "But the things can get out," Alice gently remarked. "Do you know the lid's open?" (Lewis Carroll, 1899)

Figure 1.1 records some historical time series for the UK over the period from about 1700 to 1991 (the dates differ for the various variables). Many other variables manifesting dramatic non-stationarities are shown in Hendry (2001a, 2001b, 2005) and Clements and Hendry (2001), where the first and last examine UK industrial output in more detail. Here we focus on numbers of bankruptcies and patents, industrial output per capita, and real equity prices (deflated by a cost of living price index) (see Feinstein, 1972; Mitchell, 1988; Crafts and Mills, 1994, *inter alia*).

These four variables were selected from a range of alternatives as being related to advances in technology and medicine, their implementation, incentives for progress through intellectual property, and one source of financing (see Siegel and Wright, 2007, for a recent review and bibliographic perspective). Technological change is sometimes modeled as an "exogenous" random walk. While that is an improvement over a deterministic trend, it is hardly a convincing representation of a process which requires substantial inputs of human and physical capital, as highlighted by endogenous growth models (see, e.g., Crafts, 1997). At the very

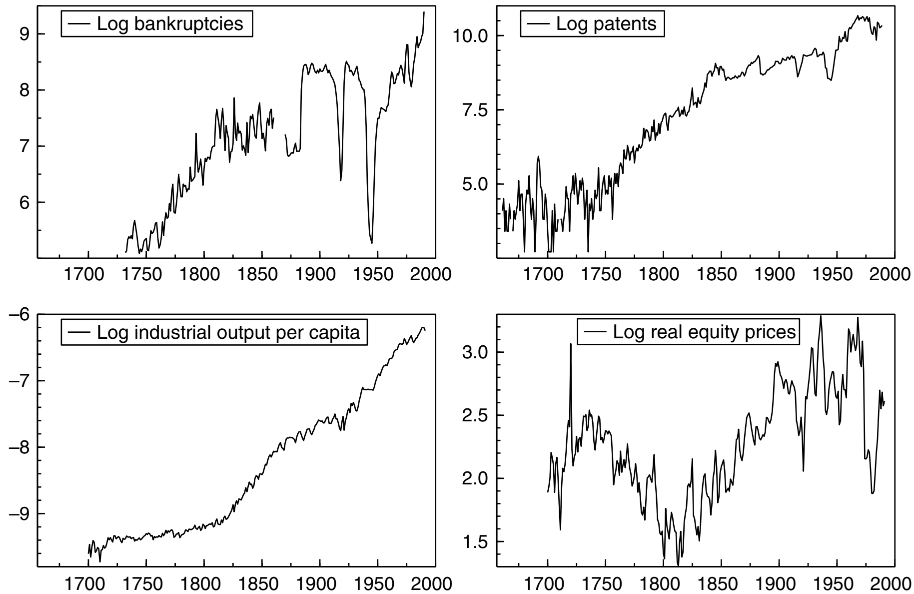


*Figure 1.2* Logs of historical UK time series

start of the industrial revolution, Smith (1776, Ch. 1) notes the key development of specialists in R&D: "In the progress of society, philosophy or speculation becomes, like every other employment, the principal or sole trade and occupation of a particular class of citizens."

The non-stationarities in Figure 1.1 are blatant, and reflect more than just unit roots. Major historical events are clear: e.g., real equity prices exploded in the South Sea Bubble, not regaining such levels again for 200 years, collapsed in the Napoleonic and both world wars, as well as the first oil crisis, and today are little above pre-industrial revolution levels. Frankly, it is almost infeasible to build sensible empirical models of these levels series.

Figure 1.2 records the same four series in logs. Non-stationarity remains clear, but one can at least imagine ways of sucessfully modeling such data. Figure 1.3 reports their data densities separately in each of (approximately) the three centuries involved. The shifts in means and variances are marked, even if the presence of heterogeneity and dependence within each century can distort such histograms. Nor is the non-stationarity restricted to the levels of the variables, as Figure 1.4 illustrates for the annual changes in the four variables. Variance changes are clear, which also serves to make the densities of the changes look much more alike, as seen in Figure 1.5. Differencing alone can be insufficient to induce stationarity.

We will analyze the log transformations of these data in section 1.8.

In the absence of complete theoretical guidance on all relevant and irrelevant variables, functional forms, exogeneity, dynamics, and non-stationarities, empirical determination is essential. Consequently, the initially postulated models of

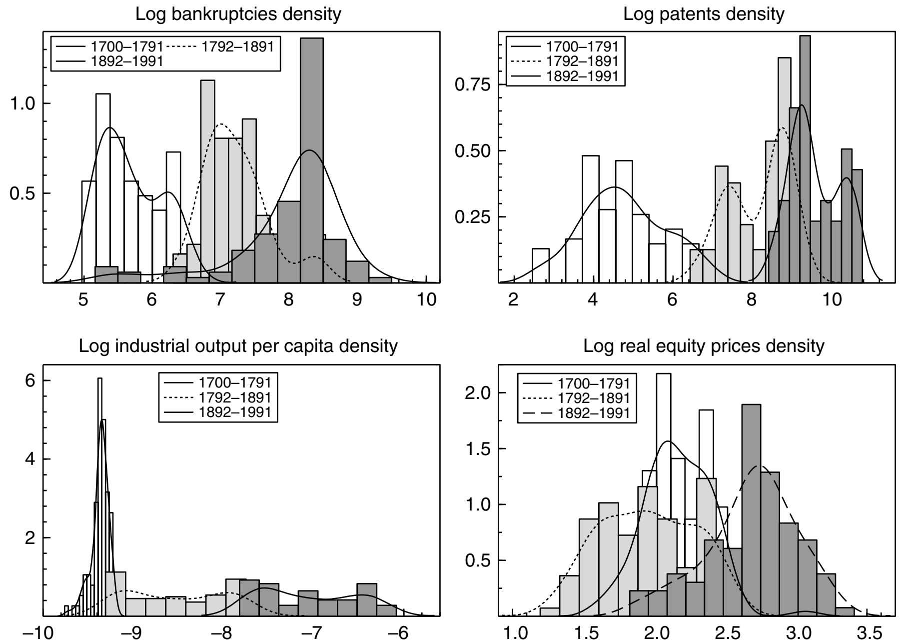


*Figure 1.3* Three centuries of data distributions of levels

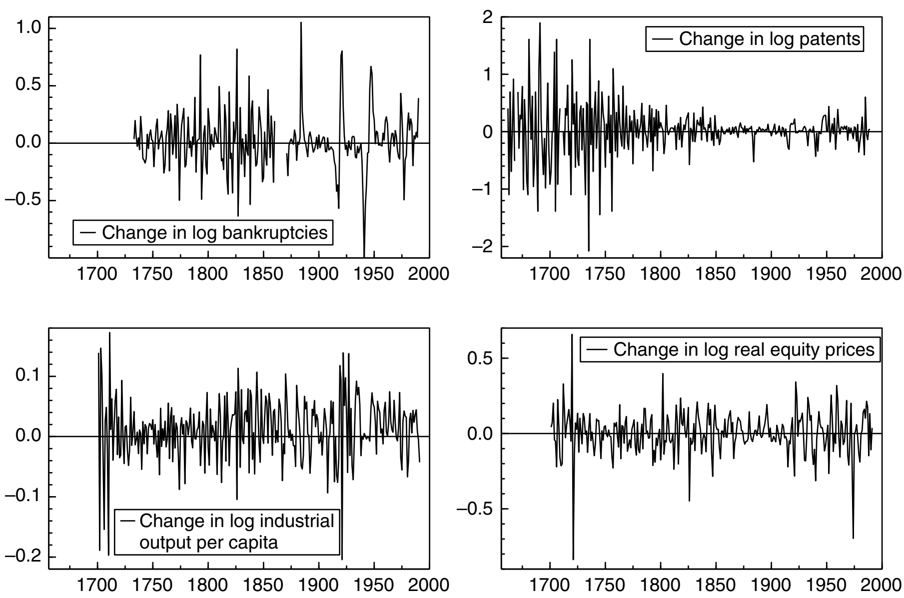


*Figure 1.4* Changes in historical time series for the UK

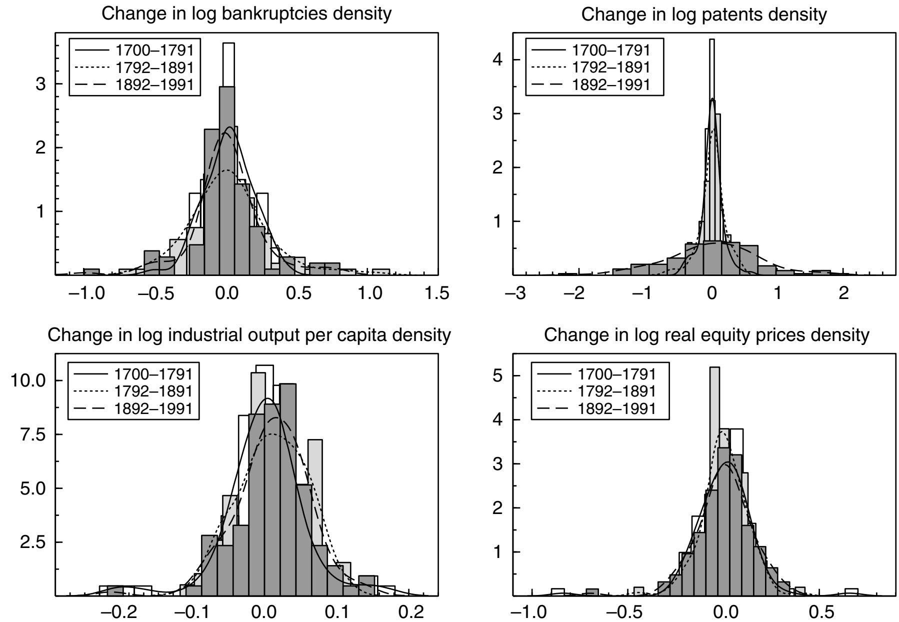


*Figure 1.5* Three centuries of data distributions of changes

empirical research should usually involve many variables, although final selections may prove to be parsimonious (an implication is considered in section 1.5): we now consider that route.

### 1.4.2 Incomplete specifications

> "What am I to do?" exclaimed Alice, looking about in great perplexity as first one round head, and then the other, rolled down from her shoulder, and lay like a heavy lump in her lap. (Lewis Carroll, 1899)

Economies are so high dimensional, interdependent, heterogeneous, and evolving that a comprehensive specification of all events is impossible: the number of economy-wide relevant variables is uncountable in a human lifetime. Reducing that high dimensionality by aggregation over any or all of time, space, commodities, agents, initial endowments, etc., is essential, but precludes any claim to "truth." So if one cannot get at the "truth," what is on offer in economics? Three alternatives are: imposing theory-based models; constructing partial models, which aim to estimate some parameters associated with a theory, usually by generalized method of moments (GMM); or seeking the local DGP guided by economic theory. All three could operate, but depend on different assumptions. We first outline how empirical models must arise, then evaluate the three approaches in general against that basis.

#### *1.4.2.1 From DGP to LDGP*

> "The prettiest are always further!" she said at last. (Quote from Alice in Lewis Carroll, 1899)

Granted a stochastic basis for individual agent decision taking, such that any economic transaction can be described as an event in an event space, which could have been different for a myriad of reasons, then outcomes are measurable random variables with (possibly different) distributions at each point in time. Let $\mathbf{U}_T^1 = (\mathbf{u}_1, \ldots, \mathbf{u}_T)$ be the complete set of random variables relevant to the economy under investigation over a time span $t = 1, \ldots T$, defined on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$, where $\Omega$ is the sample space, $\mathcal{F}$ the event space and $\mathsf{P}$ the probability measure. Denote the vast, complex, and ever-changing joint distribution of $\{\mathbf{u}_t\}$ conditional on the pre-sample outcomes $\mathbf{U}_0$ and all necessary deterministic terms $\mathbf{Q}_T^1 = (\mathbf{q}_1, \ldots, \mathbf{q}_T)$ (like constants, seasonal effects, trends, and shifts) by:

$$\mathsf{D}_{\mathsf{U}}\left(\mathbf{U}_T^1 \mid \mathbf{U}_0, \mathbf{Q}_T^1, \xi_T^1\right), \tag{1.1}$$

where $\xi_T^1 \in \Xi \subseteq \mathcal{R}^k$ are the parameters of the agents' decision rules that led to the outcomes in (1.1). Then $\mathsf{D}_{\mathsf{U}}(\cdot)$ is the unknown, and almost certainly unknowable, data-generation process of the relevant economy. The theory of reduction discussed in, *inter alia*, Hendry (1987), Florens, Mouchart and Rolin (1990) and Hendry (1995a, Ch. 9) shows that a well-defined sequence of operations leads to the "local" DGP (LDGP), which is the actual generating process in the space of the variables under analysis. The resulting LDGP may be complex, non-linear and non-constant from aggregating, marginalizing (following the relevant data partition), and sequential factorization (the order of these reductions below is not a central aspect), so the choice of the set of variables to analyze is crucial if the LDGP is to be viably "captured" by an empirical modeling exercise. In turn, that LDGP can be approximated by a "general unrestricted model" (GUM) based on truncating lag lengths, approximating the functional form (perhaps after data transformations) and specifying which parameters are to be treated as constant in the exercise. Finally, a further series of reductions, involving mapping to non-integrated data, conditioning, and simultaneity, lead to a parsimonious representation of the salient characteristics of the dataset. Tests of losses from all these reductions are feasible, as discussed in section 1.4.2.4.

*Aggregation.* Almost all econometric data are aggregated in some way, implicitly discarding the disaggregates: although some finance data relate to point individual transactions, their determinants usually depend on aggregates (such as inflation). We represent this mapping as $\mathbf{U}_T^1 \to \mathbf{V}_T^1$, where the latter matrix is a mix of the data to be analyzed and all other variables. The key issue is the impact of that mapping on $\xi_T^1 \to \phi_T^1$, where the latter set may include more or fewer constant parameters depending on the benefits or costs of aggregation. Aggregates are linear sums, so their means have variances proportional to population size: if $x_{i,t} \sim \mathsf{IN}\left[\mu_t, \sigma_t^2\right]$

then $\bar{x}_t = N_t^{-1} \sum_{i=1}^{N_t} x_{i,t} \sim \mathsf{IN}\left[\mu_t, N_t^{-1}\sigma_t^2\right]$. Log transforms of totals and means, $\bar{x} > 0$, only differ by that population size as $\ln \sum_{i=1}^{N_t} x_{i,t} = \ln \bar{x}_t + \ln N_t$, so standard deviations of log aggregates are proportional to scaled standard deviations of means: $\mathsf{SD}[\ln \sum_{i=1}^{N_t} x_{i,t}] \simeq N_t^{-1}\sigma_t/\mu_t$ (see, e.g., Hendry, 1995a, Ch. 2). Thus logs of aggregates can be well behaved, independently of the underlying individual economic behavior.

*Data transformations.* Most econometric models also analyze data after transformations (such as logs, growth rates, etc.), written here as $\mathbf{W}_T^1 = \mathbf{g}(\mathbf{V}_T^1)$. Again, the key impact is on $\phi_T^1 \to \varphi_T^1$ and the consequences on the constancy of, and cross-links between, the resulting parameters. At this stage we have created:

$$\mathsf{D}_{\mathsf{W}}\left(\mathbf{W}_T^1 \mid \mathbf{U}_0, \mathbf{Q}_T^1, \varphi_T^1\right). \tag{1.2}$$

The functional form of the resulting representation is determined here by the choice of $\mathbf{g}(\cdot)$. Many economic variables are intrinsically positive in levels, a property imposed in models by taking logs, which also ensures that the error standard deviation is proportional to the level.

*Data partition.* No reduction is involved in specifying that $\mathbf{W}_T^1 = (\overline{\mathbf{W}}_T^1 : \mathbf{R}_T^1)$, where $\mathbf{R}_T^1$ denotes the $n \times T$ data to be analyzed and $\overline{\mathbf{W}}_T^1$ the rest. However, this decision is a fundamental one for the success of the modeling exercise, in that the parameters of whatever process determines $\mathbf{R}_T^1$ must deliver the objectives of the analysis.

*Marginalizing.* To implement the choice of $\mathbf{R}_T^1$ as the data under analysis necessitates discarding all the other potential variables, which corresponds to the statistical operation of marginalizing (1.2) with respect to $\overline{\mathbf{W}}_T^1$:

$$\mathsf{D}_{\mathsf{W}}\left(\overline{\mathbf{W}}_T^1, \mathbf{R}_T^1 \mid \mathbf{U}_0, \mathbf{Q}_T^1, \varphi_T^1\right) = \mathsf{D}_{\overline{\mathsf{W}}}\left(\overline{\mathbf{W}}_T^1 \mid \mathbf{R}_T^1, \mathbf{U}_0, \mathbf{Q}_T^1, \overline{\varphi}_T^1\right) \mathsf{D}_{\mathsf{R}}\left(\mathbf{R}_T^1 \mid \mathbf{U}_0, \mathbf{Q}_T^1, \omega_T^1\right). \tag{1.3}$$

While such a conditional-marginal factorization is always possible, a viable analysis requires no loss of information from just retaining $\omega_T^1$. That will occur only if $\left(\overline{\varphi}_T^1, \omega_T^1\right)$ satisfy a cut, so their joint parameter space is the cross-product of their individual spaces, precluding links across those parameters. At first sight, such a condition may seem innocuous, but it is very far from being so: implicitly, it entails Granger non-causality of (all lagged values of) $\overline{\mathbf{W}}_T^1$ in $\mathsf{D}_{\mathsf{R}}(\cdot)$, which is obviously a demanding requirement (see Granger, 1969; Hendry and Mizon, 1999). Spanos (1989) calls the marginal distribution $\mathsf{D}_{\mathsf{R}}(\cdot)$ in (1.3) the Haavelmo distribution.

*Sequentially factorizing.* Next, letting $\mathbf{R}^1_{t-1} = (\mathbf{r}_1, \ldots, \mathbf{r}_{t-1})$, the retained marginal density from (1.3) can be sequentially factorized as (see, e.g., Doob, 1953):

$$\mathsf{D}_{\mathsf{R}}\left(\mathbf{R}^1_T \mid \mathbf{U}_0, \mathbf{Q}^1_T, \omega^1_T\right) = \prod_{t=1}^{T} \mathsf{D}_{\mathsf{r}_t}\left(\mathbf{r}_t \mid \mathbf{R}^1_{t-1}, \mathbf{U}_0, \mathbf{q}_t, \lambda_t\right). \tag{1.4}$$

The right-hand side of (1.4) completes the intrinsic reductions from the DGP to the LDGP for the set of variables under analysis (generally, the effects of the initial conditions $\mathbf{U}_0$ are ignored and assumed to be captured by $\mathbf{R}_0$). The sequential densities in (1.4) create a martingale difference (or innovation) process:

$$\epsilon_t = \mathbf{r}_t - \mathsf{E}\left[\mathbf{r}_t \mid \mathbf{R}^1_{t-1}, \mathbf{U}_0, \mathbf{q}_t\right], \tag{1.5}$$

where $\mathsf{E}[\epsilon_t | \mathbf{R}^1_{t-1}, \mathbf{U}_0, \mathbf{q}_t] = \mathbf{0}$ by construction.

*Parameters of interest.* These are the targets of the modeling exercise, and are hypothesized – on the basis of prior reasoning, past studies, and institutional knowledge – to be the features of interest. We denote them by $\theta \in \Theta$, and any later reduction choices must be consistent with obtaining $\theta$ from the final specification. To the extent that the economic theory supporting the empirical analysis is sufficiently comprehensive, the $\{\lambda_t\}$ in (1.4) should still contain the required information about the agents' decision parameters, so $\theta = \mathbf{h}(\omega^1_T)$. The next stage is to formulate a general model of (1.4) that also retains the necessary information.

### *1.4.2.2 From LDGP to general unrestricted model*

The LDGP in (1.4) can be approximated by a model based on a further series of reductions, which we now discuss. Indeed, (1.4) is often the postulated basis of an empirical analysis, as in a vector autoregression, albeit with many additional assumptions to make the study operational. There are no losses when the LDGP also satisfies these reductions, and if not, evidence of departures can be ascertained from appropriate tests discussed in section 1.4.2.4, so that such reductions are then not undertaken.

*Lag truncation.* The potentially infinite set of lags in (1.4) can usually be reduced to a small number, so $\mathbf{R}^1_{t-1} \simeq \mathbf{R}^{t-s}_{t-1} = (\mathbf{r}_{t-s} \ldots \mathbf{r}_{t-1})$, where the maximum lag length becomes $s$ periods, with initial conditions $\mathbf{R}^{1-s}_0$. Long-memory and fractional integration processes are considered in, e.g., Granger and Joyeux (1980), Geweke and Porter-Hudak (1983), Robinson (1995) and Baillie (1996). Letting $\mathsf{f}_{\mathsf{r}_t}(\cdot)$ denote the resulting statistical model of the $\{\mathbf{r}_t\}$, which could coincide with the LDGP when the reduction is without loss, then the mapping is:

$$\prod_{t=1}^{T} \mathsf{D}_{\mathsf{r}_t}\left(\mathbf{r}_t \mid \mathbf{R}^1_{t-1}, \mathbf{U}_0, \mathbf{q}_t, \lambda_t\right) \Rightarrow \prod_{t=1}^{T} \mathsf{f}_{\mathsf{r}_t}\left(\mathbf{r}_t \mid \mathbf{R}^{t-s}_{t-1}, \mathbf{R}^{1-s}_0, \mathbf{q}_t, \psi_t\right). \tag{1.6}$$

The obvious check on the validity of such a reduction is whether longer lags matter; and as before, the key criterion is the impact on $\{\psi_t\}$.

*Parameter constancy.* The parameters in question are those that characterize the distribution $\mathsf{f}_{\mathsf{r}}(\cdot)$ in (1.6). Then their constancy entails that the $\{\psi_t\}$ depend on a smaller set of parameters that are constant, at least within regimes. Complete constancy requires $\psi_t = \psi_0\ \forall t$, and while unlikely in economics, is often the assumption made, at least until there is contrary evidence. When there is no loss, $\theta = \mathbf{f}(\psi_0)$, so all parameters of interest can be recovered from the model.

*Linearity.* The distribution in (1.6) may correspond to the linear Normal when the functional form is chosen appropriately to ensure that a homoskedastic process also results:

$$\mathsf{f}_{\mathsf{r}_t}\left(\mathbf{r}_t \mid \mathbf{R}_{t-1}^{t-s}, \mathbf{R}_0^{1-s}, \mathbf{q}_t, \psi_0\right) \underset{app}{\frown} \mathsf{IN}_k\left[\sum_{i=1}^{s} \Pi_i \mathbf{r}_{t-i} + \Pi_{s+1}\mathbf{q}_t, \Omega\right]. \tag{1.7}$$

The LDGP distribution need not be Normal, but that is partly dependent on the specification of $\mathbf{q}_t$, especially whether breaks in deterministic terms are modeled therein. The constancy of the coefficients of any model also depends on the functional forms chosen for all the data transformations, and an operational GUM presumes that $\{\mathbf{r}_t\}$ has been transformed appropriately, based on theoretical and empirical evidence. Checks for various nonlinear alternatives and homoskedasticity are merited.

#### 1.4.2.3 *From the general to the specific*

Providing that a viable set of basic parameters is postulated (and below we will allow for the possibility of many shifts), then a variant of (1.7) can act as the GUM for a statistical analysis. When the LDGP is nested in the GUM, so none of the reductions above led to important losses, a well-specified model which embeds the economic theory and can deliver the parameters of interest should result. When the LDGP is not nested in the GUM, so the reductions in the previous sub-section entail losses, it is difficult to establish what properties the final specific model will have, although a well-specified approximation at least will have been found. Because wide-sense non-stationarity of economic variables is such an important problem, and within that class, location shifts are the most pernicious feature, section 1.5 considers the recent approach of impulse saturation (see Hendry, Johansen and Santos, 2008; Johansen and Nielsen, 2008).

*Mapping to a non-integrated representation.* Many economic variables appear to be integrated of at least first order (denoted I(1)), so there is a mapping $\mathbf{r}_t \to (\Delta\mathbf{r}_{p,t} : \beta'\mathbf{r}_t) = \mathbf{x}_t$, where there are $n-p$ cointegrating relations and $p$ unit roots, so $\mathbf{x}_t$ is now I(0). Processes that are I(2) can be handled by mapping to second differences as well (see, e.g., Johansen, 1995). This reduction to I(0) transforms $\psi_0$ to $\rho_0$ (say) and leads from (1.6) to:

$$\prod_{t=1}^{T} \mathsf{f}_{\mathbf{x}_t}\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}^{t-s}, \mathbf{X}_0^{1-s}, \mathbf{q}_t, \rho_0\right). \tag{1.8}$$

VARs like (1.7) are often formulated for $\mathbf{r}_t$, rather than $\mathbf{x}_t$, as occurs in the first stage of some cointegration analyses.

*Contemporaneous conditioning.* Conditioning concerns both contemporaneous variables in models and current-dated instrumental variables (IVs), so let $\mathbf{x}_t' = (\mathbf{y}_t' : \mathbf{z}_t')$, where the former are the $k$ variables to be modeled and the latter $n-k$ are taken as given. Then for $\rho_0 = (\kappa_1 : \kappa_2)$:

$$\mathsf{f}_{\mathbf{x}_t}\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}^{t-s}, \mathbf{X}_0^{1-s}, \mathbf{q}_t, \rho_0\right) = \mathsf{f}_{\mathbf{y}_t|\mathbf{z}_t}\left(\mathbf{y}_t \mid \mathbf{z}_t, \mathbf{X}_{t-1}^{t-s}, \mathbf{X}_0^{1-s}, \mathbf{q}_t, \kappa_1\right)$$
$$\mathsf{f}_{\mathbf{z}_t}\left(\mathbf{z}_t \mid \mathbf{X}_{t-1}^{t-s}, \mathbf{X}_0^{1-s}, \mathbf{q}_t, \kappa_2\right). \tag{1.9}$$

A viable analysis from the conditional distribution alone in (1.9) requires that $\theta = \mathbf{h}_1(\kappa_1)$; and there will be no loss of information only if $(\kappa_1, \kappa_2)$ satisfy a cut so $(\kappa_1, \kappa_2) \in \mathcal{K}_1 \times \mathcal{K}_2$, in which case $\mathbf{z}_t$ is weakly exogenous for $\theta$. When (1.7) holds, both conditional and marginal distributions in (1.9) will be Normal, and the relationships linear. The former leads to VAR-type modeling as noted, whereas the conditional representation in (1.9) underpins more "structural" approaches when the $\mathbf{z}_t$ are instruments: we return to conditioning in section 1.4.5 below.

*Simultaneity.* Finally, at least for the order of reductions considered here, simultaneity can allow a more parsimonious representation of the conditional distribution by modeling in terms of $\Gamma\mathbf{y}_t$, where $\Gamma$ is a non-singular matrix that captures the current-dated interdependencies. If $\mathbf{z}_t$ does not enter the conditional distribution, $\Gamma\mathbf{x}_t$ could be modeled directly relative to lagged information (see, e.g., Demiralp and Hoover, 2003).

### *1.4.2.4 Implications*

Five important issues are clarified by these reductions from the DGP down to a specific model of a sub-set of the variables.

*Econometric concepts.* First, there exists an LDGP as in (1.4) for whatever choices are made of $\mathbf{x}_t$. When all reductions are without loss, the statistical model $\mathsf{f}_{\mathbf{y}_t|\mathbf{z}_t}(\cdot)$ in (1.9) could also be the LDGP. Although most empirical analyses seem to commence by specifying what is included, rather than what is eliminated, almost all the central concepts in econometrics (in italics below) correspond to when reductions (in bold face) can be achieved without loss of relevant information:

- **Aggregation** entails no loss of information on marginalizing with respect to disaggregates when the formulation retains *sufficient statistics* for $\theta$.
- **Data transformations** have no associated reduction, but relate to *parameters of interest*, $\theta$, and hence the need for these to be *invariant and identifiable*.

- **Data partition** determines which variables to include and which to omit in the model *specification*, a decision that is dependent on the purpose of the modeling exercise, but is fundamental to the success of the empirical model.
- **Marginalizing** with respect to $\mathbf{v}_t$ is without loss if $\mathbf{X}_T^1$ is *sufficient* for $\theta$; and marginalizing with respect to $\mathbf{V}_{t-1}^1$ is without loss if it is *Granger non-causal* for $\mathbf{x}_t$ and the conditional-marginal parameters satisfy a *cut*.
- **Sequential factorization** induces no loss as $\epsilon_t$ from (1.5) is an *innovation* relative to $\mathbf{R}_{t-1}^1$.
- **Parameter constancy** over time is fundamental to most uses of a model, and *invariance* (constancy across interventions to the marginal process) is essential for policy.
- **Lag truncation** leads to no loss if $\epsilon_t$ remains an *innovation* against $\mathbf{X}_{t-1}^1$.
- **Integrated data** can be reduced to I(0) by *cointegration* and *differencing*, sustaining a more *parsimonious* representation, and supporting *conventional inference*.
- **Functional form** specification may or may not entail a reduction, and does not when the two densities are equivalent (e.g., logs of log-normal variables are normal).
- **Conditional factorizations** entail no loss of information when $\mathbf{z}_t$ is *weakly exogenous* for $\theta$, addressed in section 1.4.5.
- **Simultaneity** can allow one to *parsimoniously* capture *joint dependence*.

*Testing reductions.* Second, reductions are testable against any preceding, less reduced, distributions. Indeed, there is an accompanying taxonomy of evaluation information that seeks to ascertain the statistical significance of the losses imposed by the various reductions. This leads to six major null hypotheses about the final model's specification: homoskedastic innovations $\{\epsilon_t\}$; $\mathbf{z}_t$ weakly exogenous for $\theta$; constant, invariant $\theta$; data-admissible formulations on accurate observations; theory consistent, identifiable structures; encompassing rival models. While this exhausts the nulls to test, there are many alternatives to each. Models which satisfy the first and third are well specified on the available information, and if satisfying the first three are said to be (empirically) congruent. One model (parsimoniously) variance dominates another if it has a smaller unexplained variance (and no more parameters): the notion of one model explaining the results of other models extends variance dominance to account for all other parameters. The principle of encompassing was formalized in Hendry and Richard (1982), and the theory of testing developed by Mizon (1984) and Mizon and Richard (1986) (see Hendry and Richard, 1989, and Hendry, Marcellino, and Mizon, 2008, for surveys). An admissible, theory-consistent, encompassing, congruent model satisfies all six criteria.

*Choosing the Haavelmo distribution.* Third, knowledge of the LDGP is the "optimum" one can achieve for the given set of variables. Different choices of $\{\mathbf{r}_t\}$, and hence the Haavelmo distribution, will lead to different LDGPs with more or less constancy and congruence with the available evidence. If (1.7) were indeed the LDGP, then model selection could target its variables. The congruence of an empirical model corresponds to its encompassing the LDGP (so not deviating significantly from it in any of the first five directions) (see Bontemps and Mizon, 2003). Testing the selected model against all extant models of the same variables allows a rigorous evaluation of its "closeness" to the LDGP (see, *inter alia*, White, 1990; Mayo and Spanos, 2006).

*Parameter dependence.* Fourth, the resulting coefficients in (1.7) or (1.9) remain dependent on the initial DGP parameters. If those DGP parameters change, induced shifts can occur in the parameters of the LDGP. The extent to which these shifts occur, and when they do so, whether they can be anticipated, modeled or even understood, will depend on how usefully the reduced representation captures the structure of the relevant sub-set of the economy under analysis. Here, "structure" denotes invariance under extensions of the information set over (i) time (i.e., constancy), (ii) regimes (i.e., changes to marginal distributions or policy variables) and (iii) variables (so the reductions did not eliminate any important explanatory factors). When the initial economic analysis that led to the specification of $\{\mathbf{x}_t\}$ (i.e., the transformed sub-set of data under analysis) actually captured the main features of the behavior of the agents involved, then $\rho_0$, or $\kappa_1$, should be an invariant that also throws light on the agents' decision parameters underlying $\varphi_T^1$ in (1.2). Thus, properly embedded in a general congruent model, the economics should carry through.

*Minimizing reductions.* Finally, given the inertial dynamics of a high dimensional, interdependent and non-stationary system like an economy, reductions seem likely to be costly in practice and involve real information losses. These will manifest themselves through non-constant models, empirical "puzzles" and poor forecasts, so general systems seem generically preferable. "Errors" on empirical models are created by reductions, so will be highly composite, reflecting many components. It is unclear whether that also favors disaggregation, given problems posed by measurement errors and heterogeneity difficulties as disaggregation increases, or whether a "law of large numbers" may induce substantial offsets (as discussed above).

### *1.4.2.5 Evaluating the three main approaches*

We now consider how the three basic approaches fare against the above analysis. Given their assumptions, each would of course work well; and with sufficiently rigorous testing, the choice of approach becomes a matter of research efficiency (see White, 1990). But efficiency is important, as the assumptions may not adequately characterize reality, and rigorous attempts to reject are not always undertaken.

*Imposing economic theory.* First, if one simply imposes an *a priori* theory on the data, then the outcome will be excellent when the theory is complete (relative to the issue under analysis) and "correct" (in that all omissions are relatively negligible). Otherwise, it is difficult to ascertain in general how poor the outcome will be (see, e.g., Juselius and Franchi, 2007). If no testing occurs, that strategy is both highly risky and theory dependent. The risk is that a major discrepancy is not detected, leading to a poor description of the underlying agents' behavior: we addressed the issue of "*ceteris paribus*" in section 1.4.1.1, but if economies are inherently wide-sense non-stationary, then other things will not stay constant. When theories lack precise formulations of lag lengths, functional dependencies, other potential determinants, breaks, and non-economic factors of importance, such difficulties seem all too likely. The problem with theory dependence is that since no economic analysis has yet proved immutable, the empirical results will be discarded when the theory is altered, so there is no progressive knowledge accumulation. This is the real reason that Summers (1991) finds little contribution from empirical econometrics – it was not really allowed to make one, being restricted to providing empirical cloth for a pre-designed framework.

*Partial use of economic theory.* Second, a partial use of economic theory often leads to pre-specified moment conditions linking variables, $\mathbf{x}_t$, and parameters, $\varphi$, usually being zero for the "true" value of the parameter, $\varphi_0$, in the form (sometimes without conditioning):

$$\mathsf{E}\left[\mathbf{h}\left(\mathbf{x}_t,\varphi_0 \mid \mathbf{X}_{t-1}^{1}\right)\right] = \mathbf{0}\ \forall t, \tag{1.10}$$

enabling GMM estimation of $\varphi$ (see, e.g., Smith, 2007) (also, Smith, 1992, develops non-nested tests applicable after GMM). Equally often, inference has to be based on heteroskedastic and autocorrelation-consistent covariance (HAC) matrices (see White, 1980a; Andrews, 1991), which assume that the residuals reflect precisely those problems in the errors. Unfortunately, residuals can be heteroskedastic and autocorrelated for many other reasons, including unmodeled breaks, measurement errors, incorrect dynamics, omitted variables, or an inappropriate functional form *inter alia*, most of which would invalidate the estimates derived from (1.10), and possibly refute the underlying theory. Thus, rigorous testing against that range of hypotheses would seem necessary, leading to three important difficulties. First, unless a joint test is used, non-rejection of each may occur when there are several failures. Second, if any test rejects at the chosen significance level (controlling for the number of tests undertaken), the validity of all the other tests is cast into doubt. Third, if rejection does occur, it remains a *non sequitur* to assume that the hypothesis which was rejected was the source of the failure, so model revision may require a theory revamp. Again, there seem to be distinct advantages to beginning with general formulations that can be simplified when evidence permits, subject to maintaining identifiability – which can also be a problem with GMM (section 1.4.7 discusses identification).

*Economic theory guidelines.* Finally, seeking a congruent model of the LDGP based on economic theory guidelines by embedding the theory-based model in a more general GUM for the set of candidate variables, with a range of possible specifications of lags, functional forms, breaks, etc., offers many advantages, not least avoiding restrictive assumptions dependent on hope rather than evidence. Such a general-to-specific (Gets) approach can be demanding, and while it can help mitigate problems of under-specification, it is no free lunch, as it leads to a different set of possible problems relating to data-based model selection, discussed in section 1.5. However, the criticism that the LDGP is too complicated for Gets to work well must also apply to all other approaches, as they will not fare any better in such a state of nature unless some remarkable requirements chance to hold (e.g., the complexity of the LDGP happens to lie in directions completely unrelated to the aspects under study). In general, even if one simply wants to test an economic hypothesis as to whether some effect is present, partial inference cannot be conducted alone, unless one is sure about the complete absence of all contaminating influences.

### 1.4.3 Data exactitude

> "She can't do Addition," the Red Queen interrupted. "Can you do Subtraction? Take nine from eight." "Nine from eight I can't, you know," Alice replied very readily. (Lewis Carroll, 1899)

No agency produces perfect data measures on every variable, and although some observations may be both accurate and precise (e.g., specific stock market, or foreign exchange, transactions), most are subject to measurement errors. These can be difficult to handle, especially when there are both revisions and changes in exactitude over time, which thereby introduce an additional source of non-stationarity. Moreover, in any given sample of time series, more recent data will be subject to potentially larger later revisions: section 1.7.1 considers the impact of one example of considerable data revisions.

Mapping theoretical constructs to data counterparts and measuring (or modeling) latent variables both raise further issues. Many commonly used macro variables do not have established measurements, e.g., output gaps, business cycles, capacity utilization, trade union power, etc. Even those that do, such as constructs for consumption, user costs, etc., are open to doubt. These types of measurement errors are not directly caused by inaccurate data collection, but both impinge on empirical studies, and can change over time.

Incentives to improve data quality, coverage and accuracy were noted in section 1.3.4 (see Boumans, 2007, for recent discussions of various measurement issues). In the absence of exact data, there must remain trade-offs between using theory to impose restrictions on badly-measured data, using such data to reject theory specifications, or building data-based models. Again, a balance utilizing both theory and evidence in a progressive process seems advisable.

### 1.4.4 Hidden dependencies

> "Why, it's a Looking-glass book of course! And if I hold it up to a glass, the words will all go the right way again." (Quote from Alice in Lewis Carroll, 1899)

Hidden dependencies abound in all data forms, including cross-sections, time series and panels. An important aspect of sequential conditioning in time series is to explicitly remove temporal dependence, as (1.4) showed, where a martingale difference process is created by the sequential conditioning. In principle, the same concepts apply to cross sections. It must be stressed that "random sampling" by itself does not justify factorizing a joint density or likelihood function. As an extreme form of cross-section dependence, put 1,000 copies of the number "1" in a hat, then draw a random sample of 100 therefrom: one learns nothing after the first draw, although all are "randomly drawn." Sequential factorization correctly reveals that difficulty. Denote the randomly-drawn data sample by $(r_1 \ldots r_N)$; then for any ordering when $\tau$ is the mean value of all the numbers in the hat:

$$\mathsf{D}_{\mathsf{r}}\left(r_1 \ldots r_N \mid \tau\right) = \prod_{i=1}^{N} \mathsf{D}_{\mathsf{r}_i}\left(r_i \mid r_{i-1} \ldots r_1; \tau\right) = \mathsf{D}_{\mathsf{r}_1}\left(r_1; \tau\right), \tag{1.11}$$

since all the other probabilities are precisely unity. As $r_1 = 1$, we correctly deduce $\tau = 1$. Certainly, the other $N - 1$ draws add the information that all the numbers are unity, but would do so even if not randomly drawn.

More generally, the order of an independent sample does not matter, so unlike (1.11), for any ordering the joint density should factorize as:

$$\mathsf{D}_{\mathsf{r}}\left(r_1 \ldots r_N \mid \tau\right) = \prod_{i=1}^{N} \mathsf{D}_{\mathsf{r}_i}\left(r_i \mid r_{i-1} \ldots r_1; \tau\right) = \prod_{i=1}^{N} \mathsf{D}_{\mathsf{r}_i}\left(r_i \mid \tau\right). \tag{1.12}$$

Consequently, potential dependence is testable by conditioning on $s$ "neighbors" after a suitable exogenous ordering to check if their influence is non-zero; i.e., to see whether:

$$\prod_{i=1}^{N} \mathsf{D}_{\mathsf{r}_i}\left(r_i \mid r_{i-1} \ldots r_{i-s}; \tau\right) \neq \prod_{i=1}^{N} \mathsf{D}_{\mathsf{r}_i}\left(r_i \mid \tau\right). \tag{1.13}$$

Suitable tests for the absence of dependence would seem essential before too great a weight is placed on results that base (1.12) on the claim of random sampling, especially when the units are large entities like countries. More generally, when all units are affected in part by macro-forces and their attendant non-stationarities, dependence like (1.13) is likely. If an ordering based on an outside variable is available, then models of $\mathsf{D}_{\mathsf{r}_i}\left(r_i \mid r_{i-1} \ldots r_{i-s}; \tau\right)$ could be investigated directly, similar to some cases of spatial dependence (see Anselin, 2006).

There is a large literature on panel data analysis recently discussed in Choi (2006) and Baltagi (2006).

### 1.4.5 Conditioning variables

> "I'm afraid he'll catch cold with lying on the damp grass," said Alice, who was a very thoughtful little girl. (Lewis Carroll, 1899)

Instrumental variables are a key part of any conditioning set, so require weak exogeneity as well as correlation with the relevant endogenous variables (or the auxiliary assumptions of orthogonality to any unknown vector of excluded influences and independence from the "true" model's errors).

### *1.4.5.1 Weak exogeneity*

The notion of exogeneity, or synonyms thereof, in relation to econometric modeling dates back to the origins of the discipline (see, e.g., Morgan, 1990; Hendry and Morgan, 1995), with key contributions by Koopmans (1950) and Phillips (1957). Weak exogeneity was formalized by Engle, Hendry and Richard (1983), building on Richard (1980) (see Ericsson, 1992, for an exposition), and is a fundamental requirement for efficient conditional inference, which transpires to be at least as important in integrated systems as in stationary processes (see Phillips and Loretan, 1991). Weak exogeneity is equally relevant to instrumental variables estimation, since the marginal density of $\mathbf{z}_t$ then relates to the distribution of the claimed instruments: asserting orthogonality to the error term is often inadequate, as shown by the counter-examples in Hendry (1995a).

Further, $\mathbf{z}_t$ is strongly exogenous for $\theta$ if $\mathbf{z}_t$ is weakly exogenous for $\theta$, and:

$$\mathsf{D}_{\mathbf{z}_t}\left(\mathbf{z}_t \mid \mathbf{X}_{t-1}^{t-s}, \mathbf{X}_0^{1-s}, \mathbf{q}_t, \kappa_2\right) = \mathsf{D}_{\mathbf{z}_t}\left(\mathbf{z}_t \mid \mathbf{Z}_{t-1}^{t-s}, \mathbf{X}_0^{1-s}, \mathbf{q}_t, \kappa_2\right). \tag{1.14}$$

When (1.14) is satisfied, $\mathbf{z}_t$ does not depend upon $\mathbf{Y}_{t-1}$ so $\mathbf{y}$ does not Granger-cause $\mathbf{z}$, following Granger (1969). This requirement sustains marginalizing $\mathsf{D}_{\mathbf{z}_t}(\mathbf{z}_t|\mathbf{X}_{t-1}^{t-s}, \mathbf{X}_0^{1-s}, \mathbf{q}_t, \kappa_2)$ with respect to $\mathbf{Y}_{t-1}^{1}$, but does not concern conditioning. Consequently, Granger causality alone is neither necessary nor sufficient for weak exogeneity, and cannot validate inference procedures (see Hendry and Mizon, 1999).

The consequences of failures of weak exogeneity can vary from just a loss of estimation efficiency through to a loss of parameter constancy, depending on the source of the problem (see Hendry, 1995a, Ch. 5). We now illustrate both extreme cases and one intermediate example.

*Outperforming Gauss–Markov.* First, consider a standard regression setting where Gauss–Markov conditions seem satisfied:

$$\mathbf{y} = \mathbf{Z}\beta + \epsilon \quad \text{with} \quad \epsilon \sim \mathsf{N}_T\left[\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}\right], \tag{1.15}$$

when $\mathbf{Z} = (\mathbf{z}_1 \ldots \mathbf{z}_T)'$ is a $T \times k$ matrix, rank($\mathbf{Z}$) $= k$, and $\epsilon' = (\epsilon_1 \ldots \epsilon_T)$, with:

$$\mathsf{E}[\mathbf{y} \mid \mathbf{Z}] = \mathbf{Z}\beta,$$

and hence $\mathsf{E}[\mathbf{Z}'\epsilon] = \mathbf{0}$. OLS estimates of $\beta$, the parameter of interest here, are:

$$\widehat{\beta} = \beta + \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\epsilon \sim \mathsf{N}_k\left[\beta, \sigma_\epsilon^2\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\right].$$

However, ordinary least squares (OLS) need not be the most efficient unbiased estimator of $\beta$, and an explicit weak exogeneity condition is required to preclude that possibility when $\mathbf{Z}$ is stochastic. For example, let:

$$\mathbf{z}_t = \beta + \nu_t \quad \text{where} \quad \nu_t \sim \mathsf{IN}_k\left[\mathbf{0}, \Sigma\right],$$

estimated by the mean vector:

$$\overline{\beta} = \beta + \overline{\nu} \sim \mathsf{N}_k\left[\beta, T^{-1}\Sigma\right],$$

then it is easy to construct scenarios where $\overline{\beta}$ is much more efficient than $\widehat{\beta}$. Consequently, even in simple regression, for the Gauss–Markov theorem to be of operational use one needs the condition that $\beta$ cannot be learned from the marginal distribution.

*Weak exogeneity in cointegrated systems.* Second, cointegrated systems provide a major forum for testing one aspect of exogeneity. Formulations of weak exogeneity conditions and tests for various parameters of interest in cointegrated systems are discussed in, *inter alia*, Johansen and Juselius (1990), Phillips and Loretan (1991), Hunter (1992), Urbain (1992), Johansen (1992), Dolado (1992), Boswijk (1992) and Paruolo and Rahbek (1999). Equilibrium-correction mechanisms which cross-link equations violate long-run weak exogeneity, confirming that weak exogeneity cannot necessarily be obtained merely by choosing the "parameters of interest." Conversely, the presence of a given disequilibrium term in more than one equation is testable. Consider an apparently well-defined setting with the following bivariate DGP for the I(1) vector $\mathbf{x}_t = (y_t : z_t)'$ from Hendry (1995c):

$$y_t = \beta z_t + u_{1,t} \tag{1.16}$$

$$z_t = \lambda y_{t-1} + u_{2,t}, \tag{1.17}$$

where:

$$\begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ \rho & 1 \end{pmatrix} \begin{pmatrix} u_{1,t-1} \\ u_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}, \tag{1.18}$$

and:

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \sim \mathsf{IN}_2 \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \gamma\sigma_1\sigma_2 \\ \gamma\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right] = \mathsf{IN}_2 \left[ \mathbf{0}, \Sigma \right]. \tag{1.19}$$

The DGP in (1.16)–(1.19) defines a cointegrated vector process in triangular form (see Phillips and Loretan, 1991) which can be written in many ways, of which the following equilibrium-correction form is perhaps the most useful:

$$\begin{aligned} y_t &= \beta z_t + \epsilon_{1,t} \\ \Delta z_t &= \lambda \Delta y_{t-1} + \rho \left( y_{t-1} - \beta z_{t-1} \right) + \epsilon_{2,t}, \end{aligned} \tag{1.20}$$

where $\epsilon_t = \left( \epsilon_{1,t} : \epsilon_{2,t} \right)'$ is distributed as in (1.19).

The parameters of the DGP are $(\beta, \lambda, \rho, \gamma, \sigma_1, \sigma_2)$. When cointegration holds, $\beta$ and $\sigma_1$ can be normalized at unity without loss of generality, and we also set $\sigma_2 = 1$. The parameter of interest is $\beta$, which characterizes the long-run relationship between $y_t$ and $z_t$. Let $\mathcal{I}_{t-1}$ denote available lagged information (the $\sigma$-field generated by $\mathbf{X}_{t-1}$). Then, from (1.19) and (1.20), the conditional expectation of $y_t$ given $(z_t, \mathcal{I}_{t-1})$ is:

$$\mathsf{E}\left[ y_t \mid z_t, \mathcal{I}_{t-1} \right] = \beta z_t + \gamma \Delta z_t - \gamma\rho \left( y_{t-1} - \beta z_{t-1} \right) - \gamma\lambda \Delta y_{t-1}. \tag{1.21}$$

For some parameter values in the DGP, the conditional expectation will coincide with (1.16), whereas for other parameter configurations, (1.16) and (1.21) will differ, in which case it is unsurprising that (1.16) is not fully informative. However, an exact match between the equation to be estimated and the conditional expectation of the dependent variable given $\mathcal{I}_{t-1}$ is not sufficient to justify least squares estimation, even when the error is an innovation against $\mathcal{I}_{t-1}$. Indeed, when $\lambda = \gamma = 0$, but $\rho \neq 0$, there is a failure of weak exogeneity of $z_t$ for $\beta$, even though the conditional expectation is:

$$\mathsf{E}\left[y_t \mid z_t, \mathcal{I}_{t-1}\right] = \beta z_t. \tag{1.22}$$

Nevertheless, $z_t$ is not weakly exogenous for $\beta$ when $\rho \neq 0$ since:

$$\Delta z_t = \rho\left(y_{t-1} - \beta z_{t-1}\right) + \epsilon_{2t}, \tag{1.23}$$

so a more efficient analysis is feasible by jointly estimating (1.16) (or (1.22)) and (1.23). Here the model coincides with both the conditional expectation and the DGP equation, but as shown in Phillips and Loretan (1991) and Hendry (1995c), the violation of weak exogeneity can lead to important distortions to inference when estimating the parameters of (1.16), highlighting the important role of weak exogeneity in conditional inference.

#### *1.4.5.2 Super exogeneity and structural breaks*

Next, processes subject to structural breaks sustain tests for super exogeneity and the Lucas (1976) critique (following Frisch, 1938): (see, e.g., Hendry, 1988; Fischer, 1989; Favero and Hendry, 1992; Engle and Hendry, 1993; Hendry and Santos, (2009). Formally, super exogeneity augments weak exogeneity with the requirement that the parameters of the marginal process can change (usually over some set) without altering the parameters of the conditional. Reconsider (1.9), written with potentially non-constant parameters as:

$$\begin{aligned}\mathsf{D}_{\mathbf{x}_t}\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}^{t-s}; \mathbf{X}_0^{1-s}, \mathbf{q}_t, \rho_t\right) &= \mathsf{D}_{\mathbf{y}_t|\mathbf{z}_t}\left(\mathbf{y}_t \mid \mathbf{z}_t, \mathbf{X}_{t-1}^{t-s}, \mathbf{X}_0^{1-s}, \mathbf{q}_t, \kappa_{1,t}\right) \\ &\quad \mathsf{D}_{\mathbf{z}_t}\left(\mathbf{z}_t \mid \mathbf{X}_{t-1}^{t-s}; \mathbf{X}_0^{1-s}, \mathbf{q}_t, \kappa_{2,t}\right).\end{aligned} \tag{1.24}$$

When $\theta$ enters both $\kappa_{1,t}$ and $\kappa_{2,t}$ in (1.24), inference can again be distorted if weak exogeneity is falsely asserted. When conditional models are constant despite data moments changing considerably, there is *prima facie* evidence of super exogeneity for that model's parameters; whereas, if the model as formulated does not have constant parameters, resolving that failure ought to take precedence over issues of exogeneity. However, while super exogeneity tests are powerful in detecting location shifts, changes to "reaction parameters" of mean-zero stochastic variables are difficult to detect (see, e.g., Hendry, 2000b). Hendry and Santos (2009) propose a test for super exogeneity based on impulse saturation (see Hendry, Johansen and Santos, 2008) to automatically select breaks in the marginal processes, then test their relevance in the conditional. When none of the breaks enters the conditional model, that provides evidence in favor of $\mathbf{z}_t$ causing $y_t$, since the same response occurs to changes across different "regimes" (see, e.g., Heckman, 2000; Hendry, 2004; and the references therein).

### *1.4.5.3 Weak exogeneity and economic theory*

Much economic theory concerns relationships between means such as:

$$\mu_y = \beta' \mu_z. \tag{1.25}$$

A famous example is the permanent income hypothesis (PIH), where $\mu_y$ is permanent consumption and $\mu_z$ is permanent income, so the income elasticity of consumption is unity $\forall\beta$. Most demand and supply functions relate to expected plans of agents; expectations and Euler equation models involve conditional first moments, as do GMM approaches; policy relates planned instruments to expected targets, etc. Since constructs like $\mu_y$ and $\mu_z$ are inherently unobservable, additional assumptions are needed to complete the model. For example, Friedman (1957) uses:

$$y_t = \mu_y + \epsilon_{y,t} \text{ and } z_t = \mu_z + \epsilon_{z,t} \text{ where } \mathsf{E}\left[\epsilon_{y,t}\epsilon_{z,t}\right] = 0, \tag{1.26}$$

which precludes weak exogeneity of $z_t$ for $\beta$ given the dependence between the means in (1.25). Allowing $\mu_y$ to also depend on second moments would not alter the thrust of the following analysis.

Econometrics, however, depends on second moments of observables. Consider the regression:

$$y_t = \gamma' \mathbf{z}_t + v_t \text{ where } v_t \sim \mathsf{IN}\left[0, \sigma_v^2\right]. \tag{1.27}$$

For $\mathbf{z}_t \sim \mathsf{IN}_n\left[\mu_z, \Sigma_{zz}\right]$ with $\mathsf{E}\left[\mathbf{z}_t v_t\right] = \mathbf{0}\ \forall t$:

$$\mathsf{E}\left[y_t \mid \mathbf{z}_t\right] = \gamma' \mathbf{z}_t, \tag{1.28}$$

then, for $\mathbf{y}' = (y_1 \ldots y_T)$ and $\mathbf{Z}' = (\mathbf{z}_1 \ldots \mathbf{z}_T)$:

$$\widehat{\gamma} = \left(\mathbf{Z}'\mathbf{Z}\right)^{-1} \mathbf{Z}'\mathbf{y}, \tag{1.29}$$

so that second moments are used to estimate $\gamma$. Here, (1.27) entails $\mathsf{E}[y_t] = \gamma' \mathsf{E}[\mathbf{z}_t]$, and from (1.28):

$$\mathsf{E}\left[\mathbf{z}_t y_t\right] = \mathsf{E}\left[\mathbf{z}_t \mathbf{z}_t'\right]\gamma \text{ or } \sigma_{yz} = \Sigma_{zz}\gamma, \tag{1.30}$$

both of which involve $\gamma$. Thus, there seems to be no difference between how means and variances are related, which is why second moments can be used to infer about links between first moments. However, when any relation like (1.25) holds, then $\sigma_{yz}$ and $\Sigma_{zz}$ in (1.30) must be connected by $\beta$, not $\gamma$, if valid inferences are to result about the parameters of interest $\beta$. Weak exogeneity is needed, either directly in (1.27), or indirectly for "instrumental variables." This is more easily seen from the joint distribution:

$$\begin{pmatrix} y_t \\ \mathbf{z}_t \end{pmatrix} \sim \mathsf{IN}_{n+1}\left[\begin{pmatrix} \mu_y \\ \mu_z \end{pmatrix}, \begin{pmatrix} \sigma_{yy} & \sigma_{yz}' \\ \sigma_{zy} & \Sigma_{zz} \end{pmatrix}\right], \tag{1.31}$$

so given (1.25):

$$\mathsf{E}\left[y_t \mid \mathbf{z}_t\right] = \mu_y - \gamma'\mu_z + \gamma'\mathbf{z}_t = (\beta - \gamma)'\,\mu_z + \gamma'\mathbf{z}_t, \tag{1.32}$$

which coincides with (1.28) only if $\gamma = \beta$, so means and variances are then related by identical parameters.

Note from (1.31):

$$\mathbf{z}_t = \mu_z + \mathbf{u}_t, \tag{1.33}$$

so impulse responses cannot be identified uniquely as originating from perturbing $\mathbf{u}_t$ or $\mu_z$. But from (1.32), the response of $y_t$ to these perturbations in (1.33) will differ unless $\gamma = \beta$, so weak exogeneity is essential for unique impulse responses, which cannot be based on an arbitrary choice of Cholesky decompositions (only one variant could coincide with valid conditioning).

### 1.4.6 Functional form

> Alice began to remember that she was a Pawn, and that it would soon be time to move. (Lewis Carroll, 1899)

In practice, one cannot expect every functional form specification to coincide with that which generated the data, however well-based its logic or theory credentials. There are theories of what various linear and other approximations deliver (see, e.g., White, 1980b, 2008), but such approximations cannot ensure non-systematic residuals. Automatic model selection has been extensively applied to select functional forms from quite large classes using data evidence (see, *inter alia*, Perez-Amaral, Gallo and White, 2003, 2005; Castle, 2005; Castle and Hendry, 2005, and section 1.5). In low-dimensional models, semiparametric and nonparametric methods are often used to avoid specifying the functional form, but can be susceptible to unmodeled outliers and breaks.

### 1.4.7 Identification

> ... watching one of them that was bustling about among the flowers, poking its proboscis into them, "just as if it was a regular bee," thought Alice. However, this was anything but a regular bee: in fact, it was an elephant. (Lewis Carroll, 1899)

Identification has three attributes of uniqueness, interpretation, and correspondence to reality (see, e.g., Hendry, 1995a), which we discuss in turn. Since unidentified parameters entail a non-unique model specification – so what is estimated need not match the parameters of the generating process – identification is a fundamental attribute of a parametric specification.

First, a general understanding of identification as uniqueness has been developed (see, e.g., Fisher, 1966; Rothenberg, 1971; Sargan, 1983: Hsiao, 1983, provides an overview), building on the rank and order conditions so well known to be necessary and sufficient in simultaneous systems when the restrictions are given by subject-matter theory: we could call this a technical issue. Cowles Commission researchers showed that the "reduced form" (or statistical system) was always identified in their formulation, and that all just-identified models were isomorphic to that statistical system, hence tests of overidentified "structural forms" could be derived by comparing their two likelihoods. Their analysis, therefore, entailed that the "structural form" is actually a *reduction* of the statistical system, so logically can be obtained from it without any prior knowledge of the relevant restrictions. Thus, when a model is identified relative to an identified system, the identification restrictions in question do not have to be known *a priori*, but can be found by a suitable algorithm (see, e.g., Hendry and Krolzig, 2005) – indeed, several overidentified, but distinct, representations can coexist (see, e.g., Hendry, Lu and Mizon, 2008). Such a conclusion is predicated on the statistical system itself being identified, which requires sufficient explanatory variables – the vexed topic of "exogeneity" discussed in section 1.4.5 above. Moreover, for an overidentification test to be valid, the statistical system must be well specified, so needs to be modeled and evaluated first (see, e.g., Spanos, 1990), after which it can be reduced to a "structural form." Nevertheless, that prior identification restrictions must be known in advance remains the dominant belief, which if true, would preclude empirical modeling not preceded by a rigorous theory derivation that entailed sufficient restrictions.

Second, interpretation is a regular seminar question along the lines: "How do you know you have identified the demand curve" (as opposed to some other entity)? This is essentially an economic theory issue, and only substantive theory can resolve such a debate. It is separate from uniqueness: a regression of price on quantity is always unique, but hardly qualifies as a demand curve just because the regression coefficient is negative.

Third, even if both uniqueness and interpretation are confirmed, the result still need not correspond to reality, which is an empirical issue (and related to the usage of the word "identification" in, say, Box and Jenkins', 1976, analysis, as well as the quote above). An estimated equation may be unique and interpretable but not the relevant relation. Thus, all aspects of model building are involved in establishing satisfactory identification.

Recently, problems of weak instruments, and the resulting issue of identification, have become salient (see, among others, Staiger and Stock, 1997; Stock and Wright, 2000; Stock, Wright and Yogo, 2002; Kleibergen, 2002; Mavroeidis, 2004).

### 1.4.8 Parameter constancy

> "Yes, all his horses and all his men," Humpty Dumpty went on. "They'd pick me up again in a minute, they would!" (Lewis Carroll, 1899)

Parameters are the entities which must be constant if the specified model is to be a useful characterization of reality. However, that does not preclude the coefficients in any model formulation from changing, as in "random coefficients" models or "structural time series" (see, e.g., Hildreth and Houck, 1968; Harvey, 1993). The main problem for economic forecasting using econometric models is that coefficients of deterministic terms do not seem to stay constant, but suffer location shifts, which in turn induce forecast failure (see, e.g., Clements and Hendry, 2005, 2006). While changes in zero-mean variables seem less damaging to forecasts (see, e.g., Hendry and Doornik, 1997), such breaks nevertheless remain pernicious for policy analyses.

### 1.4.9 "Independent" homoskedastic errors

> "Contrariwise," continued Tweedledee, "if it was so, it might be; and if it were so, it would be: but as it isn't, it ain't. That's logic." (Lewis Carroll, 1899)

Joint densities can always be factorized into sequential forms, as with martingale difference sequences. Moreover, equations can often be standardized to be homoskedastic by dividing by contemporaneous error variances (when these exist), so this category may be one of the least stringent requirements.

### 1.4.10 Expectations formation

> "What sort of things do you remember best?" Alice ventured to ask. "Oh, things that happened the week after next" the Queen replied in a careless tone. (Lewis Carroll, 1899)

Surprisingly little is known about how economic agents actually form their expectations for variables relevant to their decisions. Almost no accurate expectations data exist outside financial market traders, so resort is usually needed to proxies for the unobserved expectations, or to untested assumptions, such as "rational" expectations (RE), namely the correct conditional expectation $\mathsf{E}[\cdot]$ of the variable in question ($y_{t+1}$) given the available information ($\mathcal{I}_t$). There is a large gap between economic theory models of expectations – which often postulate that agents hold RE – and the realities of economic forecasting, where forecast failure is not a rare occurrence. The "rational" expectation is often written as (see Muth, 1961):

$$y^{re}_{t+1} = \mathsf{E}\left[y_{t+1} \mid \mathcal{I}_t\right], \tag{1.34}$$

which implicitly assumes free information and free computing power as available information is vast. The usual argument, perhaps loosely worded to avoid contradictions, is that otherwise there would be arbitrage opportunities, or agents would suffer unnecessary losses. But expectations are instrumental to agents' decisions, and the accuracy thereof is not an end in itself, so agents should just equate the marginal benefits of improved forecast accuracy against the extra costs of achieving that, leading to "economically rational expectations" (ERE) (see Aghion *et al.*, (2002)). "Model consistent expectations" instead impose the expectations formation process as the solved estimated model specification, so – unless the model is perfect – suffer the additional drawback of imposing invalid restrictions.

While ERE may be more realistic than RE, it still assumes knowledge of the form of dependence of $y_{t+1}$ on the information used: as expressed in $\mathsf{E}\left[y_{t+1}|\mathcal{I}_t\right]$ in (1.34), that assumption corresponds to agents knowing precisely what the conditioning operator is. In a stationary world, one could imagine learning mechanisms that eventually led to its discovery (see, e.g., Evans and Honkapohja, 2001). However, in a wide-sense non-stationary environment, an explicit statement of the form of (1.34) is:

$$y_{t+1}^{re} = \mathsf{E}_{t+1}\left[y_{t+1} \mid \mathcal{I}_t\right] = \int y_{t+1} f_{t+1}\left(y_{t+1} \mid \mathcal{I}_t\right) \mathrm{d}y_{t+1}. \tag{1.35}$$

Thus, when $f_t(\cdot) \neq f_{t+1}(\cdot)$, agents need to *know the future* conditional density function $f_{t+1}(y_{t+1}|\mathcal{I}_t)$, given present information, to obtain the appropriate conditioning relation, since only then will $y_{t+1}^{re}$ be an unbiased predictor of $y_{t+1}$. That $f_t(\cdot) \neq f_{t+1}(\cdot)$ is precisely why forecasting is so prone to problems. Unfortunately, knowing $f_{t+1}(\cdot)$ virtually requires agents to have crystal balls that genuinely "see into the future." When distributions are changing over time, agents can at best form "sensible expectations," $y_{t+1}^{se}$, based on forecasting $f_{t+1}(\cdot)$ by $\widehat{f}_{t+1}(\cdot)$ from some rule, such that:

$$y_{t+1}^{se} = \int y_{t+1} \widehat{f}_{t+1}\left(y_{t+1} \mid \mathcal{I}_t\right) \mathrm{d}y_{t+1}. \tag{1.36}$$

There are no guaranteed good rules for estimating $f_{t+1}(y_{t+1}|\mathcal{I}_t)$ when $\{y_t\}$ is wide-sense non-stationary. In particular, when the conditional moments of $f_{t+1}(y_{t+1}|\mathcal{I}_t)$ are changing in unanticipated ways, setting $\widehat{f}_{t+1}(\cdot) = f_t(\cdot)$ could be a poor choice, yet that underlies most of the formal derivations of RE, which rarely distinguish between $f_t(\cdot)$ and $f_{t+1}(\cdot)$. Outside a stationary environment, agents cannot solve (1.34), or often even (1.35). The drawbacks of (1.34) and (1.35), and the relative success of robust forecasting rules (see, e.g., Clements and Hendry, 1999; Hendry, 2006), suggest agents should use them, an example of imperfect-knowledge expectations (IKE) (see Aghion *et al.*, 2002; Frydman and Goldberg, 2007). IKE acknowledges that agents cannot know how $\mathcal{I}_t$ enters $f_t(\cdot)$ when processes are evolving in a non-stationary manner, let alone $f_{t+1}(\cdot)$, which still lies in the future. Collecting systematic evidence on agents' expectations to replace the unobservables by estimates, rather than postulates, deserves greater investment (see, e.g., Nerlove, 1983).

Finally, take expectations conditional on the available information set $\mathcal{I}_{t-1}$ in a regression model with valid weak exogeneity:

$$y_t = \beta' \mathbf{z}_t + \epsilon_t, \tag{1.37}$$

so that:

$$\mathsf{E}\left[y_t \mid \mathcal{I}_{t-1}\right] = \beta' \mathsf{E}\left[\mathbf{z}_t \mid \mathcal{I}_{t-1}\right], \tag{1.38}$$

as $\mathsf{E}\left[\epsilon_t|\mathcal{I}_{t-1}\right] = 0$. Writing (1.38) as $y_t^e = \beta' \mathbf{z}_t^e$, the conditional model (1.37) always has an expectations representation, although the converse is false. Importantly, therefore, contemporaneous conditioning variables can also be expectations variables, and some robust forecasting rules like $\Delta\widehat{p}_{t+1} = \Delta p_t$ have that property.

The New Keynesian Phillips curve is perhaps the best-known model which includes expected inflation to explain current inflation. Models of this type are usually estimated by replacing the expected value by the actual future outcome, then using IV or GMM to estimate the resulting parameters, as in, say, Galí, Gertler and Lopez-Salido (2001). As shown in Castle *et al.* (2008), since breaks and regime shifts are relatively common, full-sample estimates of equations with future values can deliver spuriously significant outcomes when breaks are not modeled, a situation detectable by impulse saturation (see section 1.5).

### 1.4.11 Estimation

> "I was wondering what the mouse-trap was for," said Alice. "It isn't very likely there would be any mice on the horse's back."
>
> "Not very likely, perhaps," said the Knight; "but if they do come, I don't choose to have them running all about." (Lewis Carroll, 1899)

Developing appropriate estimators comprises a major component of extant econometric theory, and given any model specification, may seem an uncontentious task. However, only in recent decades has it been clear how to avoid (say) nonsense correlations in non-stationarity data, or tackle panel dependencies, so unknown pitfalls may still lurk.

## 1.5 Model selection

> In another moment Alice was through the glass, and had jumped lightly down into the Looking-glass room. (Lewis Carroll, 1899)

Model selection is the empirical route whereby many of the simplifications in sections 1.4.2.2 and 1.4.2.3 are implemented in practice. In that sense, it is not a distinct step *per se*, but a way of carrying out some of the earlier steps, hence our treating the topic in a separate section.

Selection remains a highly controversial topic. It must be granted that the best approaches cannot be expected to select the LDGP on every occasion, even when the GUM nests the LDGP, and clearly cannot do so ever when the LDGP is not a nested special case. However, that statement remains true when the GUM is exactly the LDGP, but conventional inference is nevertheless undertaken to check that claim. If the LDGP were known at the outset of a study, apart from the unknown values of its parameters, then if any specification or misspecification testing was undertaken, one could only end by doubting the claim that the initial formulation was indeed the LDGP. The least worst outcome would be weak confirmation of the prior specification, and otherwise either some included variables will be found insignificant, or some assumptions will get rejected, casting doubt on the claim. That is the risk of undertaking statistical inference. The alternative of not testing claimed models is even less appealing, namely never learning which ones are useless. To quote Sir Francis Bacon: "If a man will begin with certainties, he shall end in doubts; but if he will be content to begin with doubts he shall end in certainties."

Conversely, the list at the beginning of section 1.4 makes it clear that "model uncertainty" comprises much more than whether one selected the "correct model" from some set of candidate variables that nested the LDGP. If, say, 1,000 possibly lagged, nonlinear functions of a set of candidate exogenous variables in a model with many breaks are checked for relevance at a significance level of 0.1%, and all are indeed irrelevant, then on average *one* will be retained adventitiously, so uncertainty is greatly reduced by eliminating about 999 potential influences. The entire point of model selection is to reduce some of the uncertainties about the many aspects involved in model specification, and the cost for doing so is a "local increase" in uncertainty as to precisely which influences should be included and which excluded around the margin of significance. Thus, embedding the claimed theory in a more general specification that is congruent with all the available evidence offers a chance to both utilize the best available theory insights and learn from the empirical evidence. Since such embedding can increase the initial model size to a scale where a human has intellectual difficulty handling the required reductions, we next consider computerized, or automatic, methods for model selection.

### 1.5.1 Automatic model selection

> "Does – the one – that wins – get the crown?" she asked, as well as she could, for the long run was putting her quite out of breath.
>
> "Dear me, no!" said the King. "What an idea!" (Alice to the White King in Lewis Carroll, 1899)

The many alternatives now available include, but are not restricted to, Phillips (1994, 1995, 1996), Tibshirani (1996), Hoover and Perez (1999, 2004), Hendry and Krolzig (1999, 2001), White (2000), Krolzig (2003), Kurcewicz and Mycielski (2003), Demiralp and Hoover (2003), and Perez-Amaral *et al.* (2003); also see the special issue on model selection edited by Haldrup, van Dijk and Hendry (2003) (the references cited therein provide bibliographic perspective on this huge literature). Complaints about model selection have a long pedigree, from Keynes (1939) about "data-based modeling" and Koopmans (1947) on "measurement without theory," through "pre-test biases" from test-based selection in Judge and Bock (1978); "repeated testing" inducing adventitious significance in Leamer (1978, 1983) and Lovell (1983) criticizing selection rules seeking "significance," to Pagan (1987) on the potential "path dependence of any selection"; Hendry, Leamer and Poirier (1990) debating "arbitrary significance levels"; Chatfield (1995) criticizing "ignoring selection effects" as misrepresenting uncertainty, and Faust and Whiteman (1997) on "lack of identification," but most have now been rebutted (see, e.g., Hendry, 2000a). Concerning Keynes' comment quoted above, not only should everyone get the same answer from an automatic algorithm applied to the same GUM using the same selection criteria, investigators with different GUMs, which differed only by irrelevant variables, could also end with the same model.

Here we consider Autometrics, an Ox package (see Doornik, 2006, 2007a) implementing automatic Gets modeling based on the theory of reduction discussed above. The present implementation of Autometrics is primarily for linear regression models, but extensions have been derived theoretically to automatically model dynamic, cointegrated, simultaneous systems; nonlinear equations; structural breaks; more variables ($N$) than observations ($T$); and testing exogeneity (see, e.g., Hendry and Krolzig, 2005; Castle and Hendry, 2005; Hendry, *et al.*, 2008; Johansen and Nielsen, 2008; Doornik, 2007b; and Hendry and Santos, 2009, respectively). Given any available theoretical, historical, institutional, and measurement information, as well as previous empirical evidence, a GUM must be carefully formulated, preferably with a relatively orthogonal parameterization, a subject-matter basis, and must encompass existing models. When $T \gg N$, the GUM can be estimated from all the available evidence, and rigorously tested for congruence. If congruence fails, a new formulation is required: but at least one has learned the general inadequacy of a class of models. If congruence is accepted, it is then maintained throughout the selection process by not following simplification paths which are rejected on diagnostic checking (using the same statistics), ensuring a congruent final model. When $N > T$, as must happen when impulse saturation is used and can occur more generally (discussed below), misspecification testing can only be undertaken once a feasible model size $n < T$ has been reached.

Statistically insignificant variables are eliminated by selection tests, using a tree-path search in Autometrics, which improves on the multi-path procedures in Hoover and Perez (1999) and Hendry and Krolzig (2001). Checking many paths prevents the algorithm from becoming stuck in a sequence that inadvertently eliminates a variable which actually matters, and thereby retains other variables as proxies (as in stepwise regression). Path searches terminate when no variable meets the elimination criteria. Non-rejected (terminal) models are collected, then tested against each other by encompassing: if several remain acceptable, so are congruent, undominated, mutually encompassing representations, the search is terminated using, e.g., the Schwarz (1978) information criterion, although all are reported and can be used in, say, forecast combinations.

To understand why an automatic search procedure might work, consider a case where the complete set of $N$ candidate regressors is mutually orthogonal, but which ones are relevant is unknown *a priori*, and $T \gg N$. The postulated GUM nests the LDGP. Estimate the GUM, then, squaring to eliminate signs, rank the resulting $\mathsf{t}_i^2$ statistics from the largest to the smallest. When $c_\alpha$ is the criterion for retention, let $n$ be such that $\mathsf{t}_n^2 \geq c_\alpha$ when $\mathsf{t}_{n+1}^2 < c_\alpha$. Then select the model with those $n$ regressors. That required precisely *one* decision – what to include, and hence what to exclude. No issues of search, repeated testing, path dependence, etc., arise. Goodness-of-fit is not directly used to select models; and no attempt is made to "prove" that a given number of variables matters. In practice, the role of the tree search is to ascertain "true" relevance when orthogonality does not hold; and the choice of $c_\alpha$ affects $\mathsf{R}^2$ and $n$ through retention of $\mathsf{t}_n^2$. Generalizations to other maximum likelihood estimators, or approximations thereto such as IV, are feasible (see Hendry and Krolzig, 2005; Doornik, 2007a).

However, it does matter that selection occurs: the selected model's estimates do not have the same properties as if the LDGP equation had been estimated without any testing. Sampling vagaries entail that some variables which enter the LDGP will by chance have a sample $t^2 < c_\alpha$ (low power). Since they are only retained when $t^2 \geq c_\alpha$, their estimated magnitudes will be biased away from the origin, and hence selected coefficients need to be bias corrected, which is relatively straightforward (see Hendry and Krolzig, 2005). Some variables which are irrelevant will have $t^2 \geq c_\alpha$ (adventitiously significant), where the probability of that event is $\alpha\left(N - n^*\right)$ when $n^*$ variables actually matter. Fortunately, bias correction will also drive such estimates sharply towards the origin. Thus, despite selecting from a large set of potential variables, nearly unbiased estimates of coefficients and equation standard errors can be obtained with little loss of efficiency from testing many irrelevant variables, and some loss for relevant, from the increased value of $c_\alpha$. The normal distribution has "thin tails," so the power loss from tighter significance levels is usually not substantial, whereas financial variables may have fat tails, so power loss could be more costly at tighter $\alpha$.

Impulse saturation is described in Hendry *et al.* (2008) and Johansen and Nielsen (2008) as including an indicator for every observation, entered (in the simplest case) in blocks of $T/2$, with the significant outcomes retained. This approach both helps remove outliers, and is a good example of why testing large numbers of candidate regressors does not cost much efficiency loss under the null that they are irrelevant. Setting $c_\alpha \leq 1/T$ maintains the average false null retention at one "outlier," and that is equivalent to omitting one observation, so is a tiny efficiency loss despite testing for the relevance of $T$ variables. Since all regressors are exact linear functions of $T$ impulses, that effect carries over directly in the independent and identically distributed (i.i.d.) setting, and in similar ways more generally. Thus, $N > T$ is not problematic for automatic model selection, opening the door to large numbers of new applications.

Since an automatic selection procedure is algorithmic, simulation studies of its operational properties are straightforward. In the Monte Carlo experiments reported in Hendry and Krolzig (2005), commencing from highly overparameterized GUMs (between 8 and 40 irrelevant variables; zero and 8 relevant), PcGets recovered the LDGP with an accuracy close to what one would expect if the LDGP specification were known initially, but nevertheless coefficient tests were conducted. To summarize its simulation-based properties, false rejection frequencies of null hypotheses (measured as retention rates for irrelevant variables) can be controlled at approximately the desired level; correct rejections of alternatives are close to the theoretical upper bound of power (measured as retention rates for relevant variables); model selection is consistent for a finite model size as the sample size grows without bound; nearly unbiased parameter estimates can be obtained for all variables by bias-correction formulae, which also reduce the mean square errors of adventitiously retained irrelevant variables; and reported equation standard errors are nearly unbiased estimates of those of the correct specification (see, e.g., Hendry and Krolzig, 2005). Empirically, automatic Gets selects (in seconds) models at least as good as those developed over several years by their authors (see Ericsson, 2007, for several examples). Although automatic model selection is in its infancy, exceptional progress has already been achieved (see Hoover and Perez, 1999; Hoover and Perez, 2004, provide additional evidence).

### 1.5.2 Costs of inference and costs of search

> "Don't keep him waiting, child! Why, his time is worth a thousand pounds a minute!" (Train passengers to Alice in Lewis Carroll, 1899)

Costs of inference are inevitable when tests have non-zero size and non-unit power, even if investigators commence from the LDGP – but do not know that is the correct specification, so have to test for congruence and significance. Costs of search are due to commencing from any GUM that is over-parameterized relative to the LDGP. Under-specification ensures that an invalid model of the LDGP will result. Given the many criticisms of model selection, it may surprise readers that costs of search are small in comparison to costs of inference: the main difficulty is not selection *per se*, but the vagaries of sampling. In selecting a model from a GUM, there are two possible mistakes. The first is including irrelevant variables (ones not in the LDGP), the second is omitting relevant variables. Since the first group are absent when the DGP is the GUM, that is purely a cost of search. The second is primarily a cost of inference, with possible additional search costs if there are lower probabilities of retaining relevant variables when commencing from the GUM.

When the nominal rejection frequency of individual selection tests is set at $\alpha \leq 1/N \to 0$ as $T \to \infty$, on average at most one irrelevant variable will be retained as adventitiously significant out of $N$ candidates. Thus, there is little difficulty in eliminating almost all of the irrelevant variables when starting from the GUM (a small cost of search). The so-called overall "size" of the selection procedure, namely $1-(1-\alpha)^N$, can be large, but is uninformative about the success of a simplification process that on average correctly eliminates $(1-\alpha)N$ irrelevant variables.

Conversely, even for a loose significance level like $\alpha = 0.05$, and commencing from the LDGP, there is only a 50% chance of keeping a relevant variable where the t-test on its coefficient has a non-centrality of 2 (a high cost of inference). A more stringent critical value (say $\alpha = 0.01$, so $c_\alpha \simeq 2.63$) worsens the costs of inference as the retention probability falls to 27% despite the correct specification being postulated. Costs of inference usually exceed costs of search, the exception being when all relevant variables have large non-central t-statistics (in excess of about $\pm 5$), so there are no costs of inference. The probabilities of locating the LDGP commencing from the GUM are reasonably close to the corresponding outcomes when the search commences from the LDGP. Since the LDGP is sometimes never retained even when it is the initial specification, the apparent problem of a search algorithm may be a cost of inference.

The limits of automatic model selection must also be clarified. If the LDGP equation would not be reliably selected by the given inference rules applied to itself as the initial specification, then selection methods cannot rectify that. Many apparent criticisms of selection have failed to note that key limitation. In the simulations described above, the same algorithm and selection criteria were always applied to commencing from both the GUM and the LDGP, and only the additional costs attributable to starting from the former comprise search costs. Also, when there are relevant variables with small t-statistics because the parameters are $\mathsf{O}(1/\sqrt{T})$, especially if they are highly correlated with other regressors (see Pötscher, 1991; Leeb and Pötscher, 2003, 2005), then selection is not going to work well: one cannot expect success in selection if a parameter cannot be consistently estimated. Thus, although uniform convergence seems infeasible, selection works for parameters larger than $\mathsf{O}(1/\sqrt{T})$ (as they are consistently estimable) or smaller than $\mathsf{O}(1/T)$ (as they vanish), yet $1/\sqrt{T}$ and $1/T$ both converge to zero as $T \to \infty$, so "most" parameter values are unproblematic. If the LDGP would always be retained by the algorithm when commencing from it, then a close approximation will generally be selected when starting from a GUM which nests that LDGP.

Additional problems for any empirical modeling exercise arise when the LDGP is not nested in the GUM, due to the regressor set being incomplete, the functional form misspecified or structural breaks and other non-stationarities not being fully accommodated, as well as serious measurement errors contaminating the data or endogenous variables being incorrectly treated as regressors. For very high levels of collinearity between relevant and irrelevant variables, the selected approximation may use the incorrect choice if that is undominated, but in a progressive research strategy when there are intermittent structural breaks in both relevant and irrelevant variables, such a selection will soon be dominated. Phillips (2003) provides an insightful analysis of the limits of econometrics.

## 1.6 Teaching "Applied Econometrics"

> "Manners are not taught in lessons," said Alice. "Lessons teach you to do sums, and things of that sort."
>
> "And you do Addition?" the White Queen asked. "What's one and one and one and one and one and one and one and one and one and one?"
>
> "I don't know," said Alice. "I lost count." (Lewis Carroll, 1899)

Both economic theory and theoretical econometrics are relatively structured subjects to teach, whereas applied econometrics is not, so many approaches are extant. The obvious way might be to include substantive empirical findings in the relevant subject-matter part of other economics courses, and so effectively abolish the need to teach what applied econometrics has established. This certainly happens in part, usually with a lag after the relevant study was published, but seems less common than courses specifically oriented to applied econometrics. I was taught in such a course, the bulk of which concerned studying how the "masters" had conducted their investigations, and what they found – essentially an apprenticeship. Other courses focus more on the economic and econometric theory behind key studies, with less attention to their empirical outcomes: systems of demand equations seem to be addressed that way. Presumably the aim is to explicate the relation between economics and its applications in particular cases. Finally, some courses require students to undertake empirical work themselves, often replicating or evaluating existing studies rather than novel research. Combinations of some or all of these also happen.

If the objective is one where completing students are to be able to reliably tackle a new application, then teaching applied econometrics becomes very demanding. A wide range of skills and insights need to be conveyed, many of which concern "auxiliary" issues such as data availability, its quality and its correspondence to the target of the analysis, including frequency, seasonality, transformations, etc.; institutions and policy agencies that impinge on the economic process; important historical and non-economic contingencies that occurred; the specification of the candidate list, their dynamics, exogeneity, functional forms and constancy of possible models; and the use of software. When the first attempt fails on the desired criteria, a revision process is needed, so difficulties of expanding searches and sequentially correcting problems with a model must be confronted, all too often leaving the student floundering.

A key job of an applied econometrician is to formulate the general model that underpins the analysis, which includes the specification of all candidate variables that might influence the "target" variables of interest in the modeling problem, their general functional forms (e.g., logs), and putative exogeneity assumptions. General economic reasoning plays a substantive part at this stage. Further, one must collect and carefully check all the data series to be modeled, and investigate their historical context. Finally, the easier part is using appropriate software to congruently model the relevant series. Yet, many studies by "experts" remain clever "detective exercises" in which a feel for the evidence helped point towards a viable conclusion. The approach in Hendry and Nielsen (2007a), summarized in Hendry and Nielsen (2007b), is to first prepare students to understand the elements of likelihood theory, using likelihood ratio tests for inference and evaluation – testing the assumptions for the validity of those inferences – leading to model selection in the econometric theory part of the course. A sequence of increasingly realistic theoretical models is developed from i.i.d. binary data through to cointegrated equations with structural breaks. On the applied part of the course, we thoroughly examine an agreed data set, and after teaching the relevant software, students can rapidly move from simple models to general ones using automatic methods. Our focus is on developing well-specified empirical models of interesting economic issues. Given a new problem, students then have a structured approach to follow in their investigations. We consider there has been a marked improvement in their resulting empirical studies.

An example of my own approach is recorded in Hendry (1999): there may be some "tacit knowledge" therein (and I hope there is value added), but most of the above steps can be formalized without the need for an extensive apprenticeship. The next section focuses on comparing how well automatic model selection does without any "prior" historical knowledge or empirical modeling experience. The results reported in section 1.7 took about 20 minutes of real time, including the write-up: even granted that the data were pre-prepared and the log transformations were entailed by the pre-analysis, the efficiency gains over my previous "handcrafted" study are huge. In addition, hypotheses that were previously imposed or sequentially investigated can be evaluated jointly.

## 1.7 Revisiting the "experiment in applied econometrics"

> "If I wasn't real," Alice said – half-laughing though her tears, it all seemed so ridiculous – "I shouldn't be able to cry."
>
> "I hope you don't suppose those are real tears?" Tweedledum interrupted in a tone of great contempt. (Lewis Carroll, 1899)

The recent huge increases in the prices of many foods makes the exercise of re-examining US food expenditure over 1931–89 (based on the update of Tobin, 1950, by Magnus and Morgan, 1999) of more than just historical interest. If the various price and income elasticities estimated below are approximately correct, then substantial responses can be anticipated (indeed, could this be the long-sought solution to society's burgeoning obesity problem?). Hendry (1999) sought to explain why other contributors to Magnus and Morgan (1999) had found their models were non-constant over the combined inter-war and post-war samples, so had eschewed modeling the 1930s data. Impulse dummies for a food program and post-war de-rationing allowed a constant equation to be developed over the sample 1931–89. While an indicator variable is a crude level of measurement, the converse strategy of not modeling major institutional interventions seems even less attractive. Theory and common sense suggest that food programs and switches in rationing matter; but few theory models allow for such factors in a way suitable for empirical implementation (although the original analyst of this data also published on rationing in Tobin, 1952).

The per capita variables are as follows (lower case denotes logs):

$e_f$: constant-price expenditure on food
$p_f - p$: real price of food
$e$: total constant-price expenditure
$s = \log Y - \log E$: (an approximation to the savings ratio)
$a$: average family size.

Figure 1.6 shows the time series, and reveals considerable changes over the period. After falling sharply at the commencement of the Great Depression, both $e_f$ and $e$ rise substantially till World War II, fall after, then resume a gentle rise (panels (a) and (c)), so $\Delta e_{f,t}$ is vastly more volatile pre-war (panel (e)) ($\Delta e_t$ has a similar but less pronounced pattern). Next, $p_f - p$ is quite volatile till after the war, then is relatively stable (panel (b)), whereas the dramatic rise in $s$ from "forced saving" during the war is manifest (panel (d)).

The earlier study of a cointegrated VAR for the system established that $e$, $s$, and $p_f - p$ were weakly exogenous in the food demand equation. Here, the general conditional model allowed for two lags on each of $e_f$, $e$, $p_f - p$, $s$ and one lag

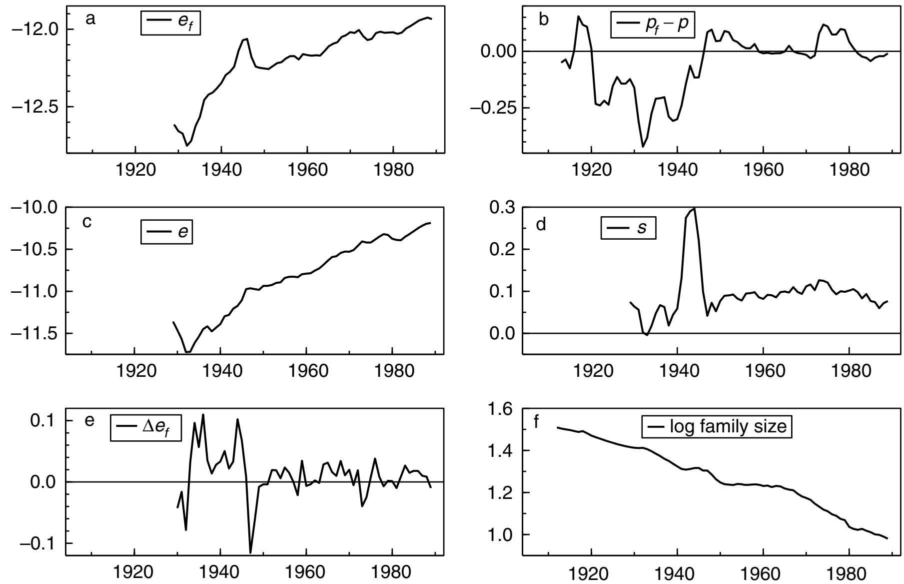


*Figure 1.6* Food expenditure and related time series

on $a$, and was selected by Autometrics at 1% for all candidate variables, including impulse saturation. All diagnostic tests were insignificant, and the PcGive unit root test strongly rejected the null of no cointegration ($t_{ur} = -11.37^{**}$: (see Banerjee and Hendry, 1992; Ericsson and MacKinnon, 2002) with the long-run solution:

$$c_0 = e_f + 7.99 - 0.4e + 0.36(p_f - p). \tag{1.39}$$

Transforming to differences and the equilibrium-correction term from (1.39), Autometrics selected over 1931–89 (at 2.5%, again including impulse saturation):

$$\begin{aligned}
\Delta e_{f,t} = \; & \underset{(0.02)}{0.34}\, s_{t-1} - \underset{(0.02)}{0.32}\, c_{0,t-1} + \underset{(0.04)}{0.67}\, \Delta e_t + \underset{(0.03)}{0.13}\, \Delta e_{t-1} \\
& - \underset{(0.03)}{0.64}\, \Delta(p_f - p)_t - \underset{(0.01)}{0.09}\, \mathrm{I}_{31} - \underset{(0.01)}{0.10}\, \mathrm{I}_{32} + \underset{(0.01)}{0.04}\, \mathrm{I}_{34} \\
& + \underset{(0.01)}{0.03}\, \mathrm{I}_{41} + \underset{(0.01)}{0.05}\, \mathrm{I}_{42} + \underset{(0.01)}{0.03}\, \mathrm{I}_{51} + \underset{(0.01)}{0.02}\, \mathrm{I}_{52} + \underset{(0.01)}{0.03}\, \mathrm{I}_{70}
\end{aligned}$$

$$\begin{aligned}
&\left(\mathsf{R}^*\right)^2 = 0.96 \;\; \mathsf{F_M}(13, 45) = 94.9^{**} \;\; \widehat{\sigma} = 0.0078 \;\; \mathsf{F_{ar}}(2, 44) = 1.34 \\
&\chi^2(2) = 1.04 \;\; \mathsf{F_{arch}}(1, 44) = 2.25 \;\; \mathsf{F_{reset}}(1, 45) = 0.35 \\
&\mathsf{F_{het}}(18, 27) = 0.48 \;\; \mathsf{F_{Chow}}(9, 37) = 0.99.
\end{aligned} \tag{1.40}$$

In (1.40), $\left(\mathsf{R}^*\right)^2$ is the squared multiple correlation when a constant is added, $\mathsf{F}_{\mathsf{M}}(13,45)$ is the associated test of the null, and $\widehat{\sigma}$ is the residual standard deviation, with coefficient standard errors shown in parentheses. The diagnostic tests are of the form $\mathsf{F}_{\mathsf{j}}(k, T-l)$, which denotes an approximate F-test against the alternative hypothesis j for: $k$th-order serial correlation ($\mathsf{F}_{\mathsf{ar}}$: see Godfrey, 1978), $k$th-order autoregressive conditional heteroskedasticity ($\mathsf{F}_{\mathsf{arch}}$: see Engle, 1982), heteroskedasticity ($\mathsf{F}_{\mathsf{het}}$: see White, 1980a); the RESET test ($\mathsf{F}_{\mathsf{reset}}$: see Ramsey, 1969); parameter constancy ($\mathsf{F}_{\mathsf{Chow}}$ (see Chow, 1960) over $k$ periods; and a chi-square test for normality ($\chi^2_{\mathsf{nd}}(2)$ (see Doornik and Hansen, 2008). No misspecification test rejects.

The result in (1.40) is to be contrasted with the equation reported earlier, which had a similar equilibrium correction term based on Johansen (1988):

$$c_2 = e_f + 7.88 - 0.4e + 0.4\left(p_f - p\right), \tag{1.41}$$

leading to (D3133 and D4446 are dummies with the value unity over the periods 1931–33 and 1944–46 respectively):

$$\begin{aligned}
\Delta e_{f,t} = {} & \underset{(0.04)}{0.27}\, s_{t-1} - \underset{(0.02)}{0.34}\, c_{2,t-1} - \underset{(0.004)}{0.019} + \underset{(0.05)}{0.24}\, \Delta s_t \\
& + \underset{(0.05)}{0.53}\, \Delta e_t - \underset{(0.04)}{0.46}\, \Delta(p_f - p)_t - \underset{(0.01)}{0.12}\, \mathrm{D3133} + \underset{(0.010)}{0.038}\, \mathrm{D4446}
\end{aligned}$$

$$\mathsf{R}^2 = 0.936 \quad \mathsf{F}_{\mathsf{M}}(7,51) = 107.2^{**} \quad \widehat{\sigma} = 0.0098 \quad \mathsf{F}_{\mathsf{ar}}(2,49) = 0.18$$

$$\chi^2(2) = 0.21 \quad \mathsf{F}_{\mathsf{arch}}(1,49) = 0.59 \quad \mathsf{F}_{\mathsf{reset}}(1,50) = 0.26 \quad \mathsf{F}_{\mathsf{het}}(13,37) = 0.47. \tag{1.42}$$

Thus, six additional outliers have been detected in (1.40), whereas none of the components of D4446 was found, nor was $\mathrm{I}_{33}$: neither dummy is remotely significant if added to (1.40). Consistent with that result, when (1.42) and (1.40) are denoted models 1 and 2 on encompassing tests, $\mathsf{F}_{\mathsf{Enc}_{1,2}}(10,41) = 4.83^{**}$ and $\mathsf{F}_{\mathsf{Enc}_{2,1}}(5,41) = 2.18$, so (1.42) is encompassed by (1.40) but not vice versa. Nevertheless, both models are rejected against the other on Cox (1961) and Ericsson (1983) non-nested tests with $\widehat{\sigma}_J = 0.0074$.

Another recent development that can be implemented based on impulse saturation is to test for the super exogeneity of the parameters of the conditional model in response to changes in the LDGPs of the two main conditioning variables, $\Delta e_t$ and $\Delta(p_f - p)_t$ (see section 1.4.5 and Hendry and Santos, 2009). The latter's equation revealed no significant breaks, but commencing from one lag of $\Delta e$, $\Delta(p_f - p)$, $\Delta s$ and $\Delta a$, the former produced:

$$\begin{aligned}
\Delta e_t = {} & \underset{(0.003)}{0.016} + \underset{(0.081)}{0.256}\, \Delta e_{t-1} - \underset{(0.083)}{0.302}\, \Delta(p_f - p)_{t-1} - \underset{(0.02)}{0.11}\, \mathrm{I}_{31} \\
& - \underset{(0.02)}{0.19}\, \mathrm{I}_{32} + \underset{(0.02)}{0.10}\, \mathrm{I}_{34}
\end{aligned}$$

$$+ \underset{(0.02)}{0.06} \, \mathrm{I}_{35} + \underset{(0.02)}{0.07} \, \mathrm{I}_{36} - \underset{(0.02)}{0.08} \, \mathrm{I}_{38} + \underset{(0.02)}{0.07} \, \mathrm{I}_{41} + \underset{(0.02)}{0.08} \, \mathrm{I}_{43}$$

$$+ \underset{(0.02)}{0.10} \, \mathrm{I}_{46} - \underset{(0.02)}{0.06} \, \mathrm{I}_{80}$$

$$\mathsf{R}^2 = 0.86 \;\; \mathsf{F_M}(12, 46) = 23.9^{**} \;\; \widehat{\sigma} = 0.019 \;\; \mathsf{F_{ar}}(2, 44) = 0.14$$

$$\chi^2(2) = 1.27 \;\; \mathsf{F_{arch}}(1, 44) = 4.42^{*} \;\; \mathsf{F_{reset}}(1, 45) = 0.01 \;\; \mathsf{F_{het}}(14, 31) = 0.69. \quad (1.43)$$

Almost all the inter-war and war years are revealed as discrepant, with four impulses in common with (1.40). Adding the 6 additional impulses found in (1.43) to (1.40) and testing their significance yields $\mathsf{F_{SExog}}(6, 40) = 1.14$, not rejecting. Nevertheless, that there are four impulses in common is strongly against the exogeneity of $\Delta e_t$ in (1.40), especially as their signs all match, and even the magnitudes are not too far from 0.67 times those in (1.43). It is not surprising that major shifts in total expenditure are associated with shifts in expenditure on a subcomponent, but since $\Delta e_t$ is included in (1.40), the conclusion must be that agents altered their decision rules more than that effect. Since a food program was in place for several of the common impulses and rationing for the other, additional shifts do not necessarily invalidate the economics behind the equation, so the overall outcome is inconclusive.

An alternative check on the commonality of the inter-war and post-war periods is to use the former to predict the latter, given the actual outcomes for the regressors. We have implicitly done so via impulse saturation, which revealed only one post-war outlier in 1970. The F-test of constancy, $\mathsf{F_{Chow}}(37, 10) = 2.02$, does not reject. Figure 1.7 shows the outcomes: panel (a) reports the fitted and actual values till 1952 and the predicted thereafter, with the full-sample fit shown immediately below in panel (c), the residuals and forecast errors in panel (b), and one-step 95% forecast intervals in panel (d). The outlier in 1970 is obvious, and otherwise there is little difference between the sub-sample and full-sample fit. Such constancy in the face of changing data behavior supports both the specification in (1.40) and the use of the whole sample to estimate and evaluate these models.

### 1.7.1 An update

> Everything was happening so oddly that she didn't feel a bit surprised. (Lewis Carroll, 1899)

The obvious extension is to update the data, and test the model on the extended information. Unfortunately, Applied Econometrics is never that easy: the data have been extensively revised. It came as a surprise even to an experienced empirical modeler that data back to 1929 could differ so much between a 1989-based set (denoted by a subscript $_0$ in the graphs) and a 2008 update when extending the data to 2000 (denoted $_1$), but Figures 1.8 (data) and 1.9 (deviations) show the extent of the revisions. Both food and total real expenditure have changed, the latter by up to 15%, and savings have shifted by up to 5%, whereas the relative price of

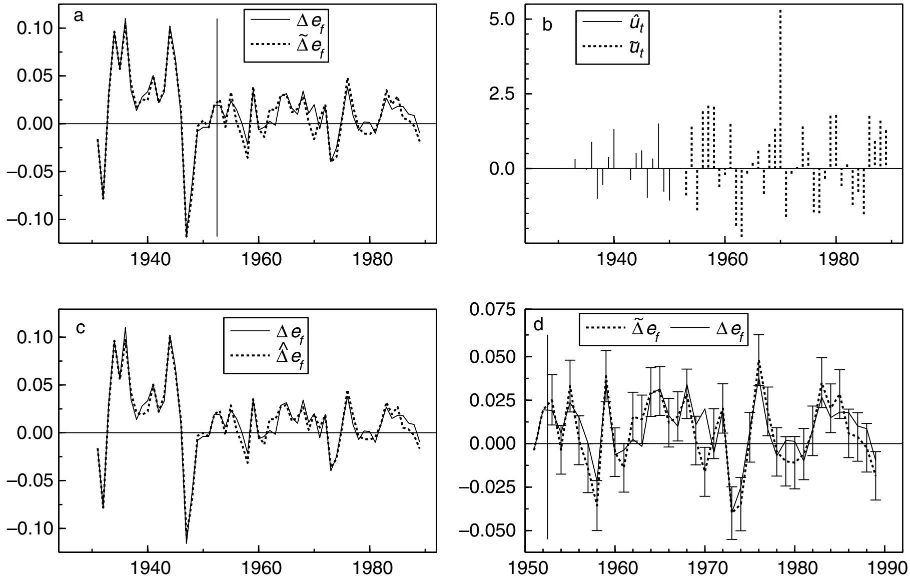


*Figure 1.7* Fitted and actual values, residuals and forecasts for $\Delta e_{f,t}$

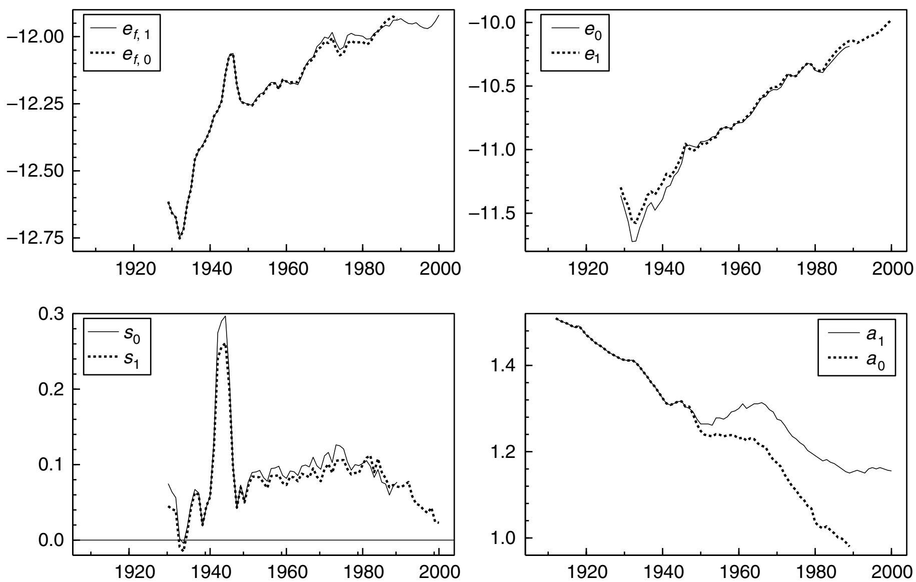


*Figure 1.8* Revised data on food expenditure time series

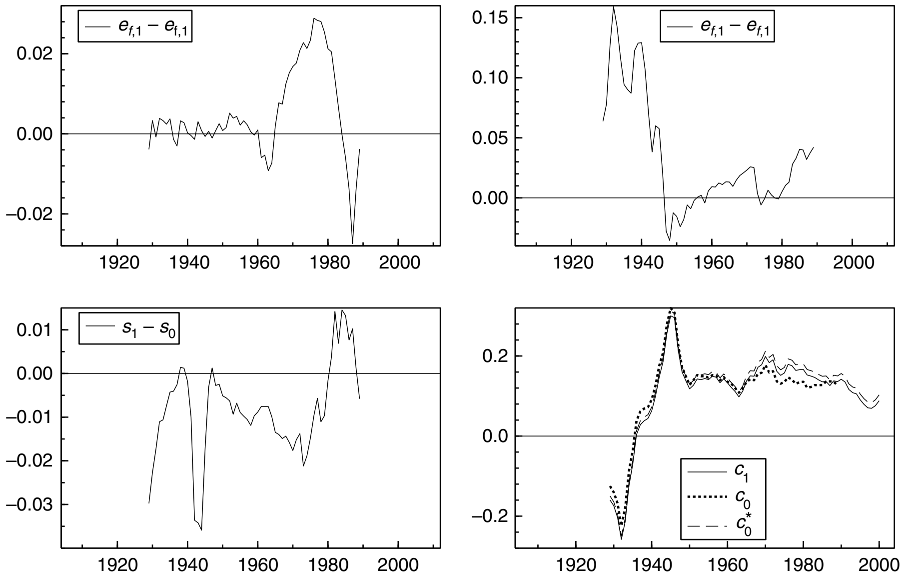


*Figure 1.9* Deviations between old and revised data on food expenditure time series

food is unaltered – yet family size is unrecognizably different. The impacts on the equilibrium-correction terms, $c_0$ in (1.39), that calculated for the revised data $c_0^*$, and $c_1$ in (1.45) below, are also shown (see Hendry, 1994; Cook, 2008, on possible approaches for cross-data-vintage encompassing).

First, enforcing the identical specification to (1.40) but on the revised data over 1930–89, testing on the 11 new years led to:

$$\begin{aligned}\Delta e_{f,t} = {} & \underset{(0.04)}{0.34}\, s_{t-1} - \underset{(0.02)}{0.27}\, c_{0,t-1} + \underset{(0.06)}{0.57}\, \Delta e_t + \underset{(0.05)}{0.09}\, \Delta e_{t-1} \\ & - \underset{(0.04)}{0.40}\, \Delta(p_f - p)_t - \underset{(0.01)}{0.10}\, \mathrm{I}_{31} - \underset{(0.01)}{0.12}\, \mathrm{I}_{32} + \underset{(0.01)}{0.04}\, \mathrm{I}_{34} \\ & + \underset{(0.01)}{0.02}\, \mathrm{I}_{41} + \underset{(0.01)}{0.05}\, \mathrm{I}_{42} + \underset{(0.01)}{0.03}\, \mathrm{I}_{51} + \underset{(0.01)}{0.02}\, \mathrm{I}_{52} + \underset{(0.01)}{0.04}\, \mathrm{I}_{70}\end{aligned}$$

$$\begin{aligned}&\left(\mathsf{R}^*\right)^2 = 0.93 \;\; \mathsf{F_M}(13,45) = 94.9^{**} \;\; \widehat{\sigma} = 0.011 \;\; \mathsf{F_{ar}}(2,44) = 5.74^{**} \\ &\chi^2(2) = 2.40 \;\; \mathsf{F_{arch}}(1,44) = 2.79 \;\; \mathsf{F_{reset}}(1,45) = 0.01 \\ &\mathsf{F_{het}}(18,27) = 0.82 \;\; \mathsf{F_{Chow}}(11,46) = 1.42. \qquad (1.44)\end{aligned}$$

The revisions have altered the coefficients to some extent, the fit is poorer and there is significant residual autocorrelation, but (without "correcting" the standard errors for that problem), the Chow test does not reject, although as Figure 1.10 reveals, the forecast errors are clearly autocorrelated.

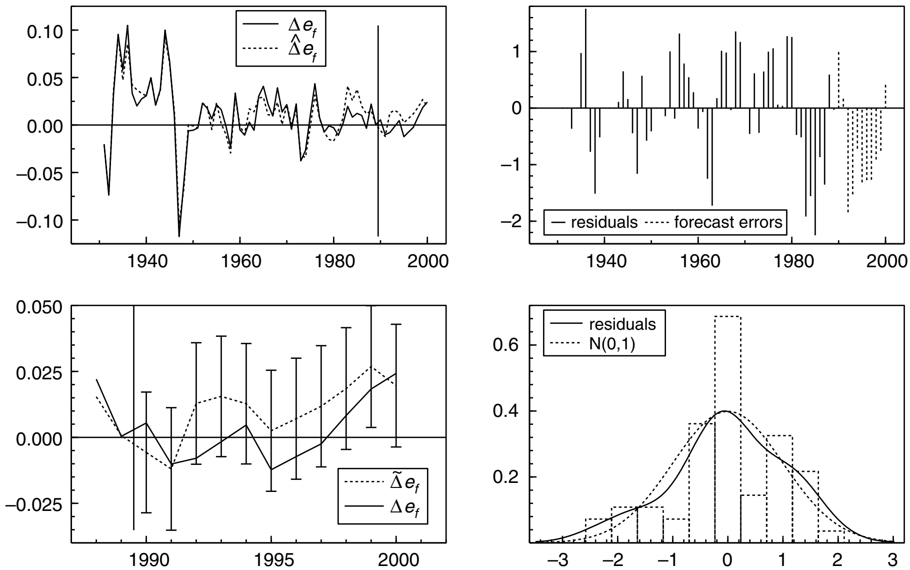


*Figure 1.10* Old model revised data fitted and actual values, residuals and forecasts for $\Delta e_{f,t}$

Next, automatic remodeling at 1% on the revised data up to 1989 (with impulse saturation to remove the outliers) led to:

$$c_1 = e_f + 8.49 - 0.35e + 0.21(p_f - p), \tag{1.45}$$

with a much simpler final equation being selected:

$$\Delta e_{f,t} = \underset{(0.032)}{0.35}\ s_t - \underset{(0.02)}{0.25}\ c_{1,t-1} + \underset{(0.05)}{0.65}\ \Delta e_t - \underset{(0.04)}{0.29}\ \Delta(p_f - p)_t$$

$$- \underset{(0.01)}{0.05}\ \mathrm{I}_{30} - \underset{(0.01)}{0.08}\ \mathrm{I}_{31} - \underset{(0.01)}{0.08}\ \mathrm{I}_{32} + \underset{(0.01)}{0.03}\ \mathrm{I}_{70}$$

$$\left(\mathsf{R}^*\right)^2 = 0.90 \quad \mathsf{F}_{\mathsf{M}}(8,51) = 60.4^{**} \quad \widehat{\sigma} = 0.012 \quad \mathsf{F}_{\mathsf{ar}}(2,50) = 0.65$$

$$\chi^2(2) = 2.09 \quad \mathsf{F}_{\mathsf{arch}}(1,50) = 0.13 \quad \mathsf{F}_{\mathsf{reset}}(1,51) = 0.11$$

$$\mathsf{F}_{\mathsf{het}}(18,33) = 0.89 \quad \mathsf{F}_{\mathsf{Chow}}(11,52) = 0.68\ (1990 - 2000). \tag{1.46}$$

Nevertheless, despite the revisions, the model in (1.46) has many features in common with both its predecessors, and is constant over the next 11 years as Figure 1.11 reports, and $\mathsf{F}_{\mathsf{Chow}}(11,52)$ confirms. The short-run elasticities still exceed their long-run counterparts, but by less than previously.

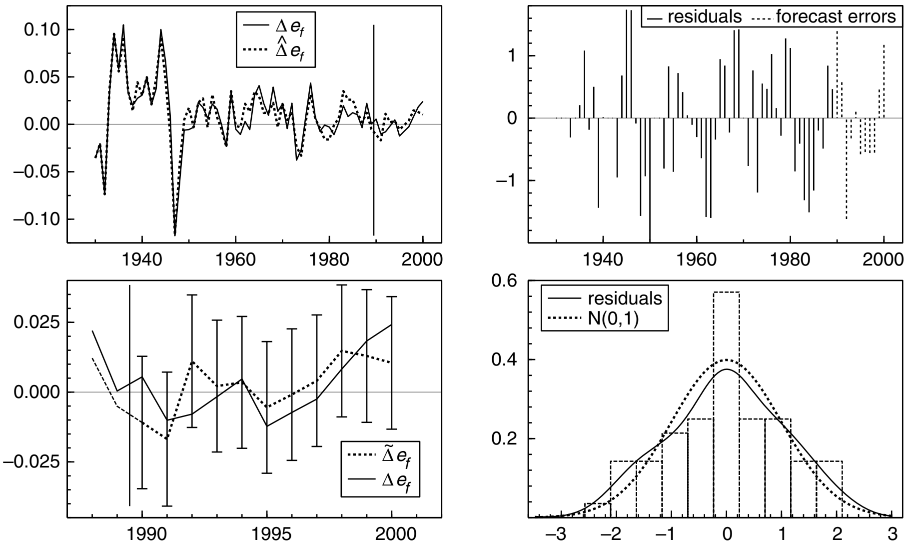


*Figure 1.11* New model on revised data fitted and actual values, residuals and forecasts for $\Delta e_{f,t}$

## 1.8 Automatic modeling of a VAR$_4$ (25)

> "The Eighth Square at last!" she cried as she bounded across ... "Oh, how glad I am to get here! And what is this on my head?" she exclaimed ... It was a golden crown. (Quote from Alice in Lewis Carroll, 1899)

To illustrate that automatic modeling is not restricted to single equations (see, e.g., Krolzig, 2003), we now model the four variables in section 1.4.1.1, namely industrial output per capita, $y_{c,t}$, numbers of bankruptcies, $b_t$, and patents, $p_t$, and real equity prices (deflated by a cost of living index), $e_t$, using a VAR with 25 lags, augmented by impulse saturation over the common sample $T$ =1757–1989 at $\alpha = 0.0025$ (so on average about one variable will be retained by chance as there are 337 candidates in the initial general model). The marginal critical t-ratio is about 3.1, and only about 3 regressors (other than impulses) were near or below that in the four finally-selected models. Most diagnostic tests were insignificant in those final models (but not computable at the start). The entire exercise took under two hours, including this write-up: technical progress in undertaking empirical econometrics is huge, as such an analysis would have been simply impossible (conceptually and practically) when I first started empirical modeling in 1967.

$$\Delta y_{c,t} = -\underset{(0.031)}{0.128}\, y_{c,t-1} + \underset{(0.037)}{0.143}\, y_{c,t-5} - \underset{(0.037)}{0.152}\, y_{c,t-19} + \underset{(0.034)}{0.135}\, y_{c,t-23}$$

$$+ \underset{(0.004)}{0.016}\, p_t - \underset{(0.005)}{0.016}\, b_{t-3} + \underset{(0.013)}{0.081}\, e_t - \underset{(0.013)}{0.084}\, e_{t-2}$$

$$+ \underset{(0.038)}{0.129} \mathrm{I}_{1827} - \underset{(0.039)}{0.206} \mathrm{I}_{1921} + \underset{(0.037)}{0.119} \mathrm{I}_{1927}$$

$$\widehat{\sigma} = 0.0368 \;\; \mathsf{F}_{\mathsf{AR}1-2}(2,220) = 1.08 \;\; \mathsf{F}_{\mathsf{ARCH}1}(1,220) = 0.14$$

$$\chi^2_{\mathsf{nd}}(2) = 1.09 \;\; \mathsf{F}_{\mathsf{Het}}(19,202) = 2.24^{**} \;\; \mathsf{F}_{\mathsf{RESET}}(1,221) = 5.97^{*}$$

$$\begin{aligned}
p_t = {} & \underset{(0.65)}{3.32} + \underset{(0.03)}{0.87} p_{t-1} - \underset{(0.22)}{0.46} y_{c,t-4} + \underset{(0.22)}{0.68} y_{c,t-5} \\
& - \underset{(0.04)}{0.19} b_{t-11} + \underset{(0.04)}{0.18} b_{t-12} \\
& - \underset{(0.05)}{0.16} e_{t-17} + \underset{(0.16)}{0.67} \mathrm{I}_{1757} - \underset{(0.15)}{0.51} \mathrm{I}_{1759} - \underset{(0.15)}{0.57} \mathrm{I}_{1761} + \underset{(0.15)}{0.58} \mathrm{I}_{1766} \\
& - \underset{(0.15)}{0.43} \mathrm{I}_{1771} - \underset{(0.15)}{0.63} \mathrm{I}_{1775} + \underset{(0.15)}{0.43} \mathrm{I}_{1783} - \underset{(0.15)}{0.65} \mathrm{I}_{1793} - \underset{(0.15)}{0.46} \mathrm{I}_{1826} \\
& - \underset{(0.15)}{0.49} \mathrm{I}_{1884} - \underset{(0.15)}{0.45} \mathrm{I}_{1940} - \underset{(0.15)}{0.46} \mathrm{I}_{1942} - \underset{(0.15)}{0.52} \mathrm{I}_{1984} + \underset{(0.15)}{0.60} \mathrm{I}_{1985}
\end{aligned}$$

$$\widehat{\sigma} = 0.148 \;\; \mathsf{F}_{\mathsf{AR}1-2}(2,210) = 1.46 \;\; \mathsf{F}_{\mathsf{ARCH}1}(1,210) = 0.57$$

$$\chi^2_{\mathsf{nd}}(2) = 4.88 \;\; \mathsf{F}_{\mathsf{Het}}(26,185) = 1.41 \;\; \mathsf{F}_{\mathsf{RESET}}(1,211) = 0.19$$

$$\begin{aligned}
b_t = {} & \underset{(0.05)}{1.10} b_{t-1} - \underset{(0.05)}{0.29} b_{t-2} - \underset{(0.10)}{0.41} y_{c,t-2} - \underset{(0.18)}{0.41} y_{c,t-17} + \underset{(0.16)}{0.80} y_{c,t-23} \\
& + \underset{(0.06)}{0.42} p_{t-1} - \underset{(0.06)}{0.23} p_{t-2} - \underset{(0.09)}{0.46} e_t + \underset{(0.10)}{0.55} e_{t-1} - \underset{(0.05)}{0.23} e_{t-5} \\
& + \underset{(0.21)}{0.64} \mathrm{I}_{1757} + \underset{(0.19)}{0.50} \mathrm{I}_{1766} - \underset{(0.18)}{0.56} \mathrm{I}_{1822} - \underset{(0.19)}{0.65} \mathrm{I}_{1838} + \underset{(0.18)}{0.83} \mathrm{I}_{1884}
\end{aligned}$$

$$\widehat{\sigma} = 0.18 \;\; \mathsf{F}_{\mathsf{AR}1-2}(2,216) = 1.60 \;\; \mathsf{F}_{\mathsf{ARCH}1}(1,216) = 2.48$$

$$\chi^2_{\mathsf{nd}}(2) = 5.25 \;\; \mathsf{F}_{\mathsf{Het}}(25,192) = 1.64^{*} \;\; \mathsf{F}_{\mathsf{RESET}}(1,217) = 1.59$$

$$\begin{aligned}
e_t = {} & \underset{(0.06)}{1.12} e_{t-1} - \underset{(0.06)}{0.17} e_{t-2} + \underset{(0.16)}{0.69} y_{c,t} - \underset{(0.16)}{0.69} y_{c,t-1} \\
& - \underset{(0.03)}{0.10} b_t + \underset{(0.03)}{0.12} b_{t-1} + \underset{(0.10)}{0.35} \mathrm{I}_{1802} + \underset{(0.11)}{0.31} \mathrm{I}_{1922} \\
& + \underset{(0.10)}{0.31} \mathrm{I}_{1959} - \underset{(0.10)}{0.31} \mathrm{I}_{1973} - \underset{(0.11)}{0.58} \mathrm{I}_{1974}
\end{aligned}$$

$$\widehat{\sigma} = 0.10 \;\; \mathsf{F}_{\mathsf{AR1-2}}(2, 220) = 2.59 \;\; \mathsf{F}_{\mathsf{ARCH1}}(1, 220) = 0.01$$
$$\chi^2_{\mathsf{nd}}(2) = 3.72 \;\; \mathsf{F}_{\mathsf{Het}}(17, 204) = 1.21 \;\; \mathsf{F}_{\mathsf{RESET}}(1, 221) = 3.72.$$

Most of the effects found make economic sense in the context of the limited information set used here as an illustration. In reverse order, real equity prices are near a random walk, but respond positively to changes in output, and negatively to changes in bankruptcies. In turn, bankruptcies fall with increased output or equity prices, but rise with patent grants. Neither equation has many outliers, whereas the patents equation does, especially in the eighteenth century. Patents fall initially as output, equity prices, bankruptcies rise, but adjust back later. Finally, changes in output respond positively to patents and changes in equity prices, but negatively to bankruptcies.

A substantive exercise would involve additional variables like interest rates and human and physical capital; would check whether bankruptcies and patents should also be per capita; and investigate cointegration reductions. Are the long lags 'spurious'? The general historical record suggests that major innovations are both creative and destructive of output, the former by the enlargement of the production frontier, and the latter through the negative impact on those already engaged in the occupations concerned (spinners, weavers, etc., initially; clerks and secretaries in more modern times), so a "generation" is required for the new state to dominate – that motivated the original choice of 25 lags. Innovations take time to develop and be adopted; and the seeds for bankruptcy are often sown well before the reaping, even if the span is not quite "clogs to clogs in three generations." Notably, the equation for equity prices still has short lags despite the "opportunity" to find other correlations.

## 1.9 Conclusion

> Ever drifting down the stream –
> Lingering in the golden gleam –
> Life, what is it but a dream? (Lewis Carroll, 1899)

"Applied Econometrics" has a vast range of empirical issues to investigate: the very non-stationarity of economies keeps creating new topics for analysis. However, so long as "Applied Econometrics" is just a calibration of extant economic theory, it will never make much of an independent contribution: in that sense, one must agree with Summers (1991) but for completely opposite reasons. Much of the observed data variability in economics is due to features that are absent from most economic theories, but which empirical models have to tackle. *Ceteris paribus* conditions can sometimes be justified for theoretical reasoning, but do not provide a viable basis for empirical modeling: only a "minor influence" theorem, which must be established empirically, will suffice.

This implication is not a tract for mindless modeling of data in the absence of economic analysis, but instead suggests formulating more general initial models that embed the available economic theory as a special case, consistent with our knowledge of the institutional framework, historical record, and the data properties. Once a congruent encompassing general model is established, an automatic model selection approach based on general-to-simple principles could help bring objectivity and credibility to empirical econometric modeling.

Economic observations are far from perfect, being subject to revision, and even to conceptual changes, with important variables unobserved, and available proxies of unknown quality. Theory constructs (such as "consumption," or "user cost of capital") and their measured counterparts (consumers' expenditure or after-tax real interest rates adjusted for depreciation) can differ markedly, especially after aggregation. Thus, a "pure" data-based approach can lack substance.

Economics has delivered a range of invaluable insights into individual decision taking, market functioning, and system-wide economies, with a vast body of theory, which has made rapid technical and intellectual progress – and will continue to do so. Applied econometrics cannot be conducted without an economic theoretic framework to guide its endeavors and and help interpret its findings. Nevertheless, since economic theory is not complete, correct, and immutable, and never will be, one also cannot justify an insistence on deriving empirical models from theory alone. That paradigm encourages covert data mining, so the credibility of existing evidence is unclear.

Data "mining" does not have pernicious properties when using a structured approach, using appropriate significance levels that decline with both the number of candidate variables and sample size: at 1% significance, one irrelevant variable in 100 will be significant by chance, at the cost of raising the selection t-ratio from around $\pm 2.0$ to $\pm 2.7$. Parsimony is not a justification for arbitrarily excluding many potentially relevant contenders, not even when doing so to avoid more initial variables than observations. While it is essential that the final model is much smaller than the sample size, that does not preclude starting general and making the maximum use of our best available theory and econometrics to guide our empirical endeavors and then interpret their outcomes. Thus, Frisch (1933) remains our best advice: "mutual penetration', which entails using economic analysis to guide an applied study, but letting the empirical evidence play a real role.

## Acknowledgments

Financial support from the ESRC under Research Grant RES-062-23-0061 is gratefully acknowledged. I am indebted to Gunnar Baardsen, Julia Campos, Jennifer Castle, Guillaume Chevillon, Jurgen Doornik, Neil Ericsson, Katarina Juselius, Søren Johansen, Bobby Mariano, Jaime Marquez, Terry Mills, Mary Morgan, Bent Nielsen, Ragnar Nymoen, Kerry Patterson, Duo Qin, James Reade, Aris Spanos and Pravin Trivedi for helpful comments.

## Note

1. Atkinson (2008) notes Robbins' apparent dismissal of Richard Stone (1951) as "not economics."

## References

Aghion, P., R. Frydman, J. Stiglitz and M. Woodford (eds.) (2002) *Knowledge, Information and Expectations in Modern Macroeconomics*. Princeton: Princeton University Press.

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817–58.

Anselin, L. (2006) Spatial econometrics. In T.C. Mills and K.D. Patterson (eds.), *Palgrave Handbook of Econometrics, Volume I: Econometric Theory*, pp. 901–69. Basingstoke: Palgrave Macmillan.

Atkinson, A.B. (2005) *Measurement of Government Output and Productivity for the National Accounts: The Atkinson Review Final Report*. London: Palgrave Macmillan.

Atkinson, A.B. (2008) Economic data and the distribution of income. Stone lectures, 2008, Department of Economics, University of Oxford.

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

Baillie, R.T. (1996) Long memory processes and fractional integration in econometrics. *Journal of Econometrics* **73**, 5–59.

Baltagi, B.H. (2006) Panel data models. In T.C. Mills and K.D. Patterson (eds.), *Palgrave Handbook of Econometrics, Volume I: Econometric Theory*, pp. 633–61. Basingstoke: Palgrave Macmillan.

Banerjee, A. and D.F. Hendry (1992) Testing integration and cointegration: an overview. *Oxford Bulletin of Economics and Statistics* **54**, 225–55.

Barndorff-Nielsen, O.E. and N. Shephard (2007) *Financial Volatility in Continuous Time*. Cambridge: Cambridge University Press. Forthcoming.

Bjerkholt, O. (2005) Frisch's econometric laboratory and the rise of Trygve Haavelmo's probability approach. *Econometric Theory* **21**, 491–533.

Bjerkholt, O. (2007) Writing "the probability approach" with nowhere to go: Haavelmo in the United States, 1939–1944. *Econometric Theory* **23**, 775–837.

Blundell, R. and T.M. Stoker (2005) Heterogeneity and aggregation. *Journal of Economic Literature* **43**, 347–91.

Bontemps, C. and G.E. Mizon (2003) Congruence and encompassing. In B.P. Stigum (ed.), *Econometrics and the Philosophy of Economics*, pp. 354–78. Princeton: Princeton University Press.

Boswijk, H.P. (1992) *Cointegration, Identification and Exogeneity*, Volume 37 of *Tinbergen Institute Research Series*. Amsterdam: Thesis Publishers.

Boumans, M.A. (2005) Measurement in economic systems. *Measurement* **38**, 275–84.

Boumans, M.A. (ed.) (2007) *Measurement in Economics: A Handbook*. Amsterdam: Elsevier.

Box, G.E.P. and G.M. Jenkins, (1976) *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day (first published 1970).

Burns, A.F. and W.C. Mitchell (1946) *Measuring Business Cycles*. New York: NBER.

Caldwell, B.J. (ed.) (1993) *The Philosophy and Methodology of Economics, Volume II*. Aldershot: Edward Elgar.

Camerer, C.F. (2007) Neuroeconomics: using neuroscience to make economic predictions. *Economic Journal* **117**, C26–42.

Cameron, A.C. and P.K. Trivedi (1998) *The Analysis of Count Data*. Cambridge: Cambridge University Press.

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

Carroll, L. (1899) *Through the Looking-Glass and What Alice Found There*. London: Macmillan & Co.

Cartwright, N. (1983) *How the Laws of Physics Lie*. Oxford: Clarendon Press.

Cartwright, N. (2002) In favor of laws that are not ceteris paribus after all. *Erkenntnis* **57**, 425–39.

Castle, J.L. (2005) Evaluating PcGets and RETINA as automatic model selection algorithms. *Oxford Bulletin of Economics and Statistics* **67**, 837–80.

Castle, J.L., J.A. Doornik, D.F. Hendry and R. Nymoen (2008) Testing the invariance of expectations models of inflation. Working Paper, Economics Department, Oxford University.

Castle, J.L. and D.F. Hendry (2005) Extending the boundaries of automatic Gets to non-linear models. Mimeo, Economics Department, Oxford University.

Chatfield, C. (1995) Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society, A* **158**, 419–66. With discussion.

Choi, I. (2006) Non-stationary panels. In T.C. Mills and K.D. Patterson (eds.), *Palgrave Handbook of Econometrics, Volume I: Econometric Theory,* pp. 511–39. Basingstoke: Palgrave Macmillan.

Chow, G.C. (1960) Tests of equality between sets of coefficients in two linear regressions. *Econometrica* **28**, 591–605.

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.

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

Clements, M.P. and D.F. Hendry (1999) *Forecasting Non-stationary Economic Time Series*. Cambridge, Mass.: MIT Press.

Clements, M.P. and D.F. Hendry (2001) An historical perspective on forecast errors. *National Institute Economic Review* **177**, 100–12.

Clements, M.P. and D.F. Hendry (eds.) (2002a) *A Companion to Economic Forecasting*. Oxford: Blackwell.

Clements, M.P. and D.F. Hendry (2002b) Explaining forecast failure in macroeconomics. In M.P. Clements and D.F. Hendry (eds.), *A Companion to Economic Forecasting*, pp. 539–71. Oxford: Blackwell.

Clements, M.P. and D.F. Hendry (2005) Guest Editors' introduction: Information in economic forecasting. *Oxford Bulletin of Economics and Statistics* **67**, 713–53.

Clements, M.P. and D.F. Hendry (2006) Forecasting with breaks in data processes. In G. Elliott, C.W.J. Granger and A. Timmermann (eds.), *Handbook of Ecometrics on Forecasting*, pp. 605–57. Amsterdam: Elsevier.

Cook, S. (2008) Cross-data vintage encompassing. *Oxford Bulletin of Economics and Statistics*. Forthcoming.

Cox, D.R. (1961) Tests of separate families of hypotheses. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1*, pp. 105–23. Berkeley: University of California Press.

Crafts, N.F.R. (1997) Endogenous growth: lessons for and from economic history. In D.M. Kreps and K.F. Wallis (eds.), *Advances in Economics and Econometrics: Theory and Applications. Seventh World Congress, Volume 2*. Cambridge: Cambridge University Press.

Crafts, N.F.R. and T.C. Mills (1994) Trends in real wages in Britain, 1750–1913. *Explorations in Economic History* **31**, 176–94.

Darnell, A. (ed.) (1994) *The History of Econometrics*. Aldershot: Edward Elgar.

Demiralp, S. and K.D. Hoover (2003) Searching for the causal structure of a vector autoregression. *Oxford Bulletin of Economics and Statistics* **65**, 745–67.

Dolado, J.J. (1992) A note on weak exogeneity in VAR cointegrated systems. *Economics Letters* **38**, 139–43.

Doob, J.L. (1953) *Stochastic Processes* (1990 edition). New York: John Wiley Classics Library.

Doornik, J.A. (2006) *Object-Oriented Matrix Programming using Ox* (fourth edition). London: Timberlake Consultants Press.

Doornik, J.A. (2007a) Autometrics. Working Paper, Economics Department, University of Oxford.

Doornik, J.A. (2007b) Econometric modelling when there are more variables than observations. Working paper, Economics Department, University of Oxford.

Doornik, J.A. and H. Hansen (2008) A practical test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*. Forthcoming.

Elliott, G., C.W.J. Granger and A. Timmermann (eds.) (2006) *Handbook of Econometrics on Forecasting*. Amsterdam: Elsevier.

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

Engle, R.F. and D.F. Hendry (1993) Testing super exogeneity and invariance in regression models. *Journal of Econometrics* **56**, 119–39.

Engle, R.F., D.F. Hendry and J.-F. Richard (1983) Exogeneity. *Econometrica* **51**, 277–304.

Epstein, R.J. (1987) *A History of Econometrics*. Amsterdam: North-Holland.

Ericsson, N.R. (1983) Asymptotic properties of instrumental variables statistics for testing non-nested hypotheses. *Review of Economic Studies* **50**, 287–303.

Ericsson, N.R. (1992) Cointegration, exogeneity and policy analysis: an overview. *Journal of Policy Modeling* **14**, 251–80.

Ericsson, N.R. (2007) *Econometric Modeling*. Oxford: Oxford University Press. Forthcoming.

Ericsson, N.R. and D.F. Hendry (1999) Encompassing and rational expectations: how sequential corroboration can imply refutation. *Empirical Economics* **24**, 1–21.

Ericsson, N.R. and J.G. MacKinnon (2002) Distributions of error correction tests for cointegration. *Econometrics Journal* **5**, 285–318.

Ermini, L. and D.F. Hendry (2008) Log income versus linear income: an application of the encompassing principle. *Oxford Bulletin of Economics and Statistics*. Forthcoming.

Evans, G.W. and S. Honkapohja (2001) *Learning and Expectations in Macroeconomics*. Princeton: Princeton University Press.

Farebrother, R.W. (2006) Early explorations in econometrics. In T.C. Mills and K.D. Patterson (eds.), *Palgrave Handbook of Ecometrics, Volume I: Econometric Theory,* pp. 88–116. Basingstoke: Palgrave Macmillan.

Faust, J. and C.H. Whiteman (1997) General-to-specific procedures for fitting a data-admissible, theory-inspired, congruent, parsimonious, encompassing, weakly-exogenous, identified, structural model of the DGP: a translation and critique. *Carnegie–Rochester Conference Series on Public Policy* **47**, 121–61.

Favero, C. and D.F. Hendry (1992) Testing the Lucas critique: a review. *Econometric Reviews* **11**, 265–306.

Fehr, E. and A. Falk (2002) Psychological foundations of incentives. *European Economic Review* **46**, 687–724.

Fehr, E., U. Fischbacher and M. Kosfeld (2005) Neuroeconomic foundation of trust and social preferences. Dp. 5127, CEPR, London.

Feige, E.L. and D.K. Pearce, (1976) Economically rational expectations. *Journal of Political Economy* **84**, 499–522.

Feinstein, C.H. (1972) *National Income, Expenditure and Output of the United Kingdom, 1855–1965*. Cambridge: Cambridge University Press.

Fischer, A.M. (1989) Policy regime changes and monetary expectations: testing for super exogeneity. *Journal of Monetary Economics* **24**, 423–36.

Fisher, F.M. (1966) *The Identification Problem in Econometrics*. New York: McGraw-Hill.

Florens, J.-P., M. Mouchart and J.-M. Rolin (1990) *Elements of Bayesian Statistics*. New York: Marcel Dekker.

Foxwell, H.S. (ed.) (1884) *Investigations in Currency and Finance*. London: Macmillan.

Friedman, M. (1957) *A Theory of the Consumption Function*. Princeton: Princeton University Press.

Frisch, R. (1933) Editorial. *Econometrica* **1**, 1–4.

Frisch, R. (1938) Statistical versus theoretical relations in economic macrodynamics. Mimeograph dated July 17, 1938, League of Nations Memorandum. Reprinted in D.F. Hendry and M.S. Morgan (eds.), *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press, 1995.

Frydman, R. and M.D. Goldberg (eds.) (2007) *Imperfect Knowledge Economics: Exchange Rates and Risk*. Princeton: Princeton University Press.

Galí, J., M. Gertler and J.D. Lopez-Salido (2001) European inflation dynamics. *European Economic Review* **45**, 1237–70.

Geweke, J.F. and S. Porter-Hudak (1983) The estimation and application of long memory time series models. *Journal of Time Series Analysis* **4**, 221–38.

Gilbert, C.L. and D. Qin (2006) The first fifty years of modern econometrics. In T.C. Mills and K.D. Patterson (eds.), *Palgrave Handbook of Econometrics, Volume I: Econometric Theory,* pp. 117–55. Basingstoke: Palgrave Macmillan.

Gilbert, C.L. and D. Qin (2007) Representation in econometrics: a historical perspective. In M.A. Boumans (ed.), *Measurement in Economics: A Handbook,* Ch. 10. Amsterdam: Elsevier.

Godfrey, L.G. (1978) Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica* **46**, 1303–13.

Granger, C.W.J. (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–38.

Granger, C.W.J. (1980) Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics* **14**, 227–38.

Granger, C.W.J. (1981) Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* **16**, 121–30.

Granger, C.W.J. (1987) Implications of aggregation with common factors. *Econometric Theory* **3**, 208–22.

Granger, C.W.J. and R. Joyeux (1980) An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis* **1**, 15–30.

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**, 351–7.

Haavelmo, T. (1989) *Prize Lecture*. Sveriges Riksbank: Prize in Economic Sciences in Memory of Alfred Nobel.

Hald, A. (1990) *A History of Probability and Statistics and Their Applications before 1750*. New York: Wiley.

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

Haldrup, N., H. H. van Dijk and D.F. Hendry (eds.) (2003) Model selection and evaluation. *Oxford Bulletin of Economics and Statistics* **65**.

Hall, R.E. (1978) Stochastic implications of the life cycle-permanent income hypothesis: evidence. *Journal of Political Economy* **86**, 971–87.

Hamouda, O.F. and J.C.R. Rowley (1997) *The Reappraisal of Econometrics*. Aldershot: Edward Elgar.

Harvey, A.C. (1993) *Time Series Models* (second edition; first edition 1981). Hemel Hempstead: Harvester Wheatsheaf.

Heckman, J.J. (1976) The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* **5**, 475–92.

Heckman, J.J. (2000) Causal parameters and policy analysis in economics: a twentieth century retrospective. *Quarterly Journal of Economics* **115**, 45–97.

Hendry, D.F. (1980) Econometrics: alchemy or science? *Economica* **47**, 387–406.

Hendry, D.F. (1987) Econometric methodology: a personal perspective. In T.F. Bewley (ed.), *Advances in Econometrics*, pp. 29–48. Cambridge: Cambridge University Press.

Hendry, D.F. (1988) The encompassing implications of feedback versus feedforward mechanisms in econometrics. *Oxford Economic Papers* **40**, 132–49.

Hendry, D.F. (1994) HUS revisited. *Oxford Review of Economic Policy* **10**, 86–106.
Hendry, D.F. (1995a) *Dynamic Econometrics*. Oxford: Oxford University Press.
Hendry, D.F. (1995b) Econometrics and business cycle empirics. *Economic Journal* **105**, 1622–36.
Hendry, D.F. (1995c) On the interactions of unit roots and exogeneity. *Econometric Reviews* **14**, 383–419.
Hendry, D.F. (1999) An econometric analysis of US food expenditure, 1931–1989. In J.R. Magnus and M.S. Morgan (eds.), *Methodology and Tacit Knowledge: Two Experiments in Econometrics,* pp. 341–61. Chichester: John Wiley and Sons.
Hendry, D.F. (2000a) Epilogue: The success of general-to-specific model selection. In *Econometrics: Alchemy or Science?* (new edition), pp. 467–90. Oxford: Oxford University Press.
Hendry, D.F. (2000b) On detectable and non-detectable structural change. *Structural Change and Economic Dynamics* **11**, 45–65.
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**, 255–75.
Hendry, D.F. (2004) Causality and exogeneity in non-stationary economic time series. In A. Welfe (ed.), *New Directions in Macromodelling*, pp. 21–48. Amsterdam: North-Holland.
Hendry, D.F. (2005) Bridging the gap: linking economics and econometrics. In C. Diebolt and C. Kyrtsou (eds.), *New Trends in Macroeconomics*, pp. 53–77. Berlin: Springer Verlag.
Hendry, D.F. (2006) Robustifying forecasts from equilibrium-correction models. *Journal of Econometrics* **135**, 399–426. Special Issue in Honor of Clive Granger.
Hendry, D.F. and J.A. Doornik (1997) The implications for econometric modelling of forecast failure. *Scottish Journal of Political Economy* **44**, 437–61. Special Issue.
Hendry, D.F. (with J.A. Doornik and B. Nielsen) (2007) *Econometric Model Selection: Arne Ryde Lectures*. Lund University, Sweden.
Hendry, D.F., S. Johansen and C. Santos (2008) Automatic selection of indicators in a fully saturated regression. *Computational Statistics* **33**, 317–35. Erratum, 337–9.
Hendry, D.F. and K. Juselius (2000) Explaining cointegration analysis: Part I. *Energy Journal* **21**, 1–42.
Hendry, D.F. and K. Juselius (2001) Explaining cointegration analysis: Part II. *Energy Journal* **22**, 75–120.
Hendry, D.F. and H.-M. Krolzig (1999) Improving on "Data mining reconsidered" by K.D. Hoover and S.J. Perez. *Econometrics Journal* **2**, 202–19.
Hendry, D.F. and H.-M. Krolzig (2001) *Automatic Econometric Model Selection*. London: Timberlake Consultants Press.
Hendry, D.F. and H.M. Krolzig (2004) We ran one regression. *Oxford Bulletin of Economics and Statistics* **66**, 799–810.
Hendry, D.F. and H.-M. Krolzig (2005) The properties of automatic Gets modelling. *Economic Journal* **115**, C32–61.
Hendry, D.F., E.E. Leamer and D.J. Poirier (1990) A conversation on econometric methodology. *Econometric Theory* **6**, 171–261.
Hendry, D.F., M. Lu and G.E. Mizon (2008) Model identification and non-unique structure. In J.L. Castle and N. Shephard (eds.), *The Methodology and Practice of Econometrics*. Oxford: Oxford University Press. Forthcoming.
Hendry, D.F., M. Marcellino and G.E. Mizon (eds.) (2008) *Encompassing*. *Oxford Bulletin of Economics and Statistics*. Special Issue. Forthcoming.
Hendry, D.F. and M. Massmann (2007) Co-breaking: recent advances and a synopsis of the literature. *Journal of Business and Economic Statistics* **25**, 33–51.

Hendry, D.F. and G.E. Mizon (1999) The pervasiveness of Granger causality in econometrics. In R.F. Engle and H. White (eds.), *Cointegration, Causality and Forecasting*. Oxford: Oxford University Press.

Hendry, D.F. and G.E. Mizon (2000) Reformulating empirical macro-econometric modelling. *Oxford Review of Economic Policy* **16**, 138–59.

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

Hendry, D.F. and B. Nielsen (2007a) *Econometric Modeling: A Likelihood Approach*. Princeton: Princeton University Press.

Hendry, D.F. and B. Nielsen (2007b) Teaching undergraduate econometrics using *OxMetrics*. Mimeo, Economics Department, University of Oxford.

Hendry, D.F. and J.-F. Richard (1982) On the formulation of empirical models in dynamic econometrics. *Journal of Econometrics* **20**, 3–33.

Hendry, D.F. and J.-F. Richard (1989) Recent developments in the theory of encompassing. In B. Cornet and H. Tulkens (eds.), *Contributions to Operations Research and Economics. The XXth Anniversary of CORE*, pp. 393–440. Cambridge, Mass.: MIT Press.

Hendry, D.F. and C. Santos (2009) An automatic test of super exogeneity. In M.W. Watson, T. Bollerslev and J. Russell (eds.), *Volatility and Time Series Econometrics*. Oxford: Oxford University Press. Forthcoming.

Hildenbrand, W. (1994) *Market Demand: Theory and Empirical Evidence*. Princeton: Princeton University Press.

Hildenbrand, W. (1998) How relevant are the specifications of behavioural relations on the micro-level for modelling the time path of population aggregates? *European Economic Review* **42**, 437–58.

Hildreth, C. and J.P. Houck (1968) Some estimators for a linear model with random coefficients. *Journal of the American Statistical Association* **63**, 584–95.

Hooker, R.H. (1901) Correlation of the marriage rate with trade. *Journal of the Royal Statistical Society* **64**, 485–92. Reprinted in D.F. Hendry and M.S. Morgan (eds.), *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press, 1995.

Hoover, K.D. and S.J. Perez (1999) Data mining reconsidered: encompassing and the general-to-specific approach to specification search. *Econometrics Journal* **2**, 167–91.

Hoover, K.D. and S.J. Perez (2004) Truth and robustness in cross-country growth regressions. *Oxford Bulletin of Economics and Statistics* **66**, 765–98.

Hsiao, C. (1983) Identification. In Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, *Volume* 1, Ch. 4. Amsterdam: North-Holland.

Hunter, J. (1992) Cointegrating exogeneity. *Economics Letters* **34**, 33–5.

Jevons, W.S. (1875) The solar period and the price of corn. In H.S. Foxwell (ed.), *Investigations in Currency and Finance*, pp. 194–205. London: Macmillan.

Johansen, S. (1988) Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* **12**, 231–54.

Johansen, S. (1992) Testing weak exogeneity and the order of cointegration in UK money demand. *Journal of Policy Modeling* **14**, 313–34.

Johansen, S. (1995) *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Johansen, S. (2006) Cointegration: an overview. In T.C. Mills and K.D. Patterson (eds.), *Palgrave Handbook of Econometrics, Volume I: Econometric Theory*, pp. 540–77. Basingstoke: Palgrave Macmillan.

Johansen, S. and K. Juselius (1990) Maximum likelihood estimation and inference on co-integration – with application to the demand for money. *Oxford Bulletin of Economics and Statistics* **52**, 169–210.

Johansen, S. and B. Nielsen (2008) An analysis of the indicator saturation estimator as a robust regression estimator. In J.L. Castle and N. Shephard (eds.), *The Methodology and Practice of Econometrics*. Oxford: Oxford University Press. Forthcoming.

Judge, G.G. and M.E. Bock (1978) *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*. Amsterdam: North-Holland.

Juselius, K. (1993) VAR modelling and Haavelmo's probability approach to econometrics. *Empirical Economics* **18**, 595–622.

Juselius, K. and M. Franchi (2007) Taking a DSGE model to the data meaningfully. *Economics: The Open-Access, Open-Assessment E-Journal* **1**, 4.

Keynes, J.M. (1939) Professor Tinbergen's method. *Economic Journal* **44**, 558–68. Reprinted in D.F. Hendry and M.S. Morgan (eds.), *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press, 1995.

Keynes, J.M. (1940) Statistical business-cycle research: Comment. *Economic Journal* **50**, 154–6.

Kleibergen, F. (2002) Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica* **70**, 1781–803.

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

Koopmans, T.C. (1937) *Linear Regression Analysis of Economic Time Series*. Haarlem: Netherlands Economic Institute.

Koopmans, T.C. (1947) Measurement without theory. *Review of Economics and Statistics* **29**, 161–79.

Koopmans, T.C. (1950) When is an equation system complete for statistical purposes? In T.C. Koopmans (ed.), *Statistical Inference in Dynamic Economic Models*, No. 10 in Cowles Commission Monograph, Ch. 17. New York: John Wiley & Sons.

Krolzig, H.-M. (2003) General-to-specific model selection procedures for structural vector autoregressions. *Oxford Bulletin of Economics and Statistics* **65**, 769–802.

Kurcewicz, M. and J. Mycielski (2003) A specification search algorithm for cointegrated systems. Discussion paper, Statistics Department, Warsaw University.

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**, 161–78.

Le Gall, P. (2007) *A History of Econometrics in France: From Nature to Models*. London: Routledge.

Leamer, E.E. (1978) *Specification Searches. Ad-Hoc Inference with Non-Experimental Data*. New York: John Wiley.

Leamer, E.E. (1983) Let's take the con out of econometrics. *American Economic Review* **73**, 31–43.

Leeb, H. and B.M. Pötscher (2003) The finite-sample distribution of post-model-selection estimators, and uniform versus non-uniform approximations. *Econometric Theory* **19**, 100–42.

Leeb, H. and B.M. Pötscher (2005) Model selection and inference: facts and fiction. *Econometric Theory* **21**, 21–59.

Lovell, M.C. (1983) Data mining. *Review of Economics and Statistics* **65**, 1–12.

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.

Magnus, J.R. and M.S. Morgan (eds.) (1999) *Methodology and Tacit Knowledge: Two Experiments in Econometrics*. Chichester: John Wiley and Sons.

Mavroeidis, S. (2004) Weak identification of forward-looking models in monetary economics. *Oxford Bulletin of Economics and Statistics* **66**, 609–35.

Mayo, D.G. and A. Spanos (2006) Severe testing as a basic concept in a Neyman–Pearson philosophy of induction. *British Journal for the Philosophy of Science* **57**, 323–57.

Mills, T.C. (ed.) (1999) *Economic Forecasting* (two volumes). Cheltenham: Edward Elgar.

Mills, T.C. and K.D. Patterson (eds.) (2006) *Palgrave Handbook of Econometrics, Volume 1: Econometric Theory*. Basingstoke: Palgrave Macmillan.

Mitchell, B.R. (1988) *British Historical Statistics*. Cambridge: Cambridge University Press.

Mizon, G.E. (1984) The encompassing approach in econometrics. In D.F. Hendry and K.F. Wallis (eds.), *Econometrics and Quantitative Economics*, pp. 135–72. Oxford: Basil Blackwell.
Mizon, G.E. and J.-F. Richard (1986) The encompassing principle and its application to non-nested hypothesis tests. *Econometrica* **54**, 657–78.
Moene, K.A. and A. Rødseth (1991) Nobel Laureate: Trygve Haavelmo. *Journal of Economic Perspectives* **5**, 175–92.
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. (1923) *Generating Economic Cycles*. New York: Macmillan.
Morgan, M.S. (1990) *The History of Econometric Ideas*. Cambridge: Cambridge University Press.
Muth, J.F. (1961) Rational expectations and the theory of price movements. *Econometrica* **29**, 315–35.
Nerlove, M. (1983) Expectations, plans, and realizations in theory and practice. *Econometrica* **51**, 1251–79.
Nymoen, R. (2002) Faulty watch towers – "structural" models in Norwegian monetary policy analysis. Unpublished paper, University of Oslo.
Pagan, A.R. (1987) Three econometric methodologies: a critical appraisal. *Journal of Economic Surveys* **1**, 3–24.
Paruolo, P. and A. Rahbek (1999) Weak exogeneity in I(2) systems. *Journal of Econometrics* **93**, 281–308.
Peart, S.J. (2001) "Facts carefully marshalled" in the empirical studies of William Stanley Jevons. *History of Political Economy* **33**, 252–76.
Perez-Amaral, T., G.M. Gallo and H. White (2003) A flexible tool for model building: the relevant transformation of the inputs network approach (RETINA). *Oxford Bulletin of Economics and Statistics* **65**, 821–38.
Perez-Amaral, T., G.M. Gallo and H. White (2005) A comparison of complementary automatic modelling methods: RETINA and PcGets. *Econometric Theory* **21**, 262–77.
Phillips, A.W.H. (1957) Stabilization policy and the time form of lagged response. *Economic Journal* **67**, 265–77.
Phillips, P.C.B. (1994) Bayes models and forecasts of Australian macroeconomic time series. In C. Hargreaves (ed.), *Non-stationary Time-Series Analyses and Cointegration*. Oxford: Oxford University Press.
Phillips, P.C.B. (1995) Automated forecasts of Asia-Pacific economic activity. *Asia-Pacific Economic Review* **1**, 92–102.
Phillips, P.C.B. (1996) Econometric model determination. *Econometrica* **64**, 763–812.
Phillips, P.C.B. (2003) Laws and limits of econometrics. *Economic Journal* **113**, C26–52.
Phillips, P.C.B. and M. Loretan (1991) Estimating long-run economic equilibria. *Review of Economic Studies* **58**, 407–36.
Popper, K.R. (1963) *Conjectures and Refutations*. New York: Basic Books.
Pötscher, B.M. (1991) Effects of model selection on inference. *Econometric Theory* **7**, 163–85.
Qin, D. (1993) *The Formation of Econometrics: A Historical Perspective*. Oxford: Clarendon Press.
Qin, D. (2008) Consolidation of the Cowles Commission paradigm. Unpublished paper, Queen Mary College, London.
Ramsey, J.B. (1969) Tests for specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society B* **31**, 350–71.
Richard, J.-F. (1980) Models with several regimes and changes in exogeneity. *Review of Economic Studies* **47**, 1–20.
Robbins, L. (1932) *An Essay on the Nature and Significance of Economic Science*. London: Macmillan.
Robinson, P.M. (1995) Log-periodogram regression of time series with long range dependence. *Annals of Statistics* **23**, 1048–72.
Rothenberg, T.J. (1971) Identification in parametric models. *Econometrica* **39**, 577–92.

Sala-i-Martin, X.X. (1997) I have just run two million regressions. *American Economic Review* **87**, 178–83.
Samuelson, P.A. (1947) *Foundations of Economic Analysis*. Cambridge, Mass.: Harvard University Press.
Sargan, J.D. (1983) Identification and lack of identification. *Econometrica* **51**, 1605–33.
Schultz, H. (1928) *The Theory and Measurement of Demand*: Chicago: University of Chicago Press.
Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics* **6**, 461–4.
Siegel, D.S. and M. Wright (2007) Intellectual property: the assessment. *Oxford Review of Economic Policy* **23**, 529–40.
Sims, C.A. (1980) Macroeconomics and reality. *Econometrica* **48**, 1–48.
Smets, F. and R. Wouters (2003) An estimated stochastic dynamic general equilibrium model of the Euro Area. *Journal of the European Economic Association* **1**, 1123–75.
Smith, A. (1759) *Theory of Moral Sentiments*. Edinburgh: A. Kincaid & J. Bell.
Smith, A. (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: W. Strahan & T. Cadell.
Smith, A. (1795) The history of astronomy. In D. Stewart (ed.), *Essays on Philosophical Subjects by Adam Smith*, pp. 33–105. Edinburgh: W. Creech. Liberty Classics edition, by I.S. Ross, 1982.
Smith, R.J. (1992) Non-nested tests for competing models estimated by generalised method of moments. *Econometrica* **60**, 973–80.
Smith, R.J. (2007) Efficient information theoretic inference for conditional moment restrictions. *Journal of Econometrics* **138**, 430–60.
Spanos, A. (1989) On re-reading Haavelmo: a retrospective view of econometric modeling. *Econometric Theory* **5**, 405–29.
Spanos, A. (1990) The simultaneous equations model revisited: statistical adequacy and identification. *Journal of Econometrics* **44**, 87–105.
Spanos, A. (1995) On theory testing in econometric modelling with non-experimental data. *Journal of Econometrics* **67**, 189–226.
Spanos, A. (2006) Econometrics in retrospect and prospect. In T.C. Mills and K.D. Patterson (eds.), *Palgrave Handbook of Econometrics, Volume I: Econometric Theory,* pp. 3–58. Basingstoke: Palgrave Macmillan.
Spanos, A., D.F. Hendry and J.J. Reade (2008) Linear vs. log-linear unit root specification: an application of mis-specification encompassing. *Oxford Bulletin of Economics and Statistics*. Forthcoming.
Staiger, D. and J.H. Stock (1997) Instrumental variables regression with weak instruments. *Econometrica* **65**, 557–86.
Stigler, G.J. (1962) Henry L. Moore and statistical economics. *Econometrica* **30**, 1–21.
Stigler, S.M. (1986) *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge, Mass.: Harvard University Press.
Stigler, S.M. (1999) *Statistics on the Table*. Cambridge, Mass.: Harvard University Press.
Stiglitz, J. (2003) Is Keynes dead? Reviving a sensible macroeconomics. Clarendon lectures, Department of Economics, University of Oxford.
Stigum, B.P. (1990) *Towards a Formal Science of Economics*. Cambridge, Mass.: MIT Press.
Stock, J.H. and J.H. Wright (2000) GMM with weak identification. *Econometrica* **68**, 1055–96.
Stock, J.H., J.H. Wright and M. Yogo (2002) A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics* **20**, 518–29.
Stone, J.R.N. (1951) *The Role of Measurement in Economics*. Cambridge: Cambridge University Press.
Summers, L.H. (1991) The scientific illusion in empirical macroeconomics. *Scandinavian Journal of Economics* **93**, 129–48.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, B, **58**, 267–88.

Tinbergen, J. (1930) Determination and interpretation of supply curves: an example [Bestimmung und Deutung von Angebotskurven: ein Beispiel]. *Zeitschrift fur Nationalökonomie* **1**, 669–79. Reprinted in D.F. Hendry and M.S. Morgan (eds.), *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press, 1995.

Tinbergen, J. (1939) *Statistical Testing of Business-Cycle Theories. Volume I: A Method and its Application to Investment Activity*. Geneva: League of Nations.

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

Tobin, J. (1950) A statistical demand function for food in the U.S.A. *Journal of the Royal Statistical Society A* **113**(2), 113–41.

Tobin, J. (1952) A survey of the theory of rationing. *Econometrica* **26**, 24–36.

Tress, R.C. (1959) The contribution of economic theory to economic prognostication. *Economica* **26**, 194–211.

Urbain, J.-P. (1992) On weak exogeneity in error correction models. *Oxford Bulletin of Economics and Statistics* **54**, 187–207.

Vining, R. (1949a) Koopmans on the choice of variables to be studied and of methods of measurement. *Review of Economics and Statistics* **31**, 77–86.

Vining, R. (1949b) A rejoinder. *Review of Economics and Statistics* **31**, 91–4.

von Weizsacker, C.C. (2005) The welfare economics of adaptive preferences. Preprint 2005/11, Max Planck Institute, Bonn.

Weissmann, G. (1991) Asprin. *Scientific American*, January, 58–64.

White, H. (1980a) A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48**, 817–38.

White, H. (1980b) Using least squares to approximate unknown regression functions. *International Economic Review* **21**, 149–70.

White, H. (1990) A consistent model selection. In C.W.J. Granger (ed.), *Modelling Economic Series*, pp. 369–83. Oxford: Clarendon Press.

White, H. (2000) A reality check for data snooping. *Econometrica* **68**, 1097–126.

White, H. (2008) Approximate nonlinear forecasting methods. Mimeo, Economics Department, University of California at San Diego.

Working, E.J. (1927) What do statistical demand curves show? *Quarterly Journal of Economics* **41**, 212–35.

Wright, P.G. (1915) Review of Moore, "*Economic Cycles*" (1915). *Quarterly Journal of Economics* **29**, 631–41. Reprinted in D.F. Hendry and M.S. Morgan (eds.), *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press, 1995.

Wright, P.G. (1929) Review of H. Schultz: "*Statistical Laws of Demand and Supply*" (1915). *Journal of the American Statistical Association* **24**, 207–15.

Yule, G.U. (1897) On the theory of correlation. *Journal of the Royal Statistical Society* **60**, 812–38.

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