

Hendry, David F.; Mizon, Grayham E.

**Article**

## Improving the teaching of econometrics

Cogent Economics & Finance

**Provided in Cooperation with:**

Taylor & Francis Group

*Suggested Citation:* Hendry, David F.; Mizon, Grayham E. (2016) : Improving the teaching of econometrics, Cogent Economics & Finance, ISSN 2332-2039, Taylor & Francis, Abingdon, Vol. 4, Iss. 1, pp. 1-26,  
<https://doi.org/10.1080/23322039.2016.1170096>

This Version is available at:

<https://hdl.handle.net/10419/147809>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by/4.0/>



Received: 26 February 2016  
Accepted: 08 March 2016  
Published: 18 April 2016

\*Corresponding author: Grayham E. Mizon, Faculty of Social, Human, and Mathematical Sciences, University of Southampton, Southampton, UK; Economic Modelling Programme, Institute for New Economic Thinking, Oxford Martin School, University of Oxford, Oxford, UK  
E-mail: [grayham.mizon@soton.ac.uk](mailto:grayham.mizon@soton.ac.uk)

Reviewing editor:  
Steve Cook, Swansea University, UK

Additional information is available at the end of the article

## GENERAL & APPLIED ECONOMICS | RESEARCH ARTICLE

# Improving the teaching of econometrics

David F. Hendry<sup>1</sup> and Grayham E. Mizon<sup>2,3\*</sup>

**Abstract:** We recommend a major shift in the Econometrics curriculum for both graduate and undergraduate teaching. It is essential to include a range of topics that are still rarely addressed in such teaching, but are now vital for understanding and conducting empirical macroeconomic research. We focus on a new approach to macro-econometrics teaching, since even undergraduate econometrics courses must include analytical methods for time series that exhibit both evolution from stochastic trends and abrupt changes from location shifts, and so confront the “non-stationarity revolution”. The complexity and size of the resulting equation specifications, formulated to include all theory-based variables, their lags and possibly non-linear functional forms, as well as potential breaks and rival candidate variables, places model selection for models of changing economic data at the centre of teaching. To illustrate our proposed new curriculum, we draw on a large UK macroeconomics database over 1860–2011. We discuss how we reached our present approach, and how the teaching of macro-econometrics, and econometrics in general, can be improved by nesting so-called “theory-driven” and “data-driven” approaches. In our methodology, the theory-model’s parameter estimates are unaffected by selection when the theory is complete and correct, so nothing is lost, whereas when the theory is incomplete or incorrect, improved empirical models can

### ABOUT THE AUTHORS

David F. Hendry, Professor of Economics and Fellow of Nuffield College, Oxford University directs the Economic Modelling Programme, Institute for New Economic Thinking, Oxford Martin School. He was knighted in 2009, has received an ESRC Lifetime Achievement Award and eight Honorary Doctorates, is a Thomson Reuters Citation Laureate, and has published more than 200 papers and 25 books, including *Empirical Model Discovery and Theory Evaluation*, 2014, with Jurgen Doornik, and *Introductory Macroeconomics: A New Approach*, 2015.

Grayham E. Mizon is Emeritus Professor of Econometrics, Faculty of Social, Human, and Mathematical Sciences, University of Southampton, and Associate, Economic Modelling Programme, Institute for New Economic Thinking, Oxford Martin School, University of Oxford. He is a Foreign Honorary Member, Polish Academy of Sciences, and Honorary Research Fellow, Chinese Academy of Social Sciences. He has published more than 80 papers on econometric theory and econometric modelling.

### PUBLIC INTEREST STATEMENT

We recommend a major change in macro-econometrics teaching for both graduates and undergraduates to include methods for analysing time series that have both stochastic trends and abrupt shifts. To incorporate all theory-based variables, their lags, possibly non-linear functional forms, potential outliers, breaks and rival candidate variables, the resulting specifications place model selection facing non-stationary economic data at the centre of teaching. We illustrate our proposed curriculum using a large UK macroeconomics database over the turbulent period 1860–2011. We describe how the teaching of econometrics can be improved by nesting “theory-driven” and “data-driven” approaches, whereby the theory-model’s parameter estimates are unaffected by selection when the theory is correct, whereas improved empirical models can be discovered when the theory is incorrect. Recent software like Autometrics facilitates both the teaching and the implementation of econometrics, supported by simulation tools to examine operational performance, designed to be feasibly presented live in the classroom.

be discovered from the data. Recent software like *Autometrics* facilitates both the teaching and the implementation of econometrics, supported by simulation tools to examine operational performance, designed to be feasibly presented live in the classroom.

**Subjects:** Econometrics; Macroeconomics; Teaching & Learning

**Keywords:** teaching econometrics; model selection; theory retention; location shifts; indicator saturation; *Autometrics*; economic tools for teaching

**AMS subject classifications:** C51; C22

## 1. Introduction

Economic theories are inevitably incomplete characterizations of the complicated reality of economic life, and hence empirical models based thereon are bound to be mis-specified, and are not estimating “truth”. Because economies are very high dimensional, dynamic, evolving and subject to intermittent unanticipated shifts, students fitting regressions to real-world data often discover problems with their models, from residual autocorrelation, heteroskedasticity and non-normality through non-constant parameters and non-linearities. Although textbooks usually present procedures for handling symptoms of empirical mis-specification, such “solutions” are seriatim and separate, so often do not deal with the underlying problem. Further, trying to fix one mis-specification problem at a time can lead nowhere, as all may be due to some other drawback, such as omitted variables. Consequently, students need guidance on how to undertake empirical modelling above the theoretical material in standard textbooks. This paper is written to help teachers in that task, for students who have a basic knowledge of statistics and probability, and possibly econometrics, at the level of, say, the first six chapters of Hendry and Nielsen (2007).

There are four fundamental problems facing all empirical modelling exercises (see Doornik & Hendry, 2015). First, formulating a sufficiently general initial model to capture all the substantively relevant influences. Failure to do so precludes finding the underlying process. In particular, Hendry and Mizon (2014) explain why it is essential to address shifts of distributions in empirical macro-econometric modelling as otherwise the resulting models are doomed to be non-constant, and provide inappropriate policy advice.<sup>1</sup> Consequently, even the teaching of undergraduate econometrics must change to reflect such developments, and the “non-stationarity revolution” more generally, a theme expanded on in Hendry (2015).<sup>2</sup> Second, as such a formulation will usually be too large for humans to handle, and indeed may comprise more variables,  $N$ , than available observations,  $T$ , a powerful automatic model selection procedure is essential. Given many candidate variables, a tight significance level is needed to avoid an excess of adventitiously significant irrelevant variables. Fortunately, this need not affect retaining available theory insights. Third, it is crucial to rigorously evaluate any relationships claiming to explain the phenomena of interest. Finally, the selection algorithm must be able to sort through immense numbers of possible models in a reasonable time. These considerations are pertinent to all empirical research and in particular to economics. We seek to place the teaching of econometrics in this framework.

### 1.1. Past developments

When we began teaching undergraduate econometrics in the early 1970s, the emphasis was on the theory of estimation for linear regression models and linear simultaneous equation models, with practical demonstrations of these methods a rarity. While computations had certainly speeded up greatly compared to those on desktop calculators undertaken by e.g. Klein and his co-researchers in developing Klein model I (Klein, 1950), the available mainframe computers were relatively slow in processing the Hollerith-card delivered instructions and data. However, by the late 1970s, the rapid improvement in computing speed and associated equipment, including the development of personal computers (PCs), made it possible to use a “projection panel” that was connected to a PC and lying on top of an overhead slide projector to present live estimation of models in teaching and

seminar presentations. This was facilitated by the first author developing a suite of programmes derived from those used in his doctoral research for the analysis of autoregressive time series models (see Hendry & Srba, 1980), which was the precursor to the widely used *PcGive* software of today: Hendry and Doornik (1999) trace that development. Section 2.9 provides a brief overview of this and some of the other econometrics software that has been developed since the late 1970s. The extremely powerful computing equipment and the sophisticated, yet easy to use, software implementing the many advances in modelling strategy that are available today mean that it is now possible for empirical researchers to tackle the vast array of issues that they face in modelling economic systems. The magnitude of these developments and their success in modelling complex economic systems relative to the achievements of the widely used alternatives that pervade today's econometrics textbooks means that it is important for there to be a major shift in both undergraduate and graduate econometrics curricula. We will now present an overview of the evolution of the elements of our recommended modelling strategy, then illustrate its application in teaching.

The structure of the paper is as follows. Section 2 summarizes the many key concepts that have been clarified. Each is then amplified in Sections 2.1–2.9, followed by subsections to help in teaching undergraduates by explaining the basics of any material that involved substantive mathematics. Section 3 illustrates the new econometrics teaching that entails, considering in turn the roles of subject matter theory in Section 3.1, the database and software in Section 3.2, computing parameter estimates and evaluating models in Section 3.3, selecting a representation of the general unrestricted model in Section 3.4. We briefly note testing parameter constancy in Section 3.5, the validity of exogeneity in Section 3.6, and checking the need for a non-linear representation in Section 3.7, and using simulation to investigate the reliability of the outcome in Section 3.8. Section 4 concludes.

## 2. Key concepts

A set of inter-related concepts sustains our modelling strategy, commencing with the fundamental notion of a data generation process (DGP) and its representation. Data are obviously generated by some process, which in economics is a combination of the behaviour of the individuals in an economy (and other economies they interact with) and how that behaviour is measured. The total number of transactions in an economy far exceeds any hope of detailed modelling of them all, so macro-econometrics focuses on aggregates of variables such as gross national product (GNP), total unemployment (U) and so on. Section 2.1 considers further reductions to create manageably sized formulations. From the perspective of teaching, the first and most crucial step explaining the origin of empirical models through the theory of reduction unfortunately involves many technical concepts, some of which students will not have encountered previously. Moreover, any earlier courses probably assumed, explicitly or implicitly, that the specification of the model was given by an economic analysis. Certainly, the empirical counterpart of the theory model is the **objective** of the analysis. However, the only hope of ending with a well-specified empirical representation is for the **target** of the modelling exercise to be the process that actually generated the variables under analysis. Conflating these two distinct entities by asserting that “the model is the mechanism” is all too common, but unlikely to lead to robust empirical evidence. Our approach retains any available theory insights within the general model, so if the theory happened to be complete and correct, that is what would be found, but if it were either incomplete or incorrect, an improved representation would be discovered without the retention of the theory contaminating the final selection in any way. Thus, tackling the theory of reduction is imperative to place empirical modelling on a sound footing. A valuable collateral benefit is that many of the central concepts of modern econometrics correspond to reductions moving from the DGP to the model to be empirically analysed, including sufficient statistics, innovation errors, (Granger) causality, exogeneity, constancy and invariance, cointegration, conditioning and simultaneity.

The next question is how to judge whether such reductions lose valuable information about the objectives of the analysis, addressed in Section 2.2 by the concept of congruence between a model and the evidence. A strategy to locate congruent models is described in Section 2.3, followed in

Section 2.4 by an explanation of exogeneity and its role in empirical modelling. Empirical analyses almost never exist in isolation, so to evaluate the latest findings we require that a new model encompasses or explains previous results, discussed in Section 2.5. However, economies evolve over time and often suffer abrupt shifts, as with the recent Great Recession, so empirical modelling has to take account of such non-stationarities, the topic of Section 2.6. The final two developments concern the potential non-linearity of relationships, considered in Section 2.7, and how to undertake model selection in Section 2.8 in a way that addresses all of the potential complications inherent in economic data. While we mainly consider macroeconomic time series, similar principles apply to cross section and panel observational data (see e.g. Castle & Hendry, 2011, for an example).

### 2.1. Data generation process and its representation

The most crucial decision in all empirical studies concerns the set of variables to collect observations on and then analyse, which will be a small subset of all the variables in the economy. The factors influencing this decision include the focus of interest in modelling (which could be an economic policy issue, evaluating an economic theory, or forecasting), as well as economic theory, previous empirical results, and related institutional knowledge and economic history. Denoting these  $n$  variables by  $\{\mathbf{x}_t\}$  to mean the complete set  $(\mathbf{x}_1, \dots, \mathbf{x}_T)$  for a time period  $t = 1, \dots, T$ , then the focus of attention in modelling is to learn as much as possible about the process that generates  $\{\mathbf{x}_t\}$ , called the local data generation process (LDGP). For a sample of size  $T$ , the LDGP is represented by the joint density  $D_X(\mathbf{x}_1, \dots, \mathbf{x}_T | \boldsymbol{\phi}, \mathbf{X}_0)$  where  $D_X(\cdot)$  is the form of the density function (often assumed to be normal),  $\mathbf{X}_0$  is the set of initial conditions, characterized by the vector of  $k$  parameters  $\boldsymbol{\phi} \in \Psi$ , although these might be time varying so could be denoted  $\boldsymbol{\phi}_1^T$ . The variables  $\{\mathbf{x}_t\}$  are generated as part of a much larger set of variables,  $\mathbf{w}_t$ , generated in the economy under analysis, which might be the global economy for present-day modelling. Such a DGP is far too high dimensional, heterogeneous and time varying to accurately theorize about or model in detail, whereas theorizing and learning about the LDGP for  $\{\mathbf{x}_t\}$  is often feasible. To structure the analysis, we assume that the DGP of  $\{\mathbf{w}_t\}$  can be represented by the joint density  $D_W(\mathbf{w}_1, \dots, \mathbf{w}_T | \mathbf{W}_0, \mathbf{D}_1^T, \boldsymbol{\kappa}_1^T)$  where  $\mathbf{W}_0$  are the pre-sample outcomes,  $\mathbf{D}_1^T = (\mathbf{d}_1, \dots, \mathbf{d}_T)$  contains deterministic variables including constants, seasonal effects, trends and shifts, and  $\boldsymbol{\kappa}_1^T$  are the parameters of the decisions which led to the outcomes in  $D_W(\cdot)$ .

Underlying the decision to model the LDGP rather than the DGP is a series of reductions which are inevitable and intrinsic (see inter alia, Florens, Mouchart, & Rolin, 1990; Hendry, 1995a). Firstly, the decision to model the variables in  $\{\mathbf{x}_t\}$  and not those of the DGP  $\{\mathbf{w}_t\}$  entails discarding the remaining variables  $\bar{\mathbf{w}}_t$  when  $\mathbf{w}_t' = (\bar{\mathbf{w}}_t', \mathbf{x}_t')$ . Although it is possible to factor  $D_W(\mathbf{w}_1, \dots, \mathbf{w}_T | \cdot) = D_W(\mathbf{W}_1^T | \cdot)$  into  $D_W(\mathbf{W}_1^T | \cdot) = D_{\bar{\mathbf{w}}|\mathbf{x}}(\bar{\mathbf{W}}_1^T | \mathbf{X}_1^T, \cdot) D_X(\mathbf{X}_1^T | \cdot)$ , the decision to model just  $\{\mathbf{x}_t\}$  requires that there is no substantive loss of information incurred. This in turn requires that there is no Granger causality from the lagged values of  $\bar{\mathbf{w}}_t$  to  $\mathbf{x}_t$ , which is a demanding requirement (see Granger, 1969; Hendry & Mizon, 1999). Secondly, almost all economic data in  $\{\mathbf{x}_t\}$  are aggregated across commodities, agents, space and time, and in some cases, they are estimates of the “correct” aggregate based on small samples of activity. Thirdly, most econometric studies analyse data after transformations such as logarithms and growth rates, which can affect the constancy of, and cross-links between, the resulting parameters. In fact, since aggregates are linear sums of disaggregates, log transformations of the aggregates might be well behaved even though disaggregates are not.

If any of the reductions implied in moving from  $D_W(\mathbf{W}_1^T | \cdot)$  to  $D_X(\mathbf{X}_1^T | \cdot)$  are invalid or inappropriate, then the LDGP may be non-constant and so provide a poor representation of the future generation of  $\{\mathbf{x}_t\}$  despite describing the sample. By definition, the LDGP changes with changes in the set of variables  $\{\mathbf{x}_t\}$  to analyse, which means that the decision about what to include in  $\{\mathbf{x}_t\}$  is crucial in determining how good an LDGP will be. Thus, it will be valuable to include a wide range of variables from the outset, rather than beginning with a restricted set that is more likely to be inadequate. Although it is possible to include more variables as the inadequacies of a particular set are revealed, doing so incrementally is fraught with difficulties, e.g. which additional variables to consider and in what order. Once a reasonable choice of  $\{\mathbf{x}_t\}$  has been made, its LDGP may be too complicated to be recovered fully using available empirical and theoretical information. Hence, the potentially infinite lags



in  $D_{\mathbf{x}}(\mathbf{X}_1^T|\cdot)$  must be reduced to a small number of lags, and its parameters  $\phi_1^T$  have to depend on a smaller set of parameters  $\theta$  that are constant. The validity of the reduced lag length can be checked by testing if longer lags matter, and the constancy of  $\theta$ , at least within regimes, can also be tested directly. In order to proceed further empirically, a general unrestricted model (GUM) has to be specified that: (a) uses data transformations sufficiently general to capture those in the reduced LDGP; (b) includes all variables  $\{\mathbf{x}_t\}$  of the LDGP, with perhaps some additional variables that might transpire to be relevant; and (c) contains long enough lags and sufficient deterministic variables (including indicator, or dummy, variables) to be able to capture a constant parameter representation. With a judicious choice of parameters and variables, the LDGP might be nested within the GUM, and in this case, a well-specified model which embeds the economic theory and can deliver the parameters of interest should be obtainable. Alternatively, when the LDGP is not nested in the GUM, and so some of the reductions mentioned above involve important mis-specifications, it is difficult to establish what properties the final specific model will have, although a well-specified approximation can often still be found. Section 2.1.1 suggests an approach to teaching the main ideas of intrinsic reductions without the mathematics.

Other important considerations in formulating the GUM include taking into account the wide-sense non-stationarity of economic variables, the possibility of conditioning on exogenous variables  $\mathbf{z}_t$  when  $\mathbf{x}'_t = (\mathbf{y}'_t, \mathbf{z}'_t)$ , as well as possible simultaneous determination of the endogenous variables,  $\mathbf{y}_t$ , allowing the more parsimonious combinations  $\mathbf{B}\mathbf{y}_t$  (where  $\mathbf{B}$  is non-singular), rather than  $\mathbf{y}_t$ , to be modelled. The consequences of non-stationarity and simultaneous determination are discussed in Section 2.6, and the importance of valid conditioning is discussed in Section 2.4. For more details on each of the issues considered in this section, see Hendry (2009) and Hendry and Doornik (2014).

### 2.1.1. Explaining the basics of reduction theory

For undergraduates, how can a teacher get across the main ideas of intrinsic reductions without the mathematics? One approach after teaching basic regression is to use a simple linear model fitted to artificial data, with (say) 4 variables:  $y_t$  related to  $X_{1,t}, X_{2,t}, X_{3,t}$  and an intercept, where the  $X_{i,t}$  are quite highly correlated, are autocorrelated, and have large means, all with non-zero parameters in the DGP except  $X_{3,t}$ , which is in fact irrelevant. Such data are easily generated in, say, *PcNaive*, part of *OxMetrics* (see Doornik & Hendry, 2013b), and an invaluable tool for easy-to-create Monte Carlo simulations. Then, illustrate: (a) the DGP estimates; (b) those where each of  $X_{1,t}, X_{2,t}, X_{3,t}$  and the intercept are omitted in turn; and (c) where a lag of every remaining variable is added when any variable is omitted. Non-constancy could be added where the students can cope. A slight extension of the model in Section 3.1 could be used. Better still, Monte Carlo simulations can show students the estimator and test distributions, and allow comparisons between correctly specified and mis-specified cases in terms of both biased estimates and incorrect estimated standard errors, as against the correct standard deviations of the sampling distributions.

To assess the validity of reductions,  $t$ -tests for eliminating variables can be used. That for  $X_{3,t}$  should be insignificant, have little effect on the estimates of the remaining parameters, but increase the precision of the estimates, whereas the other reductions should all reject, with possibly large changes in parameter estimates (did the coefficient of  $X_{3,t}$  become “spuriously” significant?), and the appearance of residual autocorrelation. The impacts of incorrectly eliminating the intercept, a “fixed regressor”, merit discussion.

## 2.2. Congruence

Although the early econometrics textbooks (see e.g. Goldberger, 1964; Johnston, 1963) emphasized estimation theory and its analysis using matrix algebra, they also described simple hypothesis testing, such as individual coefficient Student's  $t$ , goodness of fit  $F$ , and Durbin-Watson (DW) statistics. Initially, the adequacy of a particular model specification was assessed by comparing the statistical significance, size and sign of estimated coefficients relative to economic theory, calculating the overall goodness of fit  $R^2$  and  $F$ , and inspecting mis-specification test statistics such as DW. However,

with the continuing increase in computing power, it became possible to consider models with larger numbers of explanatory variables including longer lags, and the possibility that non-linear models might be important. Equally important was the recognition that in developing empirical models, which rarely will be more than an approximation to the high dimensional and complicated LDGP, it is crucial that once  $x_t$  has been chosen, all the relevant information that is available is fully exploited. This idea is captured in the concept of congruence which requires an empirical model not to depart substantively from the evidence. In particular, it is desirable that the empirical model is theory consistent (coherent with well established a priori theory), data consistent (coherent with the observed sample information), and data admissible (consistent with the properties of the measurement system). The theory consistency of a model can be assessed via specification tests of the restrictions implied by the theory relative to the estimated model. Given that the error terms and the corresponding residuals are by definition unexplained by the model, finding that they exhibit systematic behaviour is an indication that there is potentially valuable information in the data that the model has not captured. Whether this is the case can be assessed via mis-specification tests for residual serial correlation, and heteroskedasticity, as well as for invalid conditioning and parameter non-constancies.

By definition, the DGP is congruent with itself, and provided that the LDGP is a valid reduction of the DGP, it will also be congruent. An empirical model is congruent if it is indistinguishable from the LDGP when the LDGP is a valid reduction of the DGP. Given the importance of the GUM in the modelling strategy that we advocate, it is crucial that it is congruent, and that in seeking simplifications of it, congruence is maintained for a simpler model to be acceptable. Although the diagnostic tests used to check the congruence of the GUM and subsequent model selections are designed to affect the operating characteristics of general-to-specific selection algorithms such as *Autometrics* (e.g. possibly retaining insignificant variables to avoid diagnostic test rejections), selection does not affect their null distributions (see Hendry & Krolzig, 2005). The potential costs of not testing the congruence of the GUM, and simplifications thereof, are that it may be non-congruent and so adversely affect all inferences during the selection process. To re-emphasize, a non-congruent model fails to account for, or make use of, all relevant theoretical and empirical information that is available and so is inadequate. For further details see inter alia Hendry (1995a,2009), Mizon (1995a), Bontemps and Mizon (2003), and Hendry and Doornik (2014).

### 2.2.1. Explaining the basics of congruence

Undergraduates will almost surely have been taught some Euclidean geometry so know about congruent triangles, namely ones which match perfectly, perhaps after rotation. However, they may never have thought that one of the triangles could be the cut-off top of a triangular pyramid, so the match is only two-dimensional, and the third dimension is not “explained”. That is precisely why the name congruence was originally introduced, as a model may match the DGP only where tested, and many aspects may not be matched. Thus, congruence is not “truth”, though DGPs must be congruent, so non-congruent models cannot be DGPs. However, as discussed in Section 2.5, a sequence of congruent models is feasible in a progressive research strategy, where each explains the results of all earlier models.

### 2.3. General-to-specific

The possibility that higher order autoregressive processes might be important to adequately capture the dynamics in time series models led to analyses of the relative merits of sequential testing from simple to general specifications as opposed to simplifying general models. Parsimony has long been thought to be a highly desirable property for a model to have—why include unnecessary features? However, it is also important to include necessary features in order to find a reliable and well-specified model. The tension between the costs of profligacy, including unnecessary variables, and excessive parsimony, omitting necessary variables, led to the development of a number of alternative modelling strategies. Given the computational difficulties of modelling complicated processes in the 1970s, it was tempting to start with simple formulations, possibly embodying a specific economic theory, then assess the need to generalize them. These expanding, or specific-to-general, model

selection methods require a criterion for the termination of the search, and this is often based on a measure of penalized goodness of fit or marginal significance. For example, the next most significant omitted regressor is added to the model with the expansion stopping when no further significant variables can be found. This simple search strategy can be extended, as in stepwise regression, by also removing insignificant regressors from the model. While stepwise regression can work well in some circumstances, such as independent, white-noise regressors, in others involving complex interdependencies they can fail badly. Other expanding search methods have been developed more recently, e.g. RETINA (Perez-Amaral, Gallo, & White, 2005) and Lasso (Efron, Hastie, Johnstone, & Tibshirani, 2004), with a large literature on shrinkage-based methods, but there is also a substantial literature illustrating the drawbacks of such approaches (see e.g. Anderson, 1962; Campos, Ericsson, & Hendry, 2005; Hendry, 1995a).

The benefits of working from general models and testing for the acceptability of simplification was established by Anderson (1962) in the context of ordered sequences for the determination of the order of trend polynomials, and in Anderson (1971) for autoregressive processes. Mizon (1977) extended this analysis for some non-ordered sequences of hypotheses, which is the usual case in econometrics, and pointed out the need for a structured search. A contracting search strategy begins from a general model with insignificant variables being deleted until a termination criterion is reached, e.g. as in running stepwise regression backwards by including all variables initially then eliminating insignificant terms one by one, although there are dangers in only exploring one search path. Hendry and Doornik (1994) list the advantages of a simplification strategy when modelling linear dynamic systems, Mizon (1995a) illustrated the superiority of a general-to-specific strategy in locating the LDGP over a specific-to-general strategy in a simulation study using artificially generated data, and similarly Mizon (1995b) illustrated this point in a study of quarterly aggregate UK wages, prices and unemployment data over the period 1965(1) to 1993(1). However, these examples are for small numbers of variables where there is a limited set of possible search paths, so can be implemented manually. But to capture the complex features of the typical LDGP of today's macro-econometrics, it is necessary to model in high dimensions, and even expert modellers are not capable of handling all the resulting possible search paths. Fortunately, advances in computing power and software development mean that model complexity is no longer a limitation on the choice of modelling strategy, which instead can be based on the desired properties of the resulting selected model. Indeed, as anticipated by Hendry and Mizon (2000) computer-automated search algorithms are now available that efficiently achieve results beyond human capabilities.

Despite a considerable literature arguing against the usefulness of empirical model discovery via general-to-specific searches (see inter alia Faust & Whiteman, 1997; Leamer, 1978,1983; Lovell, 1983), an impressive record has been built up for this approach. Following the stimulus given by Hoover and Perez (1999), the general-to-specific (*Gets*) algorithm *PcGets* implemented within *PcGive* by Hendry and Krolzig (1999), Hendry and Krolzig (2001) and Krolzig and Hendry (2001), quickly established the credentials of the approach via Monte Carlo studies in which *PcGets* recovered the DGP with an accuracy close to that to be expected if the DGP specification were known, but tests for significant coefficients were undertaken. Also, when *PcGets* was applied to the data-sets analysed in empirical studies by Davidson, Hendry, Srba, and Yeo (1978) and Hendry and Ericsson (1991), it selected in seconds models that were at least as good as those developed over several years by those authors (see Hendry, 2000). The general-to-specific strategy in *Autometrics* (see Doornik, 2009) employs a multi-path search algorithm which can combine expanding and contracting searches, so can handle more variables than observations, a feature that is particularly valuable for analysing non-stationary processes. Another key feature of *Gets* is that it is based on selecting variables rather than whole models and so is more flexible and open to discovery. We discuss this and the subsequent developments of the *Autometrics* algorithm in Section 2.8.

### 2.3.1. Explaining the basics of *Gets*

Although a critical decision in all empirical modelling is where to start from—general, simple or in between—all modelling strategies, including automated ones, require the formulation of a general



information set at the outset. Consequently, this general information set provides a well-defined initial model from which a contracting modelling strategy can proceed to test simplifications when  $N \ll T$ . For specific-to-general modelling strategies on the other hand, the general information set provides a well-defined list of variables from which to select expansions. Important considerations in the choice of the general information set include: the subject matter of the empirical modelling; institutional knowledge; past experience; data quality and availability; and the results of previous investigations. However, choosing a GUM to include as much as possible of the available relevant information makes it more likely that it will be congruent, which is an essential requirement for viable inferences during search procedures. We will consider the case where  $N \geq T$  below.

#### 2.4. Exogeneity

Exogeneity, in the sense of a variable being determined “outside the model under analysis”, has a long history in economics and econometrics. Early textbooks of econometrics concentrated on the estimation and testing of linear regression models in which the regressors were assumed exogenous by being fixed in repeated samples (see e.g. Goldberger, 1964, p. 162), an assumption relevant in experimental sciences but not in economics where data are largely observational. Although a convenient simplification, “fixity” is rarely appropriate in economics and has counter-examples (see e.g. Hendry, 1995a, p. 161), so a more relevant concept of exogeneity was needed. This was important for estimating dynamic models with lagged values of the regressand and regressors, and particularly in simultaneous equation models (SEMs), seeking to analyse the joint determination of several variables where the exogeneity of conditioning variables was questionable. Moreover, an appropriate form of exogeneity is critical for reliable forecasting and policy analyses from conditional models.

Re-consider the LDGP  $D_X(\mathbf{x}_1, \dots, \mathbf{x}_T | \Phi, \mathbf{X}_0)$  in Section 2.1 for  $T$  observations on the  $n$ -dimensional vector of random variables  $\mathbf{x}_t$ . Letting  $\mathbf{X}_{t-1} = (\mathbf{X}_0, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{t-1})$ , then  $D_X(\cdot)$  can be sequentially factorized as  $D_X(\mathbf{X}_1^T | \Phi, \mathbf{X}_0) = \prod_1^T D_x(\mathbf{x}_t | \mathbf{X}_{t-1}, \Phi)$  without loss of generality. Partitioning  $\mathbf{x}_t$  as  $\mathbf{x}_t' = (\mathbf{y}_t', \mathbf{z}_t')$  enables the conditional factorization  $\prod_1^T D_x(\mathbf{x}_t | \mathbf{X}_{t-1}, \Phi) = \prod_1^T D_{y|z}(\mathbf{y}_t | \mathbf{z}_t, \mathbf{X}_{t-1}, \theta_1) \prod_1^T D_z(\mathbf{z}_t | \mathbf{X}_{t-1}, \theta_2)$  when  $\theta' = (\theta_1', \theta_2') = \mathbf{f}(\Phi)'$ . If the parameters of interest to the modeller are  $\psi$ , and these can be recovered solely from  $\theta_1$  (i.e.  $\psi = \mathbf{h}(\theta_1)$ ) when the parameters  $\theta_1$  and  $\theta_2$  are variation free (so there are no parametric restrictions linking  $\theta_1$  and  $\theta_2$ ), then  $\mathbf{z}_t$  is weakly exogenous for  $\psi$  (see Engle, Hendry, & Richard, 1983, Hendry, 1995a, 1995b). Weak exogeneity is a sufficient condition for inference on  $\psi$  to be without loss of information using the conditional distribution  $D_{y|z}(\mathbf{y}_t | \mathbf{z}_t, \mathbf{X}_{t-1}, \theta_1)$  alone. However, weak exogeneity in isolation is not sufficient to sustain predictions of  $\mathbf{y}_t$  conditional on  $\mathbf{z}_t$  more than one period ahead because  $\mathbf{z}_t$  may vary with  $\mathbf{Y}_{t-1}$  when  $\mathbf{X}_{t-1}' = (\mathbf{Y}_{t-1}', \mathbf{Z}_{t-1}')$ . In order for reliable predictions of  $\mathbf{y}_t$  to be made from the conditional distribution  $D_{y|z}(\mathbf{y}_t | \mathbf{z}_t, \mathbf{X}_{t-1}, \theta_1)$ , then  $\mathbf{z}_t$  must be both weakly exogenous for  $\psi$  and must not vary with  $\mathbf{Y}_{t-1}$ . The latter condition entails that  $D_z(\mathbf{z}_t | \mathbf{X}_{t-1}, \theta_2) = D_z(\mathbf{z}_t | \mathbf{Z}_{t-1}, \theta_2)$  and is the condition for  $y$  not to Granger cause  $z$  (Granger, 1969). However, the absence of Granger causality is neither necessary nor sufficient for weak exogeneity, so cannot *per se* validate conditional inference.

Conditional econometric models are also important for assessing and predicting the likely effects of policy changes in interest rates, tax rates, welfare benefits, etc. The fact that economic processes intermittently undergo location shifts and intrinsically exhibit stochastic trends and other wide-sense non-stationarities (see Hendry & Mizon, 2014) means that parameter constancy and invariance (i.e. not changing when there is a change in policy regime) cannot be guaranteed, so must be tested. There was a time when the feasibility of such tests was doubted, for example, when Lucas (1976), following the concerns expressed earlier by Frisch (1938), Haavelmo (1944) and Marschak (1953), asserted that “any change in policy will systematically alter the structure of econometric models” and so render policy analysis infeasible. This claim, known as the Lucas critique, appeared to many to be a fatal blow to econometric policy analysis. Fortunately, the concept of super exogeneity defined in Engle et al. (1983) provides the condition for valid econometric policy analysis, and importantly it is testable as shown initially by Favero and Hendry (1992). The conditioning variables  $\mathbf{z}_t$  are super exogenous for the parameters  $\psi$  if  $\mathbf{z}_t$  is weakly exogenous for  $\psi$  and  $\theta_1$  is invariant to

changes in  $\theta_2$ . The requirement that  $\theta_1$  be invariant to changes in  $\theta_2$  entails that policy regime shifts in the marginal process for  $\mathbf{z}_t$  do not alter the parameters  $\theta_1$  of  $D_{y|z}(\mathbf{y}_t | \mathbf{z}_t, \mathbf{X}_{t-1}, \theta_1)$ , which are critical in assessing the effect on  $\mathbf{y}_t$  of those policy changes in  $\mathbf{z}_t$ . Note that super exogeneity does not require strong exogeneity, but only weak exogeneity and invariance. This is vital, as the behaviour of past values of  $\mathbf{y}$  is usually an important input into changes in the policy variables within  $\mathbf{z}$ , so  $\mathbf{z}$  cannot be strongly exogenous. The testing of super exogeneity, and in particular, invariance, requires a class of changes in  $\theta_2$  to be considered, and parameter constancy tests applied to the marginal process for  $\mathbf{z}$  are described in Engle and Hendry (1993). Hendry and Santos (2010) introduced automatic testing of super exogeneity using impulse-indicator saturation (IIS) to detect location shifts in the processes for the conditioning variables  $\mathbf{z}_t$ , then testing the relevance of the significant indicators in the conditional model. This test of super exogeneity can be computed without additional intervention from the investigator, and without knowing *ex ante* the timings, forms or magnitudes of the breaks in the marginal process for  $\mathbf{z}_t$ .

Deterministic terms such as dummy variables for seasons, outliers and structural breaks have been routinely used in econometric modelling for many years. However, when investigating why many researchers had experienced difficulties in modelling US demand for food in the 1930s and 1940s, Hendry (1999) found that by introducing zero-one indicators for 1931–1953, simplified to dummy variables for 1931–1936, 1938, and 1941–1946, led to a model that appeared to be constant over the whole period. This was tested using the Chow (1960) test of parameter constancy over the period 1953–1989, which Salkever (1976) had shown was equivalent to testing the significance of impulse indicators over that period. Hence, zero-one impulse indicators had been included for every observation from 1931 onwards, but in two large blocks for the periods 1931–1953 and 1953 on, so as many impulse indicators as observations had been used, plus all the regressors. This realization meant that models with more variables,  $N$ , than observations,  $T$ , could be investigated in a *Gets* framework, provided the variables are introduced in blocks. This discovery led to the use of IIS for detecting and removing the effects of structural shifts, outliers and data discrepancies, thus helping to ensure near normality in residuals and sustain valid inferences, and make bias correction after selection viable. IIS creates a zero-one indicator for each observation in the sample, which are then entered in blocks, noting that such indicators are mutually orthogonal. In the simplest case in which just two blocks of  $T/2$  are used, the first step is to add half the indicators and use an automatic search procedure (e.g. *Autometrics*) to select significant variables, then record these results. Next, drop the first set of indicators, and search again for significant indicators in the second set and record the results. Finally, perform the variable search again with the significant indicators combined. Setting the retention rate of irrelevant variables in the selected model (the gauge) to  $\alpha$  means that overall  $\alpha T$  indicators will be retained on average by chance. Thus, then the smaller model setting  $\alpha \leq r/T$  (with  $r$  small, such as unity) ensures on average a false null retention of  $r$  indicators, which is a small efficiency loss when testing for any number of breaks at  $T$  points. More details of IIS are given by Hendry, Johansen, and Santos (2008), who proposed IIS for detecting and removing outliers when they are present. Johansen and Nielsen (2009) provided a comprehensive theoretical justification of IIS, in particular extending the analysis to dynamic models. When testing exogeneity, IIS can have low power if changes in the conditional process are infrequent, but this problem can be circumvented using step indicator saturation (SIS) instead (see Castle, Doornik, Hendry, & Pretis, 2015).

Ericsson and Irons (1994) reprint many of the key papers on exogeneity, Ericsson and Irons (1996) provide an overview of the literature on super exogeneity and its relationship to the Lucas critique, and Hendry and Doornik (2014) give more details of the testing of super exogeneity using IIS and SIS.

#### 2.4.1. Explaining the basics of exogeneity

For undergraduates, how can a teacher get across the main ideas of exogeneity without too much distributional mathematics? Again after teaching basic regression, use the simplest linear model  $\mathbf{y}_t = \beta \mathbf{x}_t + \epsilon_t$  where  $\epsilon_t \sim \text{IN}[0, \sigma_\epsilon^2]$  which is independent of all values of  $\{\mathbf{x}_t\}$ , so the Gauss–Markov theorem apparently applies with the least-squares estimator  $\hat{\beta}$  of  $\beta$  being best linear unbiased. Now,

introduce the distribution of  $\{x_t\}$  as  $x_t \sim \text{IN}[\beta, \sigma_x^2]$  where  $\sigma_x^2$  is tiny, but large enough to allow  $\hat{\beta}$  to be calculated. Then,  $\bar{\beta} = T^{-1} \sum_1^T x_t$ , which is linear in  $x_t$ , can be a far better unbiased estimator of  $\beta$ .

This “contradiction” with Gauss–Markov arises because  $x_t$  is not weakly exogenous in the conditional model for  $y_t$  given the cross-link of the parameter  $\beta$  between the conditional and marginal distributions. Thus, independence between errors and regressors is insufficient, and even the venerable Gauss–Markov theorem needs to be supplemented by a weak exogeneity condition.

This example also has implications for super exogeneity. Consider a policy context where an agency controls the  $x_t$  process (e.g. interest rates) and can change the parameter  $\beta$  by setting the level of the interest rate. Doing so when  $\beta$  is also the parameter of the conditional model can be seen to always alter that model when there are cross-links. If its parameters change every time the policy changes, then clearly a model is not useful for policy—this is essentially an extreme “Lucas critique”—so failures of super exogeneity have important implications.

### 2.5. Encompassing

Encompassing is a principle that aims to reconcile the plethora of empirical models that often can be found to “explain” any given phenomenon. The infamous ability of economists as a profession to develop multiple theories for the explanation of a single phenomenon provides a rich source of potential interpretations of empirical evidence. Equally, in other areas of research such as epidemiology, experts cite polar opposite evidence, as regularly occurs in the TV program “Trust me—I’m a Doctor”. Generally, there seems to be a lack of an encompassing approach in other observational disciplines, although “meta analyses” are an approximation. Some interpretations might be complementary, and so could be amalgamated in a single theory, but the majority are usually alternatives, so it is necessary to discriminate between them. Indeed, when there are several distinct competing models, all but one must be either incomplete or incorrect, and all may be false. Adopting the encompassing principle in such situations enables testing whether any model can account for the results of the alternative models and so reduce the set of admissible models, and in addition reveal the directions in which a model under-performs relative to its rivals. This lays the foundations for a progressive modelling strategy in which theory and evidence mutually interact to learn about the LDGP, noting that empirical modelling is not a once-for-all event, but a process in which models evolve to supersede earlier ones.

The early development of the encompassing approach can be found in Davidson et al. (1978) and Davidson and Hendry (1981) which informally sought to find a model that was capable of accounting for the behaviour of alternative competing models in the context of an aggregate UK consumption function. Mizon (1984) provided a formal discussion of the encompassing principle, which was further developed in Mizon and Richard (1986). Adopting their statistical framework, consider two distinct empirical models  $\mathcal{M}_1$  and  $\mathcal{M}_2$  with parameters  $\alpha$  and  $\beta$  respectively, each purporting to provide an explanation of the process that generates a variable  $y_t$  conditional on  $z_t$  and lagged values of both, namely the LDGP  $D_{y|z}(y_t|z_t, \mathbf{X}_{t-1}, \theta_1)$ . Then,  $\mathcal{M}_1$  encompasses  $\mathcal{M}_2$  (denoted  $\mathcal{M}_1 \mathcal{E} \mathcal{M}_2$ ) if and only if  $\hat{\beta} = \beta(\hat{\alpha})$  where  $\hat{\beta}$  is the estimator of  $\beta$  under  $\mathcal{M}_2$  and  $\beta(\hat{\alpha})$  is the estimator of the pseudo-true value of  $\beta$  under  $\mathcal{M}_1$ . The test of encompassing is whether  $\mathcal{M}_2$  captures features of the LDGP beyond those already embodied in  $\mathcal{M}_1$ . If  $\mathcal{M}_2$  does not offer any new insights into the LDGP beyond those of  $\mathcal{M}_1$ , then  $\mathcal{M}_1 \mathcal{E} \mathcal{M}_2$ . The encompassing principle implies that if  $\mathcal{M}_1 \mathcal{E} \mathcal{M}_2$ , then  $\mathcal{M}_1$  ought to be capable of explaining the predictions of  $\mathcal{M}_2$  about all features of the LDGP, so  $\mathcal{M}_1$  can accurately characterize these, making  $\mathcal{M}_2$  redundant in the presence of  $\mathcal{M}_1$ . Equally, when  $\mathcal{M}_1 \mathcal{E} \mathcal{M}_2$ , then  $\mathcal{M}_1$  ought to be able to indicate some of the mis-specifications of  $\mathcal{M}_2$  such as omitted variables, residual serial correlation and heteroskedasticity, invalid conditioning, or predictive failure.

An important distinction can be drawn between situations in which  $\mathcal{M}_1$  is nested within  $\mathcal{M}_2$ , when encompassing tests the validity of the reductions leading from  $\mathcal{M}_2$  to  $\mathcal{M}_1$ , and those in which  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are non-nested so that neither model is a restricted version of the other. Cox (1961, 1962) are the

seminal papers on testing non-nested hypotheses, with the many related encompassing developments since then reviewed in Mizon (2012). When  $\mathcal{M}_1$  is nested within  $\mathcal{M}_2$  ( $\mathcal{M}_1 \subseteq \mathcal{M}_2$ ) and  $\mathcal{M}_1 \mathcal{E} \mathcal{M}_2$ , then the smaller model explains the results of the larger nesting model and so  $\mathcal{M}_1$  is a parsimonious representation of the LDGP relative to  $\mathcal{M}_2$ . The principle of parsimony has long been an important ingredient of model selection procedures that seek to find the simplest undominated model, but with many different penalty functions adopted for lack of parsimony (e.g. the AIC or Schwarz criteria, see Judge, Griffiths, Hill, Lütkepohl, & Lee, 1985). In the context of the encompassing principle,  $\mathcal{M}_1$  parsimoniously encompasses  $\mathcal{M}_2$  (denoted  $\mathcal{M}_1 \mathcal{E}_p \mathcal{M}_2$ ) when  $\mathcal{M}_1 \subseteq \mathcal{M}_2$  and  $\mathcal{M}_1 \mathcal{E} \mathcal{M}_2$ , making it a suitable strategy for checking reductions from a GUM within the Gets procedure.  $\mathcal{E}_p$  also satisfies the three conditions for a partial ordering (see Hendry, 1995a, chapter 14) as it is (i) reflexive since  $\mathcal{M}_1 \mathcal{E}_p \mathcal{M}_1$ ; (ii) asymmetric since  $\mathcal{M}_1 \mathcal{E}_p \mathcal{M}_2$  implies that  $\mathcal{M}_2$  does not  $\mathcal{E}_p \mathcal{M}_1$  when  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are distinct; and (iii) transitive since  $\mathcal{M}_1 \mathcal{E}_p \mathcal{M}_2$  and  $\mathcal{M}_2 \mathcal{E}_p \mathcal{M}_3$  imply that  $\mathcal{M}_1 \mathcal{E}_p \mathcal{M}_3$ . Thus parsimonious encompassing is a vital principle to incorporate in a modelling strategy such as Gets as it will enable the gradual accumulation of knowledge, and plays a key role in *Autometrics* (see Doornik, 2008).

The outcome of empirical analysis may suggest that a more general formulation is needed to obtain a better approximation to an LDGP, or that a larger set of variables is required to define a different LDGP that is more constant and interpretable. Importantly, though it is shown by White (1990) that sufficiently rigorous testing followed by suitable model re-specification ensures the selection of an acceptable data representation of a constant LDGP as the sample size tends to infinity, provided that the significance level of the complete testing process is controlled and in particular declines as the sample size increases. Although any approach might eventually converge on a constant LDGP as the sample size increases, the Gets strategy can do so relatively quickly. Commencing from a sufficiently general GUM that nests, or closely approximates, the LDGP has the advantage of reducing the chance that an extension of the data-set will be required later. In addition, by always requiring models to be congruent ensures that seeking parsimonious encompassing of successive models sustains a progressive modelling strategy. For further details see inter alia Mizon (2008) and the papers in Hendry, Marcellino, and Mizon (2008), particularly Bontemps and Mizon (2008).

### 2.5.1. Explaining the basics of encompassing

Even for undergraduates, the concept of encompassing should be intuitive: if one model cannot explain the results of another model on the same data, it must be incomplete or incorrect; and if it can, the second model is redundant. The illustration in Section 2.1.1 can be re-interpreted as an exercise in encompassing, and reveals the dangers of under-specification and how it can lead to conflicting claims. Slightly more formally, the DGP is:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + 0x_{3,t} + \epsilon_t \quad (1)$$

so  $x_{3,t}$  is irrelevant, but estimates of the other three parameters are highly significant. However, an investigator mistakenly fits:

$$M_1: y_t = \gamma_0 + \gamma_1 x_{1,t} + \gamma_3 x_{3,t} + v_t \quad (2)$$

and finds that  $\hat{\gamma}_0$ ,  $\hat{\gamma}_1$  and  $\hat{\gamma}_3$  are significant. Another investigator chooses to fit:

$$M_2: y_t = \delta_0 + \delta_2 x_{2,t} + \delta_3 x_{3,t} + u_t \quad (3)$$

and finds that  $\hat{\delta}_0$ ,  $\hat{\delta}_2$  and  $\hat{\delta}_3$  are significant. Since (1) includes both  $x_{1,t}$  and  $x_{2,t}$ , neither  $M_1$  encompasses  $M_2$ , nor  $M_2$  encompasses  $M_1$ , each being inadequate. To deal with the inadequacy of both models to encompass the other, estimation of:

$$M_3: y_t = \lambda_0 + \lambda_1 x_{1,t} + \lambda_2 x_{2,t} + \lambda_3 x_{3,t} + v_t \quad (4)$$

reveals that  $x_{3,t}$  is not required, but appeared to be relevant in  $M_1$  and  $M_2$  because of its correlation with  $x_{1,t}$  and  $x_{2,t}$ . Thus, the significance of  $x_{1,t}$  in (4) explains the failure of  $M_2$  to encompass  $M_1$ , and conversely the significance of  $x_{2,t}$  in (4) explains why  $M_1$  does not encompass  $M_2$ .

To enliven the coverage, a teacher could refer to the history of scientific discovery where encompassing has implicitly been prevalent: examples include Newton's theory of universal gravitation explaining Descartes' vortices, and Einstein's theory of general relativity explaining Newton; or Priestley discovering what he called "dephlogisticated air", explained by Lavoisier as oxygen, thereby replacing the theory of phlogiston with a modern theory of combustion.

## 2.6. Non-stationarity

The world we inhabit whether viewed from an economic, political, meteorological or cultural perspective can be beautiful and full of interesting objects and events, but it provides no shortage of challenges and surprises. The extreme weather events throughout the world, the financial crash in 2007–08 and the subsequent economic recession, and the political unrest in Eastern Europe and the Middle East leading to mass migration of homeless and impoverished people, are but recent examples. Thus, change is forever present—the world is not static. However, many economic theories have at their core stable relationships or equilibria between variables, and the concept of stationarity has long played a key role in statistics. A weakly stationary process is one where its mean and variance are finite and constant over time such as  $y_t = \mu + \epsilon_t$  with  $\epsilon_t \sim \text{IN}[0, \sigma_\epsilon^2]$  for which  $E[y_t] = \mu$  and  $V[y_t] = \sigma_\epsilon^2$ . A feature of a stationary process is that it is *ahistorical*, in that a sample drawn from one period of time will have the same characteristics as another drawn from a different period, so that knowing the historical dates reveals no additional information. Clearly though, many variables evolve over time including increases in world population, average life expectancy in western countries, and UK wages and prices. Indeed, most economic variables are non-stationary, in that their distributions shift, and we will consider two of the most important sources of such changes, namely stochastic trends and location shifts. Sometimes, it is argued that this non-stationary behaviour can be represented by a trend-stationary process like  $y_t = \mu + \lambda t + \epsilon_t$ , but this ignores the fact that population could not grow without food, with similar prerequisites for growth in other variables. In any case, it is unsatisfactory to attribute non-stationary behaviour to something outside the model.

In a stationary process, the influence of past shocks  $\epsilon_{t-s}$  for  $s > 0$  must die out, otherwise the variance  $V[y_t]$  could not be constant. One form of stationary process in which past shocks initially affect  $y_t$ , but their influence declines through time, is an autoregressive process such as  $y_t = \mu + \beta y_{t-1} + \epsilon_t$  which is stationary when  $|\beta| < 1$ . This has the moving-average representation  $y_t = \sum_{i=0}^{\infty} \beta^i (\mu + \epsilon_{t-i}) = \mu/(1 - \beta) + \sum_{i=0}^{\infty} \beta^i \epsilon_{t-i}$  as  $\beta^i \rightarrow 0$  as  $t \rightarrow \infty$ . A process in which past shocks do not accumulate is said to be integrated of order zero, denoted  $I(0)$ . An important and more general form of autoregressive process is the vector autoregression of order  $m$  ( $\text{VAR}(m)$ ), which takes the form  $\mathbf{y}_t = \boldsymbol{\mu} + \sum_{i=1}^m \mathbf{A}_i \mathbf{y}_{t-i} + \boldsymbol{\epsilon}_t$  when  $\mathbf{y}_t$ ,  $\boldsymbol{\mu}$  and  $\boldsymbol{\epsilon}_t$  are  $k$ th order vectors and the  $\mathbf{A}_i$  are  $k \times k$  matrices. Considering only the case in which  $m = 1$ , since higher order processes can always be reduced to first order using the companion form (see e.g. Hendry, 1995a, p. 724), the VAR is stationary when the eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_k$  of  $|\lambda \mathbf{I}_k - \mathbf{A}| = 0$  lie inside the unit circle (see e.g. Johansen, 1995, p. 14). When this condition is satisfied, the VAR consists of  $k$   $I(0)$  processes.

However, stationarity and  $I(0)$  processes are the exception, non-stationarity is the norm. What we observe is that, as well as evolving, time series processes are greatly influenced by specific events, including key discoveries like vaccination and antibiotics; inventions like the steam engine and dynamo; major wars, pandemics and massive volcanic eruptions; financial innovations, etc., all of which can cause persistent shifts in the means and variances of the data, thereby violating stationarity. Processes in which the effects of shocks persist are therefore common, and are said to be integrated of order greater than zero. For example,  $y_t = \mu + y_{t-1} + \epsilon_t$  can be written after successive substitution for lagged  $y$ s as  $y_t = y_0 + \mu t + \sum_{i=0}^{t-1} \epsilon_{t-i}$  revealing that the shocks  $\epsilon_{t-i}$  accumulate. Indeed, neither the mean  $E[y_t] = E[y_0] + \mu t$ , nor the variance  $V[y_t] = V[y_0] + t\sigma_\epsilon^2$  is constant. This



is an example of I(1) processes that are often observed in practice, an example of which in economics is the stock of a variable such as an inventory that cumulates its net inflow. Thus, unlike an I(0) process, which varies around a constant mean, an I(1) process has an increasing variance, usually called a stochastic trend, and may “drift” in a general direction over time to induce an actual trend when  $\mu \neq 0$ . Perhaps the best known example of an I(1) process is a random walk, first proposed by Bachelier (1900) to describe the behaviour of prices set in speculative markets. Another feature of an I(1) process is that since successive observations share a large number of past inputs, the correlation between them will be high and only decline slowly as their distance apart increases. Not only will the serial correlation coefficients  $\gamma_p = E[(y_t - E[y_t])(y_{t-p} - E[y_{t-p}])]$  remain high, only declining very slowly with  $p$ , but also there can be a high correlation between different I(1) variables that should be unrelated. This is known as the “nonsense correlation” problem first identified by Yule (1926), and illustrated by Hendry (1980) who created an example between the price level in the UK and cumulative annual rainfall. Granger and Newbold (1974) emphasized that a supposedly “significant relation” between variables, but where there was serial correlation in the residuals from that relation, was a symptom associated with nonsense regressions. Phillips (1986) provided a technical analysis of the sources and symptoms of nonsense regressions. Noting that differencing is the opposite of integration suggests that differencing an I(1) variable will render it I(0), and this is indeed the case as transforming the I(1) process  $y_t = \mu + y_{t-1} + \epsilon_t$  into the I(0) process  $\Delta y_t = \mu + \epsilon_t$  illustrates. This idea underlies the approach in Box and Jenkins (1970/1976) which was very popular in the 1970s and early 1980s in economics as well as other disciplines.

Linear combinations of several I(1) processes are usually I(1) as well, which led to some researchers modelling variables in differences rather than levels. Were it the case that relationships between I(1) variables could only be developed in their differences, it would imply that there could be no stable economic equilibrium relationships between I(1) variables. However, stochastic trends can cancel between series to yield an I(0) outcome, and this is called cointegration (Engle & Granger, 1987). Consider the first-order autoregressive-distributed lag model  $y_t = a_1 y_{t-1} + b_0 z_t + b_1 z_{t-1} + u_t$  when both  $y_t$  and  $z_t$  are I(1) variables with  $|a_1| < 1$ . Then, the re-parameterized model  $\Delta y_t = b_0 \Delta z_t + \alpha(y_{t-1} - \beta z_{t-1}) + u_t$ , where  $\alpha = (a_1 - 1)$  and  $\beta = (b_0 + b_1)/(1 - a_1)$ , will consist entirely of I(0) variables if  $(y_{t-1} - \beta z_{t-1})$  is I(0), and thus forms a cointegrating relationship. In economics, integrated-cointegrated data seem almost inevitable because of the Granger (1981) Representation Theorem which shows that cointegration between variables must occur if there are fewer decision variables (e.g. your income and bank account balance) than the number of decisions (e.g. hundreds of shopping items: see Hendry, 2004, for an explanation). Cointegrated relationships define a “long-run equilibrium trajectory” for the economy, departures from which induce “equilibrium correction” that move the economy back towards that path. Prior to Granger (1981) and Engle and Granger (1987) defining and developing the concept of cointegration, Davidson et al. (1978) had been using what they called “error correction” models which had essentially the same characteristics as the cointegration “equilibrium correction” models. A model that has played an important role in the modelling of econometric time series since the publication of Engle and Granger (1987), and especially the subsequent more detailed statistical analysis of cointegrated systems including a test of the order of cointegration in Johansen (1988, 1995), is the vector equilibrium correction model (VEqCM) given by  $\Delta \mathbf{y}_t = \boldsymbol{\mu} + \sum_{i=1}^{m-1} \mathbf{A}_i \Delta \mathbf{y}_{t-i} + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{y}_{t-m} + \boldsymbol{\epsilon}_t$  when  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  are  $k \times r$  matrices of rank  $r$  and  $\boldsymbol{\beta}' \mathbf{y}_{t-m}$  are  $r$  I(0) cointegrating vectors. This reveals that modelling only in differences to take account of the I(1) non-stationarity in  $\mathbf{y}_t$  ignores important levels information in  $\boldsymbol{\beta}' \mathbf{y}_{t-1}$  and so is inefficient. PcGive and CATS in RATS (see Hansen & Juselius, 1995) provide full implementations of the statistical analysis of VEqCMs. Hendry and Juselius (2000, 2001) provide surveys of the literature.

If the only source of non-stationarity were the presence of I( $q$ ) processes with  $q = 1$  or 2, then a combination of differencing and cointegrating relationships would bring the analysis back to I(0) processes. Other sources of non-stationarity also matter, however, especially shifts in the means of data distributions of I(0) variables, including equilibrium corrections and growth rates. There is a tendency in the econometrics literature to identify “non-stationarity” with integrated data (unit roots), and so incorrectly claim that differencing a time series induces stationarity. There are many

other sources of non-stationarity, so we refer to wide-sense non-stationarity to include both stochastic trends and location shifts, the combination of which causes numerous problems for econometric modelling.

In the VEqCM above, a location shift must occur when  $\mu$  changes with other parameters constant, or those parameters shift with  $\mu$  constant. Failure to model, or remove, such shifts can have a pernicious effect on the quality of an estimated model, as shown in Castle and Hendry (2014a). Moreover, as Hendry and Mizon (2014) demonstrate, inter-temporal economic theory fails when unanticipated location shifts occur, with the law of iterated expectations no longer applying, and “rational expectations” being biased. Fortunately for empirical modelling, SIS provides an automatic selection method to detect and “neutralize” location shifts in-sample. Also, analogous to cointegration cancelling unit roots to deliver an  $I(0)$  relation, co-breaking can cancel location shifts in linear combinations of variables (see Hendry & Massmann, 2007). Such an occurrence suggests a tight connection between the variables involved.

Stochastic trends and location shifts in economic time series can also adversely affect forecast accuracy. The methods used in practical forecasting have to rely on currently available information about the past and present, to extrapolate into the future. Even if the analysis of the available information and the representation of it in models is exemplary, accurate and reliable, forecasting requires that the future resembles the present in its essential attributes. Unfortunately, intermittent unanticipated shifts entail that this is rarely true. Though attempts have been made to predict future shifts (see Castle, Fawcett, & Hendry, 2010, 2011), that still remains an important research agenda item. While the most parsimonious, congruent and encompassing model in-sample usually would dominate in forecasting out of sample if there were no location shifts, such models have to be made robust to location shifts, which leads to a different class of model, and one that need not even be congruent in-sample. Thus, while automatic Gets aims to locate the LDGP, doing so successfully need not improve forecasting in the face of unanticipated location shifts. However, a congruent encompassing model, although it may require robustification in order to forecast accurately, can still form a useful basis for doing so and help retain valuable causal information.

Further,  $I(1)$  processes lead to much higher forecast uncertainty using the correct in-sample model, compared to even mis-specified models on  $I(0)$  data; and models with deterministic linear trends on either data type seriously understate the correct uncertainty. Indeed, the poor record of econometric forecasts as compared with (say) the time series models of Box and Jenkins (1970/1976) led to the realization that it is important to robustify forecasting models by exploiting the fact that location shifts, or shifts in time trends, can be reduced to impulses by an appropriate order of differencing. Clements and Hendry (1998, 1999) provide extensive discussions.

### 2.6.1. *Explaining the basics of non-stationarity*

Hendry (2015) stresses that much of the historical variation in economic time series has been due to “non-economic” factors such as changes in social mores, legislation, technology, medicine and finance as well as wars, only partly influenced by economic variables like prices and incomes. Since change is the norm, and that book is aimed at teaching undergraduates, it offers simple explanations for unit-root non-stationarity, cointegration, location shifts and co-breaking, to which the reader is referred.

### 2.7. *Non-linearity*

So far, all the DGPs and models considered have been linear in the variables, albeit we have generally assumed that holds after log-transforms of the basic aggregate measures. There are also many models that are non-linear in parameters, such as threshold (see Teräsvirta, Tjøstheim, & Granger, 2011), and regime-switching models (see e.g. Hamilton, 2015). Non-linear in variables relations cannot be excluded *a priori*, but are everything else, so comprise an infinite number of possibilities, thereby posing an impossible modelling task. To cut that Gordian knot, Castle and Hendry (2010, 2014b) propose a low-dimensional approach, using squares, cubes and exponential functions of the

individual elements in the principal components of all  $N$  original variables in the GUM. Denoting those components by  $v_{i,t}$ , they add  $v_{i,t}^2$ ,  $v_{i,t}^3$  and  $v_{i,t} \exp(|-v_{i,t}|)$  to the set of candidate variables, in order to capture the most important sources of departure from linearity, including asymmetry and sign-preserving reactions, using “only”  $3N$  more variables.

A valuable advantage is that there is no collinearity between elements of  $v_{i,t}$ , and by demeaning the higher order terms, little between those either. A drawback is the difficulty of interpreting any non-linearities discovered, but whenever a preferred theory specification is available, such as a logistic smooth transition formulation, an encompassing test against that is easy to conduct. The outcome could reveal that the preferred model accounts for all the non-linearity captured by the low-dimensional approach, or is significant but some non-linearity remains, or is insignificant so is not the correct non-linear specification.

However, to tackle the vast numbers of candidate variables in GUMs with many variables, long lags on those, the non-linear components just described, and IIS and/or SIS, so there are many more variables than observations, a powerful selection tool is needed, the topic to which we now turn.

### 2.7.1. Explaining the basics of non-linearity

Few economics undergraduate econometrics courses tackle either principal components or selection in non-linear in variables models, but often include an explanation of the RESET test (see Ramsey, 1969). That test adds the square, or sometimes also the cube, of  $\hat{y}_t$  to the regression, which creates a non-linear function of a linear combination of the regressors weighted by their estimated coefficients. Here, we are adding non-linear functions of linear combinations of the regressors, weighted by their importance in explaining their overall variance. To illustrate, one route would be to reuse the data in Section 2.1.1, and add, say,  $x_{1,t}^2$  to the general regression to show that it is insignificant—as there is no non-linear connection. Next, find the four largest values of  $x_{1,t}$  in the data-set, and add sufficiently large impulses to  $y_t$  at those dates to create outliers that should now be “modelled” by a spuriously significant coefficient for  $x_{1,t}^2$ . Finally, add impulse dummies for those dates to demonstrate that the correct relation can be recovered (or use IIS if that is available). There are also non-parametric approaches: if these are available, it can be fun to show how they behave in this setting.

### 2.8. Model selection

When modelling economic or social systems, it is impossible to capture everything that matters empirically, so we focus on influences that “matter substantively”, albeit that must be context and sample size dependent. The previous sections have explained the framework and concepts that have led us to seek congruent, parsimonious encompassing representations, obtained by simplifying an initial general unrestricted model, or GUM, that captures the main data properties, such as autocorrelation, non-stationarity and regime shifts. In wide-sense non-stationary processes, *ceteris paribus* cannot apply empirically, so commencing with too few variables in the candidate set may make it impossible to find a constant parameter model. Undertaking selection from many variables rapidly exceeds what any human can achieve, so automatic selection methods have become essential for successful econometric modelling. They are capable of investigating empirically a much wider range of possibilities than even the greatest experts, and of doing so efficiently when the automatic searches are well structured.

Doornik (2009) explains how “general-to-specific” selection algorithms operate, of which *Autometrics* is the latest version. This uses block multi-path searches in a tree structure, essentially classifying effects into those that are significant given all other selected and retained variables, and those that are not. This approach allows *Autometrics* to select models even when there are more candidate variables than observations as discussed in Doornik and Hendry (2015). To maintain congruence, diagnostic testing is undertaken throughout the simplification process, as well as checking encompassing of the (local) GUM by each terminal model, backtracking to an earlier, less simple, model if any tests reject.

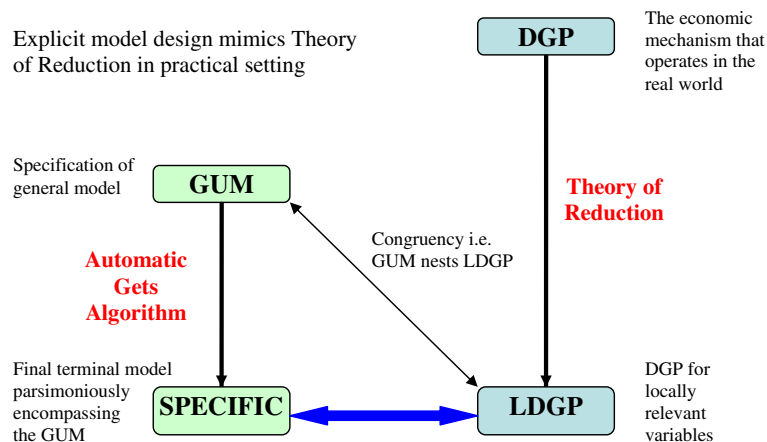
The advantages of automatic methods are described in Hendry and Doornik (2014), so here we will consider model selection when a set of variables suggested by a prior theory are to be retained without selection. Section 2.1 distinguished between the **target** of model selection, which has to be the LDGP, and the **object** of the analysis, usually an economic theory model. A natural reconciliation is to nest “data-driven” and “theory-driven” approaches in a common framework, where the theory model is retained, but not imposed, and a wide range of influences that potentially could matter are selected over. To ensure the theory model parameter estimates that result have exactly the same distribution as when a complete and correct theory is fitted directly to the same data, prior to selection, Hendry and Johansen (2015) regress all the other candidate variables on the theory variables and replace the former by the resulting residuals which are thereby orthogonal to the theory variables. As is well known from generalizations of the famous theorem in Frisch and Waugh (1933), parameter estimates are unaffected by the inclusion or exclusion of orthogonal regressors, all of which would be irrelevant when the theory model was complete and correct. In the more likely setting that the theory is incomplete, a better representation of the LDGP can be discovered, so in both states of nature, such an approach is either costless or beneficial.

### 2.8.1. Explaining the basics of model selection

If you undertook the example in Section 2.1, or used *Autometrics* at any other stage, then you have already done model selection, and presumably explained the steps involved. Thus, this part ends where it began, highlighting a perennial problem for teaching: all the concepts are closely inter-related. This usually leads to a simple to general approach in teaching, and like all such methods, the stopping point can be arbitrary. The resulting danger is leaving students with a seriously naive view of econometric modelling when they only study at an elementary level. This paper, Hendry and Nielsen (2007,2010) and Hendry (2015) attempt to convey the complexities of real-world economic time series, and provide exciting tools to build models that at least avoid the most egregious mistakes.

In teaching, we often use Figure 1 from Hendry and Doornik (2014) as a summary of all the stages above. Starting at the top right with the DGP, which is bound to be unknown however good the accompanying economic analysis, the reductions lead down to the LDGP for the variables to be modelled—the topic of Section 2.1. Moving to the upper left-hand side of the diagram, the GUM must be specified sufficiently generally to nest the LDGP and embed the theory-based variables, so is congruent in order to sustain valid inferences during selection, as discussed in Section 2.2, and for conditional models in Section 2.4. While this is ideal, often a GUM may not be sufficiently general to nest the LDGP, in which case an approximation will result, as discussed in Castle, Doornik, and Hendry (2011). In practice, with more variables than observations (e.g. when IIS or SIS are used, as discussed in Section 2.6), congruence can only be checked after some simplification. Inevitably, the GUM will contain some redundant candidate variables and indicators, so general-to-specific model selection (considered in Section 2.3) is

**Figure 1. Explaining the steps from the DGP to a specific congruent encompassing model of the LDGP.**



used to find a congruent, parsimonious encompassing representation in a final specific model, the subject of Section 2.5 and Section 2.8, possibly requiring checks of the linearity Section 2.7. When the GUM nests the LDGP, and the final model parsimoniously encompasses the GUM, then it should also parsimoniously encompass the unknown LDGP. Hence, the search has discovered what actually matters, and the researcher can legitimately evaluate the theory model.

### 2.9. Econometric software

We now present a very brief overview of some of the computer software that has been used in the teaching of econometrics since the late 1970s: Renfro (2009) provides an extensive history of econometric computing. Early mainframe computer programmes for the analysis of single equation time series models that could illustrate empirical work for teaching included TSP (see Hall & Cummins, 2005, for a recent release), MODLER (see Renfro, 1996, for a retrospective) and *Give* (see Hendry & Srba, 1980), providing estimation methods for models with endogenous explanatory variables via two and three stage least squares (TSLS and 3SLS) and instrumental variables (IV). *Fiml*, a companion to *Give* implemented full information maximum likelihood estimation of systems of simultaneous equations based on Hendry (1976). A feature of *Give* and *Fiml* was that they also incorporated mis-specification test statistics as these were developed and shown to be effective in model evaluation, particularly those related to the testing of concepts described in Section 2. With the advent of PCs, *PcGive* was developed for this more flexible medium, initially to complement *Give* but eventually to supersede it and *Fiml*. *MicroTSP* (initially developed by David M. Lilien) was introduced as the PC version of TSP, later integrated into *EViews* (see QMS, 2005). Among others, *Microfit* (see Pesaran & Pesaran, 1987) and RATS (see Enders, 1996) extended the available range, the latter more so after the development of CATS in RATS implemented multivariate cointegration analysis (see Hansen & Juselius, 1995). At a more basic level, STATA is often used in undergraduate courses, as are spreadsheets (the use of which is likely to lead to serious errors); whereas at a professional level the R language is popular, as is *Ox*.

The evolution of the *PcGive* software discussed in Hendry and Doornik (1999) can be traced via the many editions of the accompanying manuals beginning with Hendry (1984) through to the latest (Doornik & Hendry, 2013a, 2013b) which are part of the *OxMetrics* suite and incorporate *Autometrics*. Previous publications by us on the teaching of econometrics using *PcGive* and the *OxMetrics* suite of programmes include Hendry (1986, 1990) and Hendry and Nielsen (2007, 2010).

### 3. Illustrating new econometrics teaching

Hendry (2015) proposes a major change in the curriculum for undergraduate econometrics, to include a range of topics essential for understanding and undertaking empirical research, explicitly addressing the need to **discover** what influences actually matter in practice:

... the notion of empirical model discovery in economics may seem to be an unlikely idea, but it is a natural evolution from existing practices. Despite the paucity of explicit research on model discovery, there are large literatures on closely related approaches, including model evaluation (implicitly discovering what is wrong); robust statistics (discovering which sub-sample is reliable); non-parametric methods (discovering the relevant functional form); identifying time series models (discovering which model in a well-defined class best characterizes the available data); model selection (discovering which model best satisfies the given criteria), but rarely framed as discovery.

Model selection in the face of changing economies is at the centre of Hendry (2015), using a UK macroeconomics database over 1860–2011. Not only do economies evolve, so does economic theory, making it hazardous to **impose** extant theory on empirical models if that theory might be discarded shortly. As outlined in Section 2.8, rather than adopt either a “theory-driven” or a “data-driven” approach, empirical model discovery embeds the best available theory to be retained during selection while investigating many other potentially relevant variables, longer lags, non-linear functions and both outliers and location shifts. When the theory is complete and correct for the sample under analysis, the distributions of the parameter estimates will be identical to those obtained by directly fitting the theory to the data: see Hendry and Johansen (2015). However, if the theory is incorrect, or



more usually, incomplete, but the extended specification includes substantively relevant variables, an improved representation will result. As illustrated in Hendry and Mizon (2011), failing to correctly determine dynamics and outliers can lead to a model that is so seriously mis-specified the empirical results appear to reject the theory from which the model was derived, yet after taking account of those effects, is strongly consistent with the same theory. When systematically conducted, model discovery—or “data mining” as it is sometimes pejoratively called—can improve theory-based specifications. Think of undertaking extensive explorations while controlling for adventitious significance as answering all likely seminar questions in advance.

Recent advances in computer power and speed, and improvements in search algorithms, facilitate a modified general-to-specific modelling strategy even if the initial number of candidate explanatory variables,  $N$ , exceeds the available number of observations,  $T$ . Thus, we now provide an example of how one might teach a systematic approach to undertaking empirical time series research, albeit simplified to modelling a single variable dependent on a few explanatory variables using artificial data. The simplicity is to sustain live demonstrations and class participation, either directly with each student undertaking their own data generation, modelling, then simulation, or enabling questions to be addressed by showing their impact on the instructor’s models.

The aim of the following sections is to illustrate the roles of the various components discussed above. Subject-matter theory is discussed in Section 3.1, the database and software in Section 3.2, computing the estimates of the DGP parameters and testing congruence in Section 3.3, the formulation of the general unrestricted model (GUM, although we will eschew orthogonalization here) and selection with indicator saturation in Section 3.4, then testing parameter constancy in Section 3.5 and exogeneity in Section 3.6. Parsimonious encompassing is implemented automatically during selection, and we will briefly note testing for a non-linear representation in Section 3.7. Finally, the use of simulation to investigate the outcome will be described in Section 3.8.

### 3.1. The economic theory

We consider a mimic of a demand model for a perishable commodity like fish (see e.g. Graddy, 2006, and the use of her data in Hendry & Nielsen, 2007). Let  $p_t$  denote the price of the specific variety of fish in the market available in a quantity  $q_t$  where lower case letters denote the logs of the variables. The available theory suggests:

$$p_t = \beta_0 + \beta_1 q_t + \beta_2 p_{t-1} + \epsilon_t \quad (5)$$

where  $\epsilon_t \sim \text{IN}[0, \sigma_\epsilon^2]$ ,  $\beta_1 < 0$  and  $\beta_2 > 0$  with  $|\beta_2| < 1$ . Because  $q_t$  is the given volume of fish landed, it will be treated as weakly exogenous, determined by:

$$q_t = \gamma_0 + \gamma_1 q_{t-1} + \gamma_2 p_{t-1} + \gamma_3 W_t + v_t \quad (6)$$

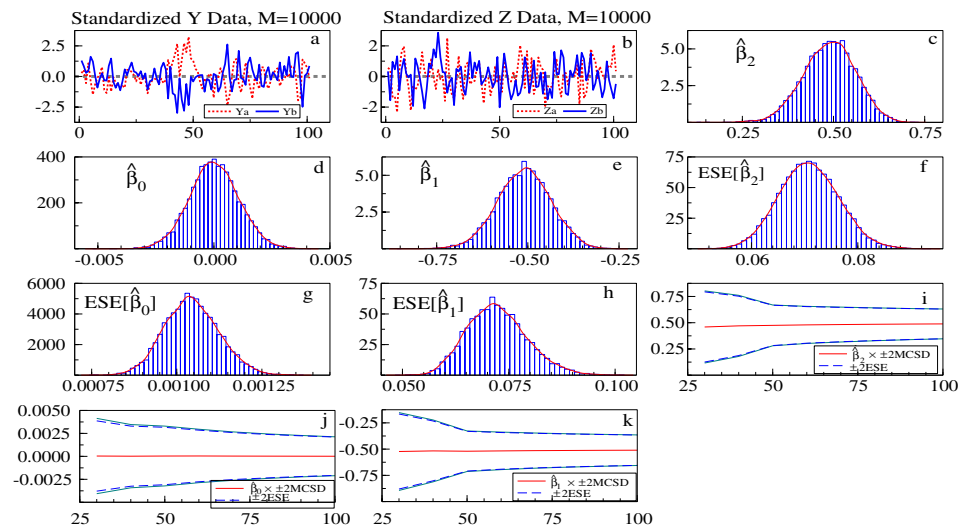
where  $v_t \sim \text{IN}[0, \sigma_v^2]$ ,  $|\gamma_1| < 1$ ,  $\gamma_2 > 0$  and  $W_t$  denotes weather variables (storms etc., acting as exogenous shocks).

### 3.2. The database and software

The database here has two components. First, the model in Hendry and Nielsen (2007) of the Fulton Fish Market time series data collected by Graddy (1995) on the daily prices and quantities of whiting sold by a wholesaler from 2 December 1991 to 8 May 1992, and associated weather-related measures sets the scene for creating a simulated data-set.<sup>3</sup> Second, students are given the task of generating an artificial data-set to mimic such a market. The reason for the artificial data is that the closeness of any claimed model to its DGP can be judged, whereas there are no known “correct answers” with empirical data.

To generate the artificial data, we use the *PcNaive* module within *PcGive*: see Doornik & Hendry, 2013b, and also the explanations for its use in Hendry (2015, Chap. 8.10). The role of *PcNaive* is to ease the design of simulation experiments, and its output is a computer programme in Ox, which

**Figure 2. Data, parameter estimation distributions for the  $Yb_t$  model, and recursive estimates.**



can be run by Ox Professional. In *PcGive*, select Monte Carlo, Advanced Experiment option, and create a DGP with two endogenous and two exogenous variables, with a break from observations 40–50, which we use to mimic a prolonged period of stormy weather that reduces the supply of fish.<sup>4</sup> Choose the “simultaneous equations” formulation and create the two equations (5) and (6), using the parameter values  $\beta_1 = -0.5$  and  $\beta_2 = 0.5$  (which deliver a long-run price elasticity of minus unity), and  $\gamma_1 = 0.4$ ,  $\gamma_2 = 0.4$  and  $\gamma_3 = 0.025$ . The intercepts matter in reality, but we will set them to zero here, an effect that could be achieved by appropriate choices of units. Finally, set  $\sigma_\epsilon = 0.01$  and  $\sigma_v = 0.01$ , again dependent on units, but in a log-linear model represent error standard deviations of 1%, so the storm is  $2.5\sigma_v$ . Set  $T = 100$ , and select “save data” so the final simulation trial can be analysed as if it were empirical data. We used  $M = 10,000$  replications for Figure 2, but only a few replications are needed.

Conceptually labelling the panels *a*–*k* along successive rows, then *a* & *b*, record the standardized data time series denoted  $Ya_t$  and  $Yb_t$ , and  $Za_t$  &  $Zb_t$  *c*–*e* show the sampling distributions of the estimators of  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ , then *f*–*h* the distributions of their conventionally computed estimated standard errors (denoted ESE), and finally *i*–*k* show the means over the  $M$  replications of the parameter estimates, with  $\pm 2\text{ESE}$  and  $\pm 2\text{MCSD}$  (the “true” standard errors based on the distributions shown in *c*–*e*). The wide scatter of the possible estimates of the parameters is noticeable, as is the potential variability in their ESEs.

Every student can be given a different draw by changing the number of replications each is assigned. We will return later to do a multi-replication study to compare the estimation of the DGP model with that resulting from selection from a much larger GUM. The role of the two exogenous variables will be to create additional irrelevant variables.

### 3.3. Computing the estimates of the DGP parameters

Load the created data-set back into *PcGive*, and use the calculator to form a dummy variable *StormDum* equal to unity from observation 41 to 51 and zero elsewhere (the date shift is due to how *PcNaive* times events). First check that the data graphs match the general form of those shown in Figure 2, then estimate the two DGP Equations (5) and (6). Here, we found:

$$Ya_t = - \begin{matrix} 0.0013 \\ (0.00096) \end{matrix} - \begin{matrix} 0.64 \\ (0.078) \end{matrix} Yb_t + \begin{matrix} 0.49 \\ (0.068) \end{matrix} Ya_{t-1}$$

$$\hat{\sigma}_\epsilon = 0.0093; R^2 = 0.554; F_{ar}(2, 95) = 0.17; F_{arch}(1, 98) = 0.001 \tag{7}$$

$$\chi_{nd}^2(2) = 3.25; F_{het}(4, 95) = 0.33; F_{reset}(2, 95) = 1.75$$

and:

$$Yb_t = 0.0002 + 0.32 Yb_{t-1} + 0.47 Ya_{t-1} - 0.024 \text{ StormDum}_t$$

(0.0011)            (0.097)            (0.095)            (0.0039)

$$\hat{\sigma}_v = 0.0097; R^2 = 0.376; F_{ar}(2, 94) = 3.25^*; F_{arch}(1, 98) = 1.90$$

$$\chi_{nd}^2(2) = 0.71; F_{het}(5, 94) = 0.68; F_{reset}(2, 94) = 1.06$$

In (7) and (8),  $\hat{\sigma}$  is the residual standard deviation, and  $R^2$  is the squared multiple correlation, with coefficient standard errors shown in parentheses. The mis-specification test statistics have the form of  $F_j(k, T - l)$ , denoting an approximate F-test against the alternative hypothesis  $j$  and comprise:  $k$ th-order serial correlation ( $F_{ar}$ : see Godfrey, 1978);  $k$ th-order autoregressive conditional heteroskedasticity ( $F_{arch}$ : ARCH, see Engle, 1982); heteroskedasticity ( $F_{het}$ : see White, 1980);  $F_{reset}$  which is the RESET test (see Ramsey, 1969); and a chi-square test for normality ( $\chi_{nd}^2(2)$ , see Doornik & Hansen, 2008).

Parsimonious encompassing of the feasible GUM will be checked during selection. Parameter constancy over  $k$  periods ( $F_{Chow}$ : Chow, see Chow, 1960), super exogeneity ( $F_{SupExog}$  based on IIS: see Hendry and Santos (2010)), and the low-dimensional test for non-linearity ( $F_{NL}$ : see Castle & Hendry, 2010) could be added as discussed in Sections 3.5–3.7. These estimates are recognizably close to the DGP parameter values used, with no mis-specification tests significant by chance at the 1% level, so there are no important departures from congruent representations.

### 3.4. Empirical model discovery

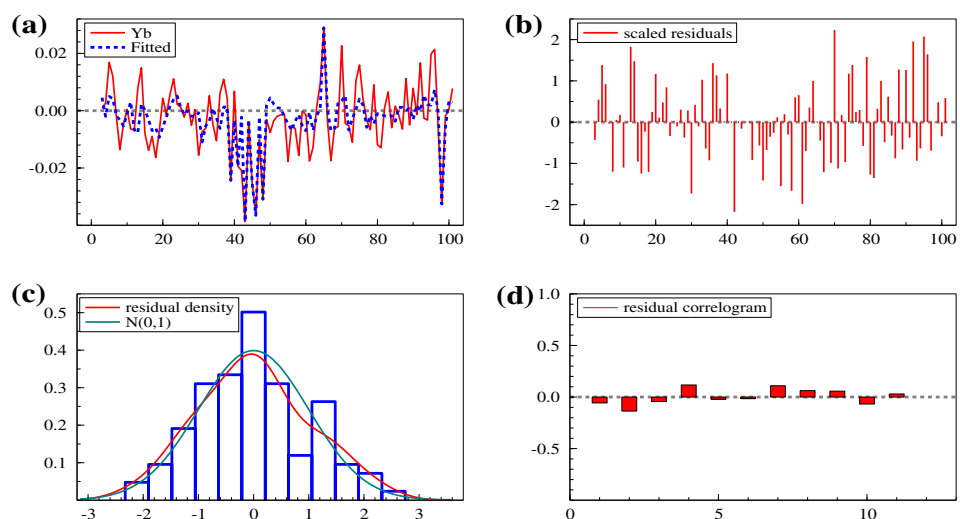
We now assume that the investigator is unsure of the validity of the theory model, the weak exogeneity status of  $Yb_t$  in the  $Ya_t$  model, the possibility of location shifts or outliers, and does not know the precise timing of the bad weather, which is correctly represented by  $\text{StormDum}$  in Equation (8). Thus, she specifies much more general initial models than those in Equations (7) and (8), including two additional variables denoted  $Za_t$  and  $Zb_t$ , both  $IN[0, 1]$ , all variables entering equations having lags of two periods, and using impulse-indicator saturation (IIS: see Hendry et al., 2008; Johansen & Nielsen, 2009). First:

$$Ya_t = \beta_0 + \beta_1 Yb_t + \beta_2 Ya_{t-1} + \beta_3 Yb_{t-1} + \beta_4 Ya_{t-2} + \beta_5 Yb_{t-2} + \beta_6 Za_t$$

$$+ \beta_7 Za_{t-1} + \beta_8 Za_{t-2} + \beta_9 Zb_t + \beta_{10} Zb_{t-1} + \beta_{11} Zb_{t-2} + \sum_{i=1}^T \delta_i 1_{(i=t)} + e_t$$

where  $e_t \sim IN[0, \sigma_e^2]$  and  $1_{(i=t)}$  denotes an indicator with the value zero except for unity when  $i = t$ . Similarly:

**Figure 3. Fitted and actual values for  $Yb_t$  after selecting from its GUM.**



$$Yb_t = \gamma_0 + \gamma_1 Ya_{t-1} + \gamma_2 Yb_{t-1} + \gamma_3 Ya_{t-2} + \gamma_4 Yb_{t-2} + \gamma_5 Za_t + \gamma_6 Za_{t-1} + \gamma_7 Za_{t-2} + \gamma_8 Zb_t + \gamma_8 Zb_{t-1} + \gamma_{10} Zb_{t-2} + \sum_{i=1}^T \kappa_i 1_{\{i=t\}} + u_t \quad (10)$$

where  $u_t \sim \text{IN}[0, \sigma_u^2]$ . Notice that (10) does not include any weather variables ( $W_t$ ), which we will try to capture by IIS.

Both GUMs have more regressors than  $T$ , but this does not pose any problems for an automatic model-selection approach like that in *Autometrics*, as explained in Hendry and Doornik (2014) and Doornik and Hendry (2015). The theory in Hendry and Johansen (2015) proposes orthogonalizing all the additional regressors against the theory-model variables, so the latter are not “contaminated” by selection, but here we will simply retain them and the intercept during selection. For the  $Ya_t$  GUM in (9), selection at 1% exactly reproduces (7) despite all the added irrelevant variables. This is slightly lucky, since with 110 candidate variables in total, on average one should be significant by chance.

For the  $Yb_t$  GUM in (10), selection at 1% finds:

$$Yb_t = - \underset{(0.00092)}{0.0005} + \underset{(0.08)}{0.29} Ya_{t-1} + \underset{(0.09)}{0.35} Yb_{t-1} - \underset{(0.0086)}{0.025} 1_{39} - \underset{(0.0087)}{0.024} 1_{41} - \underset{(0.0088)}{0.038} 1_{43} - \underset{(0.0086)}{0.024} 1_{45} - \underset{(0.0088)}{0.033} 1_{46} - \underset{(0.0088)}{0.035} 1_{48} + \underset{(0.0087)}{0.029} 1_{65} - \underset{(0.0089)}{0.024} 1_{98} \quad (11)$$

$$\hat{\sigma}_v = 0.0086; R^2 = 0.550; F_{ar}(2, 86) = 1.67; F_{arch}(1, 97) = 0.22$$

$$\chi_{nd}^2(2) = 0.70; F_{het}(4, 86) = 1.39; F_{reset}(2, 86) = 1.72$$

The storm from 41–51 is approximated by five similar magnitude same sign impulses thereby missing some of its intermediate effects but “picking up” an earlier start. When the shift is just  $2.5\sigma$ , a positive shock is likely to make an impulse indicator less significant than the 1% critical value of 2.6. Thus, finding 5 is the average probability. In addition, IIS finds outliers at observations 65 and 98, which are clearly visible in Figure 3(a). On average, roughly one should be significant by chance, but the missing storm impulses have somewhat biased the regression estimates, which may have created spurious outliers (as will transpire to be the case).

None of the irrelevant regressors was retained. Equation (11) slightly overfits when all the impulse indicators are entered freely, although that can be mitigated by a bias correction (see e.g. Johansen & Nielsen, 2009). Imposing a common coefficient across the contiguous indicators found, but from 39 to 48, leads to an almost identical outcome as using StormDum. However,  $1_{98}$  now becomes insignificant at 1%, and dropping that leads to  $\hat{\sigma}_v = 0.0093$ :

$$Yb_t = - \underset{(0.001)}{0.0002} + \underset{(0.092)}{0.47} Ya_{t-1} + \underset{(0.095)}{0.30} Yb_{t-1} - \underset{(0.004)}{0.024} \text{StormDum}_t - \underset{(0.0094)}{0.030} 1_{65}$$

$$\hat{\sigma}_v = 0.0093; R^2 = 0.43; F_{ar}(2, 92) = 3.45^*; F_{arch}(1, 97) = 1.75$$

$$\chi_{nd}^2(2) = 2.43; F_{het}(5, 92) = 1.39; F_{reset}(2, 92) = 1.42 \quad (12)$$

Thus, eliminating  $1_{98}$  creates some residual autocorrelation at 5%: such “trade-offs” between keeping insignificant variables and congruence often occur in empirical research. Notice that all the theory-based variables have been selected in Equation (12), so the same results would be delivered when those were retained.

To summarize, despite a lack of knowledge of dynamic reactions, relevant variables, location shifts or outliers, so the GUM had 110 variables for 98 observations (after lags), only one irrelevant effect, namely  $1_{65}$ , was significant by chance which is what one would anticipate at 1%. Inspecting the residuals from (8) would have shown the same outlier, and applying IIS to that equation would also have revealed that  $1_{39}$  was significant and reduced the significant (spurious) residual autocorrelation.

### 3.5. Testing parameter constancy

Like exogeneity in the next section, testing the constancy of a model's parameters usually only happens after modelling, but with indicator saturation, now can take place jointly with other selections. So far we have not applied step-indicator saturation (SIS), which uses increasing step indicators that essentially cumulate the corresponding impulses up to that time, but for a step shift like the simulated stormy period, is an effective device. For  $Yb_t$ , SIS yields:

$$Yb_t = \underset{(0.0015)}{0.0002} + \underset{(0.085)}{0.39} Ya_{t-1} + \underset{(0.096)}{0.24} Yb_{t-1} + \underset{(0.004)}{0.028} S_{40} - \underset{(0.004)}{0.022} S_{48} \\ - \underset{(0.010)}{0.035} S_{64} + \underset{(0.009)}{0.030} S_{65} \quad (13)$$

$$\hat{\sigma}_v = 0.0092; R^2 = 0.46; F_{ar}(2, 90) = 3.24^*; F_{arch}(1, 97) = 0.79 \\ \chi_{nd}^2(2) = 0.75; F_{het}(7, 90) = 0.86; F_{reset}(2, 90) = 1.06$$

The storm is captured from observation 40 to 48, and the outlier at 65 by the two offsetting steps: replacing them by  $1_{65}$  moves  $S_{40} \approx -S_{48}$  with  $\hat{\sigma}_v = 0.0093$ . So why does SIS not get the correct timing? The answer is because the shift is  $2.5\hat{\sigma}_v$  and we are selecting at 1% with a critical value of about  $c_\alpha = 2.6$ , positive draws can leave an apparent shift of less than  $c_\alpha$  so would not be selected. In effect, the storm does not show up in the data at that point, and indeed  $R^2$  is slightly lower in (12) with the correct dummy than in the variant of (13) using  $1_{65}$ ; in finite samples, the DGP need not be the “best” model, illustrating that modelling with a single sample of observations may capture a feature particular to that sample which is not part of the LDGP.

### 3.6. Testing exogeneity

The only contemporaneous regressor in the two models is  $Yb_t$  in (7), so that will be the focus of our test. Clearly, the conditional analysis was conducted under the assumption that  $Yb_t$  was weakly exogenous, as that hypothesis cannot be tested until the relevant equation has been established. The basis of our approach will be the automatic IIS test of super exogeneity, namely the joint hypothesis of constancy and weak exogeneity, proposed in Hendry and Santos (2010). This involves locating any shifts in the process of the conditional variable, here  $Yb_t$ , and testing their significance in the conditional model, here (7). Equation (11) revealed eight indicators in the former, so we test their inclusion in (7). This delivered  $F_{SupExog}(8, 89) = 1.26$ , which is insignificant, confirming the validity of conditioning on  $Yb_t$  in the model for  $Ya_t$ .

### 3.7. Testing for a non-linear representation

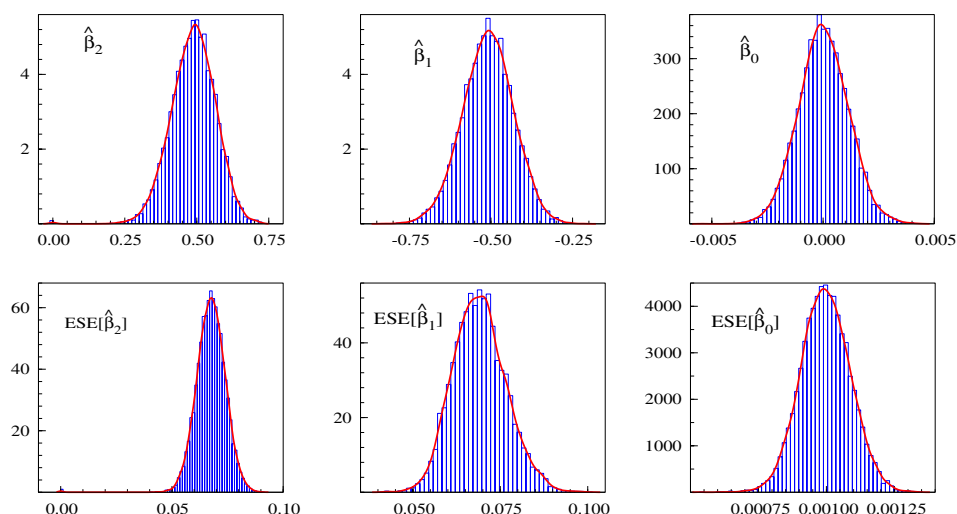
As with constancy tests, there are many available approaches, a number of which are offered within PcGive. Here, we calculate the low-dimensional test for non-linearity,  $F_{NL}$ , proposed by Castle and Hendry (2010) based on the squares, cubes and exponential functions of the principal components of the data series. These deliver  $F_{NL}(6, 90) = 1.50$  and  $F_{NL}(9, 83) = 1.06$  for (7) and (12), respectively, so neither reveals signs of non-linearity, which is the appropriate null outcome. An alternative approach we now prefer is to include those non-linear functions in the GUM and select at a tight significance level.

### 3.8. Re-simulating the model selection exercise

Since the final selections for  $Ya_t$  and  $Yb_t$  closely match their DGP equations, simulating (7) would deliver an outcome like that in Figure 2. However, simulating the GUM, or a version thereof selected



**Figure 4. Simulation distributions of parameter estimates and their estimated standard errors ESE[·] after selecting with IIS from the  $Y_{a_t}$  GUM.**



using IIS, could be worthwhile for comparison with the DGP outcome, although  $M = 10,000$  will take a considerable time on a PC. Note that the constant needs to be included as unrestricted for technical reasons, even though  $\beta_0 = 0$ , as does  $Y_{b_t}$  because the programme otherwise mishandles it as being endogenous. The *Autometrics* selection should be at 1%, which is approximately proportional to the total number of candidate variables,  $1/(T + N)$ . The chosen output should not include recursive estimates, nor need to include saving the final replication. Figure 4 records the resulting distributions of parameter estimates and their estimated standard errors for  $Y_{a_t}$ ; the output is essentially identical to Figure 2.

#### 4. Conclusions

The various manifestations of the KISS principle—keep it simple stupid—correctly emphasize parsimony, but fail to note that in his “razor”, William of Occam stressed avoiding adding **unnecessary** features. In a high-dimensional, possibly non-linear and wide-sense non-stationary world, facing both stochastic trends and distributional shifts, empirical models must be sufficiently general to capture all the substantive influences, or could end badly mis-specified. Teaching the empirical econometric analysis of large, complicated models can be demanding, but we have tried to steer a route through all the key steps, exploiting the amazing power of modern software. Not confronting the complications, and hence the need to discover what matters empirically while retaining the best available theory insights, will leave students with a dangerously naive view of how to model macro-economic time series, so we strongly advocate changing the curriculum to address all these issues. Astute readers will have noticed the gulf between the focus of our paper on concepts and model formulations, as against the usual textbook sequence of recipes for estimating pre-specified models. Appropriate estimation techniques are certainly necessary, but are far from sufficient if the model in question is not well specified. Since economic reality is complicated, pre-specification is unlikely to be perfect, so discovering a good model seems to be the only viable way ahead.

#### Acknowledgements

The authors are indebted to Jennifer L. Castle and Felix Pretis for helpful comments on an earlier version.

#### Funding

Financial support from the Institute for New Economic Thinking, Robertson Foundation, and Statistics Norway is gratefully acknowledged.

#### Author details

David F. Hendry<sup>1</sup>  
 E-mail: [david.hendry@nuffield.ox.ac.uk](mailto:david.hendry@nuffield.ox.ac.uk)  
 Grayham E. Mizon<sup>2,3</sup>

E-mail: [grayham.mizon@soton.ac.uk](mailto:grayham.mizon@soton.ac.uk)

<sup>1</sup> Economics Department and Institute for New Economic Thinking, Oxford Martin School, University of Oxford, Oxford, UK.

<sup>2</sup> Faculty of Social, Human, and Mathematical Sciences, University of Southampton, Southampton, UK.

<sup>3</sup> Economic Modelling Programme, Institute for New Economic Thinking, Oxford Martin School University of Oxford, Oxford, UK.

#### Citation information

Cite this article as: Improving the teaching of econometrics, David F. Hendry & Grayham E. Mizon, *Cogent Economics & Finance* (2016), 4: 1170096.

#### Note

1. Also see <http://www.voxeu.org/article/why-standard-macro-models-fail-crises>.
2. Available for free download at <http://www.timberlake.co.uk/intromacroeconometrics>, with teaching slides and a large UK macroeconomic database.
3. See Johansen and Neilsen (2016) for a recent re-analysis focusing on the possibility of outliers.
4. The PcNaive output in Ox code can be edited if desired for a more complicated representation.

#### References

- Anderson, T. W. (1962). The choice of the degree of a polynomial regression as a multiple-decision problem. *Annals of Mathematical Statistics*, 33, 255–265.
- Anderson, T. W. (1971). *The statistical analysis of time series*. New York, NY: Wiley.
- Bachelier, L. (1900). Théorie de la spéculation. *Annales Scientifiques de l'École Normale Supérieure*, 3, 21–86.
- Bontemps, C., & Mizon, G. E. (2003). Congruence and encompassing. In B. P. Stigum (Ed.), *Econometrics and the philosophy of economics* (pp. 354–378). Princeton, NJ: Princeton University Press.
- Bontemps, C., & Mizon, G. E. (2008). Encompassing: Concepts and implementation. *Oxford Bulletin of Economics and Statistics*, 70, 721–750.
- Box, G. E. P., & Jenkins, G. M. (1970/1976). *Time series analysis, forecasting and control*. San Francisco, CA: Holden-Day.
- Campos, J., Ericsson, N. R., & Hendry, D. F. (2005). *Readings on general-to-specific modeling*. Cheltenham: Edward Elgar.
- Castle, J. L., Doornik, J. A., & Hendry, D. F. (2011). Evaluating automatic model selection. *Journal of Time Series Econometrics*. doi:10.2202/1941-1928.1097
- Castle, J. L., Doornik, J. A., Hendry, D. F., & Pretis, F. (2015). Detecting location shifts during model selection by step-indicator saturation. *Econometrics*, 3, 240–264.
- Castle, J. L., Fawcett, N. W. P., & Hendry, D. F. (2010). Forecasting with equilibrium-correction models during structural breaks. *Journal of Econometrics*, 158, 25–36.
- Castle, J. L., Fawcett, N. W. P., & Hendry, D. F. (2011). Forecasting breaks and during breaks. In M. P. Clements & D. F. Hendry (Eds.), *Oxford handbook of economic forecasting* (pp. 315–352). Oxford: Oxford University Press.
- Castle, J. L., & Hendry, D. F. (2010). A low-dimension portmanteau test for non-linearity. *Journal of Econometrics*, 158, 231–245.
- Castle, J. L., & Hendry, D. F. (2011). Automatic selection of non-linear models. In L. Wang, H. Garnier, & T. Jackman (Eds.), *System identification, environmental modelling and control* (pp. 229–250). New York, NY: Springer.
- Castle, J. L., & Hendry, D. F. (2014a). Model selection in under-specified equations with breaks. *Journal of Econometrics*, 178, 286–293.
- Castle, J. L., & Hendry, D. F. (2014b). Semi-automatic non-linear model selection. In N. Haldrup, M. Meitz, & P. Saikkonen (Eds.), *Essays in nonlinear time series econometrics* (pp. 163–197). Oxford: Oxford University Press.
- Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28, 591–605.
- Clements, M. P., & Hendry, D. F. (1998). *Forecasting economic time series*. Cambridge: Cambridge University Press.
- Clements, M. P., & Hendry, D. F. (1999). *Forecasting non-stationary economic time series*. Cambridge, Mass: MIT Press.
- Cox, D. R. (1961). Tests of separate families of hypotheses. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 105–123). Berkeley: University of California Press.
- Cox, D. R. (1962). Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society B*, 24, 406–424.
- Davidson, J. E. H., & Hendry, D. F. (1981). Interpreting econometric evidence: The behaviour of consumers' expenditure in the UK. *European Economic Review*, 16, 177–192.
- Davidson, J. E. H., Hendry, D. F., Srba, F., & Yeo, J. S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal*, 88, 661–692.
- Doornik, J. A. (2008). Encompassing and automatic model selection. *Oxford Bulletin of Economics and Statistics*, 70, 915–925.
- Doornik, J. A. (2009). Autometrics. In J. L. Castle & N. Shephard (Eds.), *The methodology and practice of econometrics* (pp. 88–121). Oxford: Oxford University Press.
- Doornik, J. A., & Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, 70, 927–939.
- Doornik, J. A., & Hendry, D. F. (2013a). *Empirical econometric modelling using PcGive* (7th ed., Vol. I). London: Timberlake Consultants Press.
- Doornik, J. A., & Hendry, D. F. (2013b). *Interactive Monte Carlo experimentation in econometrics using PcNaive: OxMetrics 7.10* (Vol. IV). London: Timberlake Consultants Press.
- Doornik, J. A., & Hendry, D. F. (2015). Statistical model selection with big data. *Cogent Economics and Finance*. Retrieved from <http://www.tandfonline.com/doi/full/10.1080/23322039.2015.1045216#VVE5bUYsAsQ>. doi:10.1080/23322039.2015.1045216
- Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32, 407–499.
- Enders, W. (1996). *RATS handbook for econometric time series*. New York, NY: Wiley.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflation. *Econometrica*, 50, 987–1007.
- Engle, R. F., & Granger, C. W. J. (1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica*, 55, 251–276.
- Engle, R. F., & Hendry, D. F. (1993). Testing super exogeneity and invariance in regression models. *Journal of Econometrics*, 56, 119–139.
- Engle, R. F., Hendry, D. F., & Richard, J.-F. (1983). Exogeneity. *Econometrica*, 51, 277–304.
- Ericsson, N. R., & Irons, J. S. (1994). *Testing exogeneity*. Oxford: Oxford University Press.
- Ericsson, N. R., & Irons, J. S. (1996). The Lucas critique in practice: Theory without measurement. In K. D. Hoover (Ed.), *Macroeconomics: Developments, tensions and prospects* (pp. 263–312). Dordrecht: Kluwer Academic Press.
- Faust, J., & Whiteman, C. H. (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, 121–161.
- Favero, C., & Hendry, D. F. (1992). Testing the Lucas critique: A review. *Econometric Reviews*, 11, 265–306.
- Florens, J.-P., Mouchart, M., & Rolin, J.-M. (1990). *Elements of Bayesian statistics*. New York, NY: Marcel Dekker.
- Frisch, R. (1938). *Statistical versus theoretical relations in economic macrodynamics* (Mimeograph dated 17 July 1938, League of Nations Memorandum. Reproduced by University of Oslo in 1948 with Tinbergen's comments. Contained in Memorandum "Autonomy of Economic Relations", 1948, November 6). Oslo: Universitets (O)konomiske Institutt.
- Frisch, R., & Waugh, F. V. (1933). Partial time regression as compared with individual trends. *Econometrica*, 1, 221–223.
- Godfrey, L. G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica*, 46, 1303–1313.

- Goldberger, A. S. (1964). *Econometric theory*. New York, NY: Wiley.
- Graddy, K. (1995). Testing for imperfect competition at the Fulton fish market. *RAND Journal of Economics*, 26, 75–92.
- Graddy, K. (2006). The Fulton fish market. *Journal of Economic Perspectives*, 20, 207–220.
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37, 424–438.
- Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16, 121–130.
- Granger, C. W. J., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, 2, 111–120.
- Hamilton, J. D. (2015). *Macroeconomic regimes and regime shifts* (NBER Working Paper no. 21863.). San Diego: Economics Department, University of California.
- Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica*, 12, 1–118.
- Hall, B. H., & Cummins, C. (2005). *Time series Processor 5.0 user's guide*. Palo Alto, CA: TSP International.
- Hansen, H., & Juselius, K. (1995). *CATS in RATS: Cointegration analysis of time series*. Evanston: Estima.
- Hendry, D. F. (1976). The structure of simultaneous equations estimators. *Journal of Econometrics*, 4, 51–88.
- Hendry, D. F. (1980). Econometrics: Alchemy or science? *Economica*, 47, 387–406.
- Hendry, D. F. (1984). Econometric modelling of house prices in the United Kingdom. In D. F. Hendry & K. F. Wallis (Eds.), *Econometrics and quantitative economics* (pp. 211–252). Oxford: Basil Blackwell.
- Hendry, D. F. (1986). Using PC-GIVE in econometrics teaching. *Oxford Bulletin of Economics and Statistics*, 48, 87–98.
- Hendry, D. F. (1990). Using PC-NAIVE in teaching econometrics. *Oxford Bulletin of Economics and Statistics*, 53, 199–223.
- Hendry, D. F. (1995a). *Dynamic econometrics*. Oxford: Oxford University Press.
- Hendry, D. F. (1995b). 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, & M. S. Morgan (Eds.), *Methodology and tacit knowledge: Two experiments in econometrics* (pp. 341–361). Chichester: Wiley.
- Hendry, D. F. (2000). Epilogue: The success of general-to-specific model selection. In *Econometrics: Alchemy or science?* (pp. 467–490). Oxford: Oxford University Press.
- Hendry, D. F. (2004). The Nobel memorial prize for Clive W.J. Granger. *Scandinavian Journal of Economics*, 106, 187–213.
- Hendry, D. F. (2009). The methodology of empirical econometric modeling: Applied econometrics through the looking-glass. In T. C. Mills & K. D. Patterson (Eds.), *Palgrave handbook of econometrics* (pp. 3–67). Basingstoke: Palgrave MacMillan.
- Hendry, D. F. (2015). *Introductory macro-econometrics: A new approach*. London: Timberlake Consultants. Retrieved from <http://www.timberlake.co.uk/macroeconometrics.html>
- Hendry, D. F., & Doornik, J. A. (1994). Modelling linear dynamic econometric systems. *Scottish Journal of Political Economy*, 41, 1–33.
- Hendry, D. F., & Doornik, J. A. (1999). The impact of computational tools on time-series econometrics. In T. Coppock (Ed.), *Information technology and scholarship* (pp. 257–269). Oxford: Oxford University Press.
- Hendry, D. F., & Doornik, J. A. (2014). *Empirical model discovery and theory evaluation*. Cambridge, Mass: MIT Press.
- Hendry, D. F., & Ericsson, N. R. (1991). Modeling the demand for narrow money in the United Kingdom and the United States. *European Economic Review*, 35, 833–886.
- Hendry, D. F., & Johansen, S. (2015). Model discovery and Trygve Haavelmo's legacy. *Econometric Theory*, 31, 93–114.
- Hendry, D. F., Johansen, S., & Santos, C. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics*, 33, 317–335. Erratum, 337–339.
- Hendry, D. F., & Juselius, K. (2000). Explaining cointegration analysis: Part I. *Energy Journal*, 21, 1–42.
- Hendry, D. F., & Juselius, K. (2001). Explaining cointegration analysis: Part II. *Energy Journal*, 22, 75–120.
- Hendry, D. F., & Krolzig, H.-M. (1999). Improving on 'Data mining reconsidered' by K.D. Hoover and S.J. Perez. *Econometrics Journal*, 2, 202–219.
- Hendry, D. F., & Krolzig, H.-M. (2001). *Automatic econometric model selection*. London: Timberlake Consultants Press.
- Hendry, D. F., & Krolzig, H.-M. (2005). The properties of automatic Gets modelling. *Economic Journal*, 115, C32–C61.
- Hendry, D. F., Marcellino, M., & Mizon, G. E. (2008). Special issue on encompassing. *Oxford Bulletin of Economics and Statistics*, 70, 711–938.
- Hendry, D. F., & Massmann, M. (2007). Co-breaking: Recent advances and a synopsis of the literature. *Journal of Business and Economic Statistics*, 25, 33–51.
- Hendry, D. F., & Mizon, G. E. (1999). The pervasiveness of Granger causality in econometrics. R. F. Engle & H. White (Eds.), *Cointegration, causality, and forecasting* (Vol. 5, pp. 104–134). Oxford: Oxford University Press.
- Hendry, D. F., & Mizon, G. E. (2000). Reformulating empirical macro-econometric modelling. *Oxford Review of Economic Policy*, 16, 138–159.
- Hendry, D. F., & Mizon, G. E. (2011). Econometric modelling of time series with outlying observations. *Journal of Time Series Econometrics*, 3(1). doi:10.2202/1941-1928.1100
- Hendry, D. F., & Mizon, G. E. (2014). Unpredictability in economic analysis, econometric modeling and forecasting. *Journal of Econometrics*, 182, 186–195.
- Hendry, D. F., & Nielsen, B. (2007). *Econometric modeling: A likelihood approach*. Princeton, NJ: Princeton University Press.
- Hendry, D. F., & Nielsen, B. (2010). A modern approach to teaching econometrics. *European Journal of Pure and Applied Mathematics*, 3, 347–369.
- Hendry, D. F., & Santos, C. (2010). An automatic test of super exogeneity. In M. W. Watson, T. Bollerslev, & J. Russell (Eds.), *Volatility and time series econometrics* (pp. 164–193). Oxford: Oxford University Press.
- Hendry, D. F., & Srba, F. (1980). AUTOREG: A computer program library for dynamic econometric models with autoregressive errors. *Journal of Econometrics*, 12, 85–102.
- Hoover, K. D., & Perez, S. J. (1999). Data mining reconsidered: Encompassing and the general-to-specific approach to specification search. *Econometrics Journal*, 2, 167–191.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12, 231–254.
- Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive Models*. Oxford: Oxford University Press.
- Johansen, S., & Nielsen, B. (2009). An analysis of the indicator saturation estimator as a robust regression estimator. In J. L. Castle & N. Shephard (Eds.), *The methodology and practice of econometrics* (pp. 1–36). Oxford: Oxford University Press.
- Johansen, S., & Nielsen, B. (2016). Asymptotic theory of outlier detection algorithms for linear time series regression models. *Scandinavian Journal of Statistics*. doi:10.1111/sjos.12174
- Johnston, J. (1963). *Econometric methods* (1st ed.). New York, NY: McGraw-Hill.
- Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H., & Lee, T.-C. (1985). *The theory and practice of econometrics* (2nd ed.). New York, NY: Wiley.

- Klein, L. R. (1950). *Economic fluctuations in the United States, 1921–41* (11 in Cowles Commission Monograph). New York, NY: Wiley.
- Krolzig, H.-M., & Hendry, D. F. (2001). Computer automation of general-to-specific model selection procedures. *Journal of Economic Dynamics and Control*, 25, 831–866.
- Leamer, E. E. (1978). *Specification searches. Ad-hoc inference with non-experimental data*. New York, NY: Wiley.
- Leamer, E. E. (1983). Let's take the con out of econometrics. *American Economic Review*, 73, 31–43. Reprinted in Granger, C. W. J. (Eds.). (1990). *Modelling economic series*. Oxford: Clarendon Press.
- 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 & A. Meltzer (Eds.), *The Phillips Curve and labor markets, Vol. 1 of Carnegie-Rochester conferences on public policy* (pp. 19–46). Amsterdam: North-Holland.
- Marschak, J. (1953). Economic measurements for policy and prediction. In W. C. Hood & T. C. Koopmans (Eds.), *Studies in econometric method* (14 in Cowles Commission Monograph). New York, NY: Wiley.
- Mizon, G. E. (1977). Model selection procedures. In M. J. Artis & A. R. Nobay (Eds.), *Studies in modern economic analysis* (pp. 97–120). Oxford: Basil Blackwell.
- Mizon, G. E. (1984). The encompassing approach in econometrics. In D. F. Hendry & K. F. Wallis (Eds.), *Econometrics and quantitative economics* (pp. 135–172). Oxford: Basil Blackwell.
- Mizon, G. E. (1995a). Progressive modelling of macroeconomic time series: The LSE methodology. In K. D. Hoover (Ed.), *Macroeconometrics: Developments, tensions and prospects* (pp. 107–169). Dordrecht: Kluwer Academic.
- Mizon, G. E. (1995b). A simple message for autocorrelation correctors: Don't. *Journal of Econometrics*, 69, 267–288.
- Mizon, G. E. (2008). Encompassing. In S. Durlauf & L. Blume (Eds.), *New Palgrave dictionary of economics* (2nd ed.). London: Palgrave Macmillan. doi:10.1057/9780230226203.0470
- Mizon, G. E. (2012). Seconding the vote of thanks on the retrospective reading of 'A return to an old paper: Test of separate families of hypotheses' by D. R., Cox. *Journal of the Royal Statistical Society, Series B*, 75, 213–214.
- Mizon, G. E., & Richard, J.-F. (1986). The encompassing principle and its application to non-nested hypothesis tests. *Econometrica*, 54, 657–678.
- Perez-Amaral, T., Gallo, G. M., & White, H. (2005). A comparison of complementary automatic modelling methods: RETINA and PcGets. *Econometric Theory*, 21, 262–277.
- Pesaran, M. H., & Pesaran, B. (1987). *Microfit: An interactive econometric software package users' manual*. Oxford: Oxford University Press.
- Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics*, 33, 311–340.
- QMS. (2005). *EViews 5.1 User's Guide. 2005. Quantitative Micro Software*. Irvine, CA. Retrieved from <http://www.eviews.com>
- Ramsey, J. B. (1969). Tests for specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society B*, 31, 350–371.
- Renfro, C. G. (1996). On the development of econometric modeling languages: MODLER and its first twenty-five years. *Journal of Economic and Social Measurement*, 22, 241–311.
- Renfro, C. G. (2009). *The practice of econometric theory: An examination of the characteristics of econometric computing*. London: Springer.
- Salkever, D. S. (1976). The use of dummy variables to compute predictions, prediction errors and confidence intervals. *Journal of Econometrics*, 4, 393–397.
- Teräsvirta, T., Tjøstheim, D., & Granger, C. W. J. (2011). *Modelling nonlinear economic time series*. Oxford: Oxford University Press.
- White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 817–838.
- White, H. (1990). A consistent model selection. In *Modelling economic series* (pp. 369–383). Oxford: Clarendon Press.
- 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.



© 2016 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

