APPLIED ECONOMETRICS WITHOUT SINNING

David F. Hendry

Oxford University

1. Introduction

One must offer a cautious welcome to any paper (Kennedy, 2001) discussing the problems confronting the implementation of econometrics to empirical data, especially from someone who has written an elementary econometrics textbook (see Kennedy, 1985, and later editions). There is a considerable gulf between the elementary — and inappropriate — methodology inherent in many such textbooks, and what applied econometricians need to do in practice to model economic data. There is some good advice in his paper, but also some which is bad: Kennedy fails to formalize the issues, is unclear at key points, and neither draws essential distinctions nor defines central terms. Consequently, although some of his recommendations should help to narrow the theory-applications gap, others could exacerbate the difficulties. The remainder of this note amplifies these comments.

First, the topic is in fact addressed in some textbooks, albeit none of which are cited by Kennedy: see, *inter alia*, Hendry and Doornik (2001) (and earlier editions, especially chs. 11 and 15); Hendry (1995a) provides extensive analyses of the problems of matching theory and evidence, empirical modelling, and data measurement issues, as well as de-emphasizing estimation, and highlighting the role of graphs; Spanos (1986) analyzes the difficulties inherent in the conventional paradigm, and clearly explains an alternative approach, now exposited in his introductory textbook, Spanos (1999); and Charemza and Deadman (1997) describe the practice of modelling non-stationary time series; also see Hendry and Wallis (1984; ch. 1) for an earlier general discussion. There are too many relevant papers on methodology to list additional to those Kennedy mentions, but Mizon (1995) and Hendry (1997) clarify some of his difficulties.

Second, the theory and practice of model building has advanced spectacularly in the last few years (see, e.g. Hoover and Perez, 1999, 2000; Krolzig and Hendry, 2001; Hendry and Krolzig, 2001b). Many of the so-called 'sins' turn out to be false accusations, while others (formulating and checking the ingredients of the basic specification) remain essential if anything useful is to result from empirical modelling. Thus, considerable revision of the 'Ten Econometric Commandments'

0950-0804/02/04 0591–14 JOURNAL OF ECONOMIC SURVEYS Vol. 16, No. 4

is needed to clarify where in the empirical enterprise each issue arises, and hence which issues are substantive for the discipline, and which are products of non-viable methodologies.

Third, I believe a greatly improved understanding of the theory of economic forecasting has developed over the last few years, summarized in Clements and Hendry (1998, 1999a), with important implications for modelling methodology, and some of these are noted below.

The role of econometric theory in empirical economics is sketched in Section 2 to show that, despite its limitations, those are not the source of Kennedy's complaints. The flawed 'textbook' paradigm is critically evaluated in Section 3, and four non-solutions are noted in Section 4. Six aspects where methodology could improve empirical modelling procedures are analyzed in Section 5, leading to a revised list of 'ten commandments'. These apparently overlap the non-solutions: this is not a paradox — the key lies in *how* ingredients enter the empirical modelling enterprise, not *whether*. Section 6 briefly concludes.

2. Background

Econometric theory is based on:

- a postulated data generating process (DGP) — which might not characterize the reality to be modelled;
- a theoretical model specification — which might not match the DGP;
- a data set — which might be incorrectly measured and not correspond to the theory variables;
- a method of estimation — which while appropriate for the model, might even then only have a large-sample justification and known asymptotic distributional properties, but need have no useful properties relative to the data sample and the given DGP; and usually:
- a set of evaluation or inference procedures — which at best can reject sufficiently-false hypotheses, without revealing why.

We know such theory has many limitations even within the confines of the DGPs currently postulated as a basis for time-series econometrics. We also know that those DGPs are far too simple. Econometric theory advances from improved tools and increased understanding of relevant features of reality to incorporate in the postulated DGP — an already classic example is the presence of unit roots and cointegration, enforcing revisions to every aspect of specification, modelling, estimation, inference, and distributional behaviour. Such theory is an essential component of the measuring instrument of economics, and its calibration will remain a central preoccupation of econometricians, on whichever 'floor' it is taught in the USA. Although some of its limitations can be attenuated by experience, such limitations do not seem to be the source of Kennedy's complaints.

Rather, the source is the presentation in many textbooks — and indeed, most published empirical papers — of an untenable 'methodology'. Models are handed down by economists, merely needing quantitative cloth; appropriate estimation procedures then 'confirm' the original hypotheses ('a new economic law is forged': see Pagan, 1987). If pushed, it may be conceded that data are 'noisy', so models may not work instantly but need some 'practical corrections' by an econometrician. Does one need to ask why this 'textbook' approach does not work?

3. The flawed 'textbook' paradigm

First, economic analysis is based on abstract reasoning, so is inherently incorrect, incomplete, and mutable. Thus, it would require a different form of miracle from Keynes' example for the initially postulated model to be the optimal starting point for econometric modelling.

Secondly, data observations are prone to many problems rarely addressed by the theory model, and cannot be treated as a given. While the names of the variables may be the same, there remain correspondence problems to be addressed: e.g. relating 'consumption' to collected data on 'expenditure by consumers'.

Thirdly, there are many non-stationarities in economics; changes in legislation, technology, tastes, political, economic and exchange-rate regimes, wars, commodity crises and so on. These should be modelled to avoid distorting the analysis. Simple, or theory-imposed, models will not begin to capture such complexities: such models are the wrong starting point, and Kennedy does his readers a dis-service by reiterating this dimension of textbook 'methodology' — Section 4.2 amplifies.

Finally, any user of such an approach is destined to find a mis-match of model and evidence. Now the implicit textbook advice usually goes completely off the road, suggesting a variety of unjustifiable solutions to such 'problems', essentially camouflaging, rather than solving, them. A classic example is pretending that residual autocorrelation can be regarded as error autoregression (and not due to structural breaks, wrong functional form, measurement errors, omitted variables, etc.). A more subtle drawback — within its own paradigm — is the pretence that this is the only study of that data, inducing covert data-mining across studies (see the critique of 'data snooping' in White, 2000; Sullivan, Timmermann and White, 2001, and the improvements in Hansen, 2001).

No viable methodology will ever be constructed out of such manifestly invalid procedures: see Hendry (2000) and Spanos (2000) for more extensive critiques. We will first consider some proposals that transpire to be non-solutions, then consider more promising approaches.

4. Some non-solutions

We consider four proposals that seem unlikely to help: exhortations to impose more economic theory in place of data analysis; starting empirical analyses from simple models; avoiding the use of significance tests; and 'sensitivity analyses' in the senses discussed by Kennedy.

4.1. *Economic theory and common sense*

Most empirical models that are tightly based on existing economic theory are doomed not to characterize the data evidence: key issues are usually elided, rather than resolved, including heterogeneity; agents' information deficiencies; constraints on behaviour; the measurement of 'time'; aggregation across agents, commodities, and regions etc.; changes in institutional structures; and the construction of the variables. The resulting models suffer from theory dependence and will get discarded whenever the underlying theory is improved: see Hendry (1995b). Worse still, 'common sense' is all too frequently misleading — Gallileo found that out when trying to explain that the sun does not actually 'rise' — and we do not know when such cases apply. As Gottlieb (2000, p.ix) expressed the matter:

> Philosophers have regularly cocked an eyebrow at what passes for the common sense of the time; the punchline comes later, when it is 'common sense' that turns out to have been uncommonly confused.

Thus, Kennedy is effectively appealing to the first two 'golden prescriptions' in Hendry (1987), and there is no solution in that advice. 'Better' theory, and clearer thinking that transpires to be closer to reality, will undoubtedly help, but the only way to judge such properties is by a closer match of the outcome to the evidence, a point resumed in Section 5.

4.2. *Simple models*

We consider three aspects of simplicity: where it enters an analysis, what it is, and when it matters.

4.2.1. *Where does simplicity enter?*

Starting from, and ending with, simple models are entirely different methodologies. Starting with an overly simplified model is a recipe for disaster: nobody has ever produced a Monte Carlo study showing how to implement a 'generalization' modelling strategy, precisely because there is no known way to validly infer from its rejection how a model should be 'improved'. The only basis for claiming that the undefined phrase 'sensibly simple' can justify an initial specification is assertion and anecdote, not analysis and fact. The pernicious drawbacks of simple-to-general modelling procedures are so well established, it is astonishing to see that re-iterated as a 'methodology'. Even in the most structured case of an ordered sequence of nested hypotheses, Anderson (1962) demonstrated its non-optimality against a general-to-simple approach.

Moreover, if one begins with a model that fails to characterize the evidence, one cannot conduct sensible inference, nor interpret test statistics, and must indulge in a sequence of logical non-sequiturs, inferring 'solutions' to 'problems' in the light of false evidence, thereby generating a proliferation of non-encompassing models.

Indeed, the fact that model evaluation can be conducted by variable addition (see Pagan, 1984) then confirms the non-viability of commencing from an overly simple model. Such an approach cannot be commended to anyone as a way to practice, and certainly not as a way to teach, econometrics. Note *en passant* the internal contradiction in Magnus (1999), which has the implication that because it is unrealistic to allow for everything at the outset, one should therefore allow for almost nothing.

Conversely, ending with the simplest model that characterizes the evidence has a great deal to commend it, some aspects of which are discussed by Kennedy — without distinguishing whether they justify the start or the completion of the analysis. The key feature is whether or not the specified model is congruent, not 'simple'. We consider model selection in Section 5.3.1.

4.2.2. *What is a simple model anyway?*

When is a specification simple and when complex? Vector equilibrium-correction systems are hugely complicated in one sense (see Johansen, 1995, and the 'statistical technology' that underlies them), yet very simple in another, so cannot be uniquely characterized by 'measures of simplicity'. Often the 'general' specification that Kennedy objects to is actually a simple model, like an autoregressive-distributed lag conditional equation: reduction then leads to the simplest (most parsimonious) representation that the modeller can develop, so ensures a 'sophisticatedly simple' conclusion.

4.2.3. *When should a model be simple?*

Unfortunately, Kennedy also does not distinguish the purpose of a study in relation to the application of this 'principle': in particular, modelling is conflated with forecasting. For the latter, it is well established that some simple models do well in forecasting competitions: see Makridakis, Andersen, Carbone, Fildes *et al.* (1982) and Makridakis and Hibon (2000). However, $\hat{y}_t = a + bt$ is a really simple model, and does really badly in forecasting, so simplicity alone is not the criterion — simplicity may be 'sensible' in some settings, but not in others. Clements and Hendry (1999b, 2001a) explain why it is not the *simplicity* of models that matter, but their adaptability to structural breaks: it just happens that many widely used simple models, such as EWMA or $\Delta^2\hat{y}_t = 0$, are also adaptive (here, to shifts in intercepts and trends). An important implication is that *ex ante* forecasting comparisons should not be used to evaluate models (except for forecasting).

These confusions are regrettable in a paper claiming to offer 'ten commandments' for applied econometrics. One concludes that his re-defined KISS in rule #5 is neither a valid message *per se*, nor an enhanced variant of Occam's razor. A more useful rule is to commence from a specification such that, if a more general model was needed, new information would have been learned. More optimistically, I hope Kennedy is correct to claim that most econometricians 'believe in a general-to-simple approach'.

4.3. *Avoiding significance tests*

As before, key distinctions are not made: testing the assumptions underpinning the validity of a given statistical technique is quite different from a mechanical t-test of an uninteresting null. On the former — how else could one appraise that congruence held? Kennedy here gets dangerously close to dismissing a scientific approach to econometrics in favour of *ad hocery*. Of course, Milton Friedman would place little faith in significance tests — they systematically reject his models: see Hendry and Ericsson (1991). That is hardly a good reason not to rigorously evaluate claims in economics. Moreover, one must not forget the converse problem — when estimating a relation where parameters of $+1, -1$ are anticipated, imagine finding coefficients of -10 and $+10$: that their standard errors are 15 is crucial knowledge.

The issue of letting significance levels tend to zero as sample size increases is not substantive: see e.g. Akaike (1973), Schwarz (1978) and e.g. Hendry (1995a). Choosing the appropriate level is an integral aspect of any empirical analysis, and should not be 'conventionalized'.

4.4. *'Sensitivity analyses'*

Two issues are conflated in Kennedy's discussion:

1. the logic of reporting in economics (state your theory, then test it, then stop); and
2. the reliability of the reported findings under 'minor' changes in the specification.

The first is possibly an attempt to bolster a false methodology, and is indeed often mis-leading. Consequently, some 'measure' of the reliability of the results is essential. Unfortunately, 'fragility' is not well-defined: 'Leamer-fragile', as in extreme bounds, merely tells us that some variables matter empirically (see e.g. Breusch, 1990; Hendry and Mizon, 1990). Even genuine 'fragility' (e.g. in response to changing the sample by one observation) is ill-defined when unknown non-stationarities lurk. In such a setting, a more constructive approach is through the concept of extended constancy: see Ericsson, Hendry and Prestwich (1998). We re-consider the second issue in Section 5.5 below.

5. Some solutions

We consider six areas where methodology provides some guidance to improve empirical modelling procedures over the 'textbook approach', namely economic analysis, data and graphics, general to simple, rigorous testing, sensitivity analysis, and encompassing, then suggest a revised list of ten 'commandments'. As noted above, it is far from paradoxical that several of the resulting headings share the same titles as non-solutions: it is crucial how each enters an empirical analysis, not just whether it should enter.

5.1. *Economic analysis*

That economic analysis is not the perfect starting point (see § 3) does not detract from its central role in formulating what to model econometrically — it remains our best, and often only, guide to understanding the complexities of a modern economy. Unfortunately, such theory is usually too abstract to cover all the aspects of the empirical situation (however, see e.g. Hildenbrand, 1994, for some promising developments). The key distinction is between such theory being useful, and hence guiding the specification, and being optimal, when it is 'the specification'. The latter is a long time off in such a complex discipline as macro-economics.

5.2. *Data and graphics*

Data measurement is far from arbitrary, and here economic theory can help resolve some of the correspondence problems: witness its major contributions to the development of national income accounts. Nevertheless, this aspect is both down played in most textbooks and rarely taught: courses are needed on such problems, comprising explanations of index number construction, data revision processes, alternative aggregators, hedonic prices, quality adjustments and so on — all theory related. There are time opportunity costs to teaching such material, but it remains moot whether the present balance does not over-emphasize high-powered econometric technique. Here, I have sympathy for Kennedy's arguments.

Moreover, one must distinguish a full understanding of the data-collection process — including data definitions, measurement, selection, meaning, accuracy, scales, quality and so on — which are all essential pre-requisites of a viable study — from any formal role accorded to the economic theory. Both are also distinct from understanding the real-life system to be modelled. The main paragraph in rule #1 is really about matching dimensionality, transformations, selecting functional forms, data admissibility and understanding the metric of the problem. Nothing in any existing prescription of econometrics entails that these constitute 'sins' — indeed the opposite: the sin may be that they are rarely properly treated. Similar considerations apply to data graphs: the computer revolution has provided ready access to wonderful graphics. Failure to seize that opportunity could reflect a lack of thinking, but I suspect instead the pernicious influence of a false 'scientific methodology' which argues against doing so to avoid 'contaminating the analysis'. Yet Newton was well versed in Kepler's data. The converse problem is not knowing enough about the data: e.g. Davidson, Hendry, Srba and Yeo (1978) noted that a graph of consumers' expenditure was an 'eye-opener' which revealed that its dramatic seasonal fluctuations were far in excess of those of income, vitiating almost all previous models — and their own earlier research — but prompting an improved understanding.

5.3. *Model selection*

The role of selection in both modelling and forecasting needs to be considered.

5.3.1. *General to simple (Gets)*

Gets is sustained by the theory of reduction (see e.g. Hendry, 1987; Florens, Mouchart and Rolin, 1990), which establishes the existence of a local DGP (LDGP) in the space of the variables under analysis, as well as the link between encompassing that LDGP and the congruence of an empirical model (see e.g. Bontemps and Mizon, 2001). Moreover, White (1990) showed the convergence of a rigorous progressive testing strategy to that LDGP. There is also a substantive body of simulation evidence refuting the drawbacks claimed in Kennedy about Gets. First, mis-specification tests are only computed once (see § 5.4): later use does not affect their statistical properties. Secondly, the 'excessive size' (or 'high chance of false rejection') of multiple testing is badly mis-understood. Imagine including 20 irrelevant orthogonal variables; test each at 5%; the theory implies that the probability that none is included is $p = 1 - (1 - 0.05)^{20} \simeq 0.64$, which is then interpreted as the 'size' of the selection process. Instead, note that on average, only *one* of the 20 variables will be retained (0.05×20), and the other 19 will be correctly eliminated: this is a small cost for protecting oneself against erroneous omission of variables that do matter empirically. Testing at the 1% level delivers $p = (1 - 0.01)^{20} \simeq 0.18$ or one variable retained every five attempts despite such profligate over-parameterization — and this using non-optimal inference procedures (see Hendry and Krolzig, 2001b). The Golden rules from Hendry (1980) do not 'distort' Type-I errors, as the detailed studies in Hoover and Perez (1999, 2000), Hendry and Krolzig (1999, 2001b) and Krolzig (2000) demonstrate.

The Gets procedure has three main stages: formulate the general model; test its congruence, then simplify. The first is the key step, and merits the greatest share of the research time in theoretical development, data collection, literature survey for encompassing contenders, and formulation of the function forms, parameters, maximum lags (if time series) and so on. Congruence testing is to check that the general model does characterize the available data evidence. Finally, powerful search algorithms have been developed, with suitable 'pre filters' to eliminate irrelevant variables, multi-path searches for all congruent dominating reductions, encompassing selections between, and post-selection sample splits to check for adventitious significance. The algorithms have been demonstrated to have an actual size close to that pre-set, and power close to that achievable when commencing from the 'correct' model (i.e. with no redundant variables). Thus, the costs of search are small, especially relative to the unavoidable costs of inference that confront an investigator commencing from the LDGP.

5.3.2. *Forecasting*

Does model selection distort forecasts? The answer here from the studies by Clark (2000) and Clements and Hendry (2001b) is unequivocal: it does not. Once one realizes that biased parameter estimates need not entail biased forecasts, and that neither model mis-specification nor parameter estimation uncertainty can explain forecast failure, such an answer is unsurprising: see Clements and Hendry (1998, 1999a). Models will suffer forecast failure if they incorporate irrelevant variables which change, or omit relevant that shift. Thus, if forecast failure is not to persist, selection is a central component of any practical progressive research strategy in a world of deterministic shifts: neither an unrestricted, nor an *a priori* based simple model will do well — unless they are robustified to deterministic shifts.

5.4. *Testing*

Another key distinction not made by Kennedy is between test criteria used in evaluation, and those used in design. Testing for congruence needs to be rigorous yet directed (e.g. for time series, use the information taxonomy in Hendry, 1995a, to test for innovation, homoscedastic, normal errors; weakly exogenous conditioning variables for the parameters of interest; and constant parameters). Evaluation tests need nominal sizes close to actual, with power to detect relevant departures from the null, and overall size controlled (e.g. with a maximum of 5% when the tests are independent under the null). Such mis-specification tests are applied once to the general model to evaluate congruence, so have the statistical properties claimed: see e.g. Krolzig and Hendry (2001). Their repeated calculation to ensure reductions are valid does not alter those properties, merely precluding invalid reductions. Such an approach is entirely different from continually re-specifying an overly simple model that manifests a 'problem' — as detected by some test — until that test is 'insignificant' (e.g. 'Cochrane–Orcutting' till the Durbin–Watson test does not reject). Although 'Neyman–Pearson' testing of a model certainly needs fresh data, the so-called 'double' use of data (developing and testing a specification) does not 'distort': as seen in Section 5.3.1, Gets has the correct size yet power close to commencing from the LDGP.

5.5. *Sensitivity analysis*

The computer-automation revolution resolves Keynes's worry: PcGets selects a unique parsimonious, dominant, and congruent representation from a given general model — his miracle has happened. There is no 'path dependence' (see Pagan, 1987), and Gets rebuts Leamer's criticism — there is no sin with only one search which has a good chance of recovering the LDGP. Sensitivity of the selected specification can be investigated by recursive methods, dating back to Teräsvirta (1970) in econometrics, but greatly improved by modern graphics, as well as sub-sample significance, as in Hoover and Perez (1999). En-forcing the entry of some theory-relevant variables and re-selecting can also be useful in some contexts. Testing for super exogeneity (as in Engle and Hendry, 1993) and invariance (e.g. Favero and Hendry, 1992) are other valuable adjuncts, as is mis-specification encompassing or predicting anomalous results in previous studies (see e.g. Hendry and Richard, 1989).

5.6. *Encompassing*

A plethora of results from the same data set will eventuate from the advice Kennedy offers, so will need reduction: encompassing suggests how to achieve that. However, Gets would help prevent one from facing such a problem in the first place. An important missing reference is to the 'Gilbert test' (see Gilbert, 1986), whose clever analogy — namely check that no intermediate result dominates, or fails to be encompassed by, the published finding — both clarifies how to detect 'data mining', and counters the idea of publishing every regression ever run (which anyway neglects consumption costs: also see Spanos, 2000, for the role of test 'severity' in discriminating unwarranted from warranted 'data mining'; and Campos and Ericsson, 1999, for different meanings of that term).

5.7. *Useful 'rules'*

The fog on his mountain misled Kennedy into picking up the wrong tablets. A more appropriate set are:[1]

1. Begin by deciding on the substantive problem and the objective of the analysis: e.g. modelling, theory-testing, forecasting, or policy (§ 5.3);
2. Ascertain the existing knowledge state of the theory and the evidence, and extend these as appropriate (§ 5.1);
3. Obtain the relevant data and carefully check its accuracy, reliability, and suitability for the analysis envisaged (§ 5.2);
4. Understand the institutional context, and any important historical events, including major legislative, political, and technical changes (§ 3);
5. Formulate a general statistical model that is a theory-consistent, identified, evidence-encompassing representation, with substantive reasons for the functional forms, conditioning assumptions, and parameterizations (§ 5.3.1);
6. Test it for congruence in all directions relevant to the analysis (§ 5.4);
7. Develop the most parsimonious, undominated representation that remains congruent (PcGets, Hendry and Krolzig, 2001a, will accomplish this step: § 5.3.1);
8. Rigorously evaluate the resulting selection for dis-confirming evidence (§ 5.5);
9. Check that the model encompasses both theory implications and all related empirical findings (§ 5.6);
10. Return when new data accrue to evaluate the outcome in a progressive research strategy (§ 5.4).

6. Conclusion

The paper critically reviews the proposals in Kennedy (2001), then suggests more appropriate approaches. While there is some overlap in Section 5.7 with the 'Ten Commandments' in Kennedy, there are also important correctives to his proposals. The revised list can be taught — and has been the basis of a considerable body of successful empirical research — unlike vague exhortations to 'think of the correct answer at the start' implicit in Kennedy's 'Ten Commandments'.

Compromises may be needed at any stage of an empirical study, from an inability to solve the relevant economic-theory model, find adequate data, or discover constant parameters, through to adding indicators for outliers, or leaving anomalies for future resolution. However, not all compromises are equally good: e.g. conceding data admissibility (as White, 1990, suggests), is fine for a normal distribution approximation for height, but not if negative unemployment results. Perhaps this is where 'tacit knowledge' enters, and why 'apprenticeships' seem necessary: see Magnus and Morgan (1999).

One of the more unfortunate aspects of Kennedy (2001) is the number of methodological assertions which remain unsubstantiated by either sustained empirical evidence or Monte Carlo evaluation. Thus, there is a regrettable perpetuation of some folklore beliefs of econometrics based on little more than 'I have never tried it because I don't like it'. The contrast could not be more stark with the wealth of evidence that Gets works, both empirically, and from extensive Monte Carlo studies. To quote Stephen Jay Gould (2000, p. 318):

> Centuries of vain speculation dissolved in months before the resolving power of Galileo's telescope ...

Perhaps econometrics is just enough of a science for recent findings to play a similar role in our discipline.

Acknowledgements

Financial support from the U.K. Economic and Social Research Council under grant L11625015 is gratefully acknowledged. I am indebted to Mike Clements, Jurgen Doornik, Neil Ericsson, Hans-Martin Krolzig and Aris Spanos for helpful comments.

Notes

1. I assume the objective is the first of these (approaches for the others are discussed in Hendry and Mizon, 2000).

References

Akaike, A. (1973) Information theory and an extension of the maximum likelihood principle. In Petrov, B. N. and Saki, F. L. (eds), *Second International Symposium of Information Theory*, Budapest.

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.

Bontemps, C. and Mizon, G. E. (2001) Congruence and encompassing. In Stigum (2001), forthcoming.

Breusch, T. S. (1990) Simplified extreme bounds. In Granger (1990), pp. 72–81.

Campos, J. and Ericsson, N. R. (1999) Constructive data mining: Modeling consumers' expenditure in Venezuela. *Econometrics Journal*, 2, 226–240.

Charemza, W. W. and Deadman, D. F. (1997) *New Directions in Econometric Practice*. Cheltenham: Edward Elgar.
Clark, T. E. (2000) Can out-of-sample forecast comparisons help prevent overfitting? Research Division, Federal Reserve Bank of Kansas City.
Clements, M. P. and Hendry, D. F. (1998) *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.
Clements, M. P. and Hendry, D. F. (1999a) *Forecasting Non-stationary Economic Time Series*. Cambridge, Mass.: MIT Press.
Clements, M. P. and Hendry, D. F. (1999b) On winning forecasting competitions in economics. *Spanish Economic Review*, 1, 123–160.
Clements, M. P. and Hendry, D. F. (2001a) Explaining the results of the M3 forecasting competition. *International Journal of Forecasting*, 17, 550–554.
Clements, M. P. and Hendry, D. F. (2001b) Modelling methodology and forecast failure. *Econometrics Journal*, forthcoming.
Davidson, J. E. H., Hendry, D. F., Srba, F. and 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. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?*. Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000.
Engle, R. F. and Hendry, D. F. (1993) Testing super exogeneity and invariance in regression models. *Journal of Econometrics*, 56, 119–139. Reprinted in Ericsson, N. R. and Irons, J. S. (eds), *Testing Exogeneity*, Oxford: Oxford University Press, 1994.
Ericsson, N. R., Hendry, D. F. and Prestwich, K. M. (1998) The demand for broad money in the United Kingdom, 1878–1993. *Scandinavian Journal of Economics*, 100, 289–324.
Favero, C. and Hendry, D. F. (1992) Testing the Lucas critique: A review. *Econometric Reviews*, 11, 265–306.
Florens, J.-P., Mouchart, M. and Rolin, J.-M. (1990) *Elements of Bayesian Statistics*. New York: Marcel Dekker.
Gilbert, C. L. (1986) Professor Hendry's econometric methodology. *Oxford Bulletin of Economics and Statistics*, 48, 283–307. Reprinted in Granger, C. W. J. (ed.) (1990) *Modelling Economic Series*. Oxford: Clarendon Press.
Gottlieb, A. (2000) *The Dream of Reason*. London: The Penguin Press.
Gould, S. J. (2000) *The Lying Stones of Marrakech*. Random House, London: Johanthan Cape.
Granger, C. W. J. (ed.) (1990) *Modelling Economic Series*. Oxford: Clarendon Press.
Hansen, P. R. (2001) An unbiased test for superior predictive ability. Working paper 01–06, Brown University.
Hendry, D. F. (1980) Econometrics: alchemy or science?. *Economica*, 47, 387–406. Reprinted in Hendry, D. F. (1993, 2000) *op. cit*.
Hendry, D. F. (1987) Econometric methodology: A personal perspective. In Bewley, T. F. (ed.), *Advances in Econometrics*, Ch. 10. Cambridge: Cambridge University Press.
Hendry, D. F. (1995a) *Dynamic Econometrics*. Oxford: Oxford University Press.
Hendry, D. F. (1995b) Econometrics and business cycle empirics. *Economic Journal*, 105, 1622–1636.
Hendry, D. F. (1997) On congruent econometric relations: a comment. *Carnegie–Rochester Conference Series on Public Policy*, 47, 163–190.
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. New Edition.
Hendry, D. F. and Doornik, J. A. (2001) *Empirical Econometric Modelling using PcGive 10: Volume I*. London: Timberlake Consultants Press.
Hendry, D. F. and Ericsson, N. R. (1991) An econometric analysis of UK money demand

in 'Monetary Trends in the United States and the United Kingdom' by Milton Friedman and Anna J. Schwartz. *American Economic Review*, 81, 8–38.

Hendry, D. F. and 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. and Krolzig, H.-M. (2001a) *Automatic Econometric Model Selection*. London: Timberlake Consultants Press.

Hendry, D. F. and Krolzig, H.-M. (2001b) New developments in automatic general-to-specific modelling. In Stigum (2001), forthcoming.

Hendry, D. F. and Mizon, G. E. (1990) Procrustean econometrics: or stretching and squeezing data. In Granger (1990), pp. 121–136.

Hendry, D. F. and Mizon, G. E. (2000) Reformulating empirical macro-econometric modelling. *Oxford Review of Economic Policy*, 16, 138–159.

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

Hendry, D. F. and Wallis, K. F. (eds) (1984) *Econometrics and Quantitative Economics*. Oxford: Basil Blackwell.

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

Hoover, K. D. and Perez, S. J. (1999) Data mining reconsidered: Encompassing and the general-to-specific approach to specification search. *Econometrics Journal*, 2, 167–191.

Hoover, K. D. and Perez, S. J. (2000) Truth and robustness in cross-country growth regressions, unpublished paper, Economics Department, University of California, Davis.

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

Kennedy, P. E. (1985) *A Guide to Econometrics*. Oxford: Basil Blackwell. Second Edition.

Kennedy, P. E. (2001) Sinning in the basement: What are the rules? The Ten Commandments of applied econometrics. *Journal of Economic Studies*, 16, 569–589.

Krolzig, H.-M. (2000) General-to-specific reductions in vector autoregressive processes. Economics discussion paper, 2000-w34. Oxford: Nuffield College.

Krolzig, H.-M. and Hendry, D. F. (2001) Computer automation of general-to-specific model selection procedures. *Journal of Economic Dynamics and Control*, 25, 831–866.

Magnus, J. R. (1999) The success of econometrics. *De Economist*, 147, 55–71.

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

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., *et al.* (1982) The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, 1, 111–153.

Makridakis, S. and Hibon, M. (2000) The M3-competition: Results, conclusions and implications. *International Journal of Forecasting*, 16, 451–476.

Mizon, G. E. (1995) Progressive modelling of macroeconomic time series: The LSE methodology. In Hoover, K. D. (ed.), *Macroeconometrics: Developments, Tensions and Prospects*, pp. 107–169. Dordrecht: Kluwer Academic Press.

Pagan, A. R. (1984) Model evaluation by variable addition. in Hendry, and Wallis (1984), pp. 103–135.

Pagan, A. R. (1987) Three econometric methodologies: A critical appraisal. *Journal of Economic Surveys*, 1, 3–24. Reprinted in Granger, C. W. J. (ed.) (1990) *Modelling Economic Series*. Oxford: Clarendon Press.

Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.

Spanos, A. (1986) *Statistical Foundations of Econometric Modelling*. Cambridge: Cambridge University Press.

Spanos, A. (1999) *Probability Theory and Statistical Inference: Econometric Modeling with Observational Data*. Cambridge: Cambridge University Press.

Spanos, A. (2000) Revisiting data mining: Hunting with or without a license. *Journal of Economic Methodology*, 7, 231–264.

Stigum, B. (ed.) (2001) *Econometrics and the Philosophy of Economics*. Cambridge, Mass.: MIT Press, forthcoming.

Sullivan, R., Timmermann, A. and White, H. (2001) Dangers of data-driven inference: The case of calendar effects in stock returns. *Journal of Econometrics*, 105, 249–286.

Teräsvirta, T. (1970) *Stepwise Regression and Economic Forecasting*. No. 31 in Economic Studies Monograph. Helsinki: Finnish Economic Association.

White, H. (1990) A consistent model selection, in Granger (1990), pp. 369–383.

White, H. (2000) A reality check for data snooping. *Econometrica*, 68, 1097–1126.