Achievements and Challenges in Econometric Methodology

David F. Hendry*
Department of Economics, Oxford University.

April 12, 2000

Abstract

Disputes about econometric methodology have abounded in econometrics, yet their attempted resolution has attracted only a small proportion of research effort. Recently, computer automated general-to-specific reductions have been shown to perform well in Monte Carlo experiments, recovering the DGP specification from a much larger general model with size and power close to commencing from the DGP itself. Thus, future developments appear promising, with many ideas awaiting implementation and both theoretical and simulation evaluation.

As in many disciplines, computers have played an essential role in making econometrics operational. Software for data management, modelling, estimation, inference, simulation, and graphics, has underpinned most of the great empirical strides in our subject (and many other social sciences – see e.g., Coppock, 1999). A new generation of programs now bids fair to resolve some of the hotly debated issues about model selection in macro-econometrics – and 'data mining' – by demonstrating the success of general-to-specific (henceforth denoted *Gets*).

There can be little dispute that econometric methodology lacks a consensus: witness the debates in Granger (1990) and Hendry, Leamer and Poirier (1990), the diversity of approaches in d'Autume and Cartelier (1996) and Magnus and Morgan (1999), as well as the gulf in views on *Business Cycle Empirics* in the *Economic Journal* (November, 1995) and in Backhouse and Salanti (2000), among many other possible citations. Yet there has been a paucity of evidence, as against argument, on how well alternative approaches actually perform in realistic settings. A notable exception is the set of Monte Carlo experiments in Lovell (1983) – who found that none of the methods which he implemented worked well. There has, however, been a dramatic recent change with the computer implementation of several model search algorithms – see Hoover and Perez (1999), Hendry and Krolzig (1999), Hansen (1999), Sullivan, Timmermann and White (1998) and White (1997) *inter alia*. We focus on the first of these.

To program their Monte Carlo simulations for evaluating *Gets*, Hoover and Perez (1999) make several important innovations. First, they search many reduction paths, not just successively smallest t values – so their algorithm avoids getting 'stuck' in a poor choice. Secondly, each selection remains congruent. Thirdly, they choose between any resulting (non-nested) terminal models by fit. Finally, they use sub-sample reliability to assess the selection. Re-running the experiments in Lovell (1983) revealed excellent performance by *Gets* in many experiments.

Krolzig and Hendry (1999) develop an improved algorithm (written in *Ox*: see Doornik, 1998) called *PcGets*, which adds block-reduction F-tests, calibrates the diagnostic tests, applies encompassing tests between terminal contenders, repeats searches with non-unique outcomes, and uses the Schwartz (or BIC) criterion for final-model selection if required. Thus, *PcGets* embodies all the principles discussed in Hendry (1995). In the Lovell Monte Carlo experiments with highly over-parameterized models (e.g.,

*I am grateful to Jurgen Doornik, Neil Ericsson, and Hans-Martin Krolzig for comments. Financial support from the UK Economic and Social Research Council under grant R000234954 is gratefully acknowledged.

40 irrelevant variables when searching for a null model), *PcGets* improves substantially over the Hoover and Perez (1999) algorithm (e.g., correctly locating that null model more than 97% of the time: see Hendry and Krolzig, 1999). In their own Monte Carlo experiments, Krolzig and Hendry (1999) report that *PcGets* recovers the data-generation process (DGP) specification with an accuracy close to what one would expect when commencing from the DGP and conducting equivalent tests. Thus, the costs of structured searches are much lower than currently perceived – repeated testing need not induce oversized models. Applied to the macro-data in Davidson, Hendry, Srba and Yeo (1978) and Hendry and Ericsson (1991), *PcGets* selects models at least as good as those developed over several years by their authors. Thus, this operational version of the methodology confirms its power – and we have barely scratched the surface of automated model selection methodology.

Writing programs requires precise concepts and operations, and often highlights lacunae in existing theories (see the estimator-generating equation in this journal prompted by programming FIML, Hendry, 1976). Automating model selection is doing likewise. The multiple-path search strategy plays a fundamental role in improving performance over single-path 'step-wise' searches; and ensuring congruence throughout avoids incorrect inferences from inappropriate standard errors. An important distinction between the costs of inference (the subject of extensive research under the rubric 'pre-testing': see *inter alia* Bock, Yancey and Judge, 1973, and Judge and Bock, 1978) and the costs of search has been clarified: the latter is measured by the difference in *Gets* performance between commencing from the general model and from the DGP specification, whereas the former is the difference between the selected model and the DGP starting point. The statistical properties of cumulative F-tests (which also perform well in Hansen, 1999, prior to his use of BIC), of many-path searches, of split-sample reliability and recursive testing, repeated reduction rounds with encompassing choices, and the interactions of diagnostic testing and selection, are all undergoing rigorous analyses to clarify why such good selections result. Analysis of the number of paths to search is required: one is too small, as the algorithm can get stuck, but searching every possible path seems to risk over-selecting adventitious effects.

The specification of the initial general congruent model is central to the search process. Careful prior analysis remains essential, concerning the choice of variables, lag lengths, functional forms, orthogonality, parameterizations, indicators (including seasonals) etc. The *larger* the initial regressor set, the more likely adventitious effects will be retained, suggesting a key role for economic theory in 'prior simplification'. But the *smaller* that set, the more likely important variables will be omitted: given the low costs of search, relatively generous initial parameterizations seem sensible. That runs counter to much of modern macro-econometrics where 'representative-agent, tightly parameterized' models are in vogue.

Many generalizations are both needed and feasible. For example, the impact of forced search paths (to ensure the presence of some variables); system and instrumental-variables generalizations; combined diagnostic tests as suggested by Godfrey and Veale (1999), as well as studies of model revision following diagnostic-test rejections under local alternatives (not false constructivism, but using the correct test for the problem); non-linearities; and cointegration reductions, to name a few already under way. Moreover, existing automations do not embody any 'expert systems' to guide formulation and selection. 'Horse races' between these approaches and others (e.g., White, 1997) would be valuable. Little is known about cases where programs like *PcGets* might flounder (perhaps extremely collinear problems where many small effects matter). Nevertheless, the entrance to this particular gold mine is now wide open, putting model selection firmly on the research agenda with prospects of success, and suggesting that 'data mining' might yet end being a compliment for an empirical study.

References

Backhouse, R., and Salanti, A. (eds.)(2000). *Macroeconomics and the Real World*. Oxford: Oxford University Press.

Bock, M. E., Yancey, T. A., and Judge, G. C. (1973). Statistical consequences of preliminary test estimators in regression. *Journal of the American Statistical Association*, **68**, 109–116.

Coppock, T. (ed.)(1999). *Information Technology and Scholarship*. Oxford: Oxford University Press.

d'Autume, A., and Cartelier, J. (eds.)(1996). *L'Economie Devient-elle une Science Dure?* Paris: Economica. Reprinted in English as: Is Economics Becoming a Hard Science? Edward Elgar, 1997.

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.

Doornik, J. A. (1998). *Object-Oriented Matrix Programming using Ox 2.0*. London: Timberlake Consultants Press.

Godfrey, L. G., and Veale, M. R. (1999). Alternative approaches to testing by variable addition. Mimeo, York University, UK.

Granger, C. W. J. (ed.)(1990). *Modelling Economic Series*. Oxford: Clarendon Press.

Hansen, B. E. (1999). Discussion of *'Data mining reconsidered'*. *Econometrics Journal*, **2**, 26–40.

Hendry, D. F. (1976). The structure of simultaneous equations estimators. *Journal of Econometrics*, **4**, 51–88.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University 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**, 41–58.

Hendry, D. F., Leamer, E. E., and Poirier, D. J. (1990). A conversation on econometric methodology. *Econometric Theory*, **6**, 171–261.

Hoover, K. D., and Perez, S. J. (1999). Data mining reconsidered: Encompassing and the general-to-specific approach to specification search. *Econometrics Journal*, **2**, 1–25.

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

Krolzig, H.-M., and Hendry, D. F. (1999). Computer automation of general-to-specific model selection procedures. Mimeo, Economics Department, Oxford University.

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

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

Sullivan, R., Timmermann, A., and White, H. (1998). Dangers of data-driven inference: The case of calendar effects in stock returns. Mimeo, Economics Department, University of California at San Diego.

White, H. (1997). Data snooping with confidence. Mimeo, Economics Department, University of California at San Diego.