Guest Editors' Introduction: Model Selection and Evaluation in Econometrics

The 13th $(EC)^2$ meeting was held at the University of Bologna, 13–14 December 2002, with the theme *Model Selection and Evaluation*. This collection of papers originates from that conference. $(EC)^2$ is an acronym for European Conference of the Econom(etr)ics Community, focusing on *Econometrics and Quantitative Economics*. The programme chair for this meeting was Niels Haldrup with Herman K. van Dijk acting as co-chairman and with Renzo Orsi as the local organizer.

The series of $(EC)^2$ conferences was established in 1990, and each is organized about a particular theme, taking place over 2 days. Originally the conferences were intended to be relatively small in scale with no parallel sessions. To give as many participants as possible a chance to present their work, two poster sessions are also included. Given the success of the meetings, there has been an increasing number of participants at the conferences. In Bologna, there were almost 150 registered participants, and because of the growing interest in $(EC)^2$ meetings it has become necessary to turn down more papers for presentation than we would like. Although around 130 papers were submitted for presentation, only 14 contributing (plus four invited) papers were presented in the plenary programme, with 35 papers in poster sessions. Over the years the popularity of the meetings has also attracted many more participants from outside Europe. At the 2002 Bologna meeting, there were participants from 20 countries, enhancing our forum for European researchers in 'Quantitative Economics and Econometrics'. The conference series also sought to provide a forum where junior participants and senior invited presenters could meet and discuss their research, and the meetings have proved very successful in this respect. In Bologna, five of the 18 papers in the plenary programme were presented by younger researchers who were also well represented in the poster sessions.

This *Bulletin* special issue is on the same theme as the 2002 conference, namely *Model Selection and Evaluation* and includes 11 papers from the conference: merely page constraints precluded a larger selection of the papers from the conference. The papers reported here, after rigorous reviewing and

revision, cover a number of important areas within the theme of the conference. Model selection and evaluation are at the very core of econometric model building. Unless one postulates an omniscient investigator, models must be chosen in the light of data evidence and checked that these adequately characterize that evidence. The papers in this special issue address both problems, which are themselves closely related against a range of criteria, and this is usually one aspect of selection. Evaluation has a long and distinguished pedigree in econometrics from early tests such as the famous Durbin–Watson statistic for residual serial correlation (see Durbin and Watson, 1950, 1951) and the Chow (1960) test for parameter constancy. Selection has a more controversial history, and despite important contributions from Anderson (1962, 1971) (*inter alia*), was often deemed a rather disreputable aspect of empirical practice (see for example Coen, Gomme and Kendall, 1969 and the ensuing discussion; Leamer, 1978; Lovell, 1983). Nevertheless, a large literature on selection criteria has flourished, including Akaike (1969, 1973), Schwarz (1978) and Hannan and Quinn (1979). More recently, however, it has become the subject of intensive study in both statistics and econometrics, particularly computer-based or automatic methods; among many possible examples, see Phillips (1994, 1996), Hoover and Perez (1999, 2000), Hendry and Krolzig (2001, 2003), and Hjort and Claeskens (2003). Thus, our topic is timely.

    Traditionally, statistical and econometric models are formulated in terms of their first one or two moments, i.e. the conditional means or variances. A large number of different models can be defined, each of which relies on certain concepts and characteristics. These notions are relevant in relation to forecasting, models of regime switching, models of common features such as cointegration, models of seasonality, trends, breaks etc. In the paper *Time Series Concepts for Conditional Distributions*, Clive Granger examines the potential for extending traditional time series concepts to conditional and unconditional distributions, and in doing so defines a new class of models. Certain concepts prove easier to generalize than others. For several decades, Clive Granger has stimulated the econometrics profession by highlighting directions for future econometrics research, and his paper here points to an interesting topic for model selection in a broader class of models than is traditionally considered. We are delighted to add our congratulations to Clive on sharing the Nobel Prize in economics for 2003 with Rob Engle.

    The estimation of Markov-switching models for regime changes has gained increasing attention over the past 15 years since the seminal paper by Hamilton (1989). However, a criticism of many empirical applications of such models is that no sufficient tools are available for their evaluation and comparison. In their paper, *Specification Testing of Markov-switching Models*, Robert Breunig, Serinah Najarian, and Adrian Pagan contribute to the development of model selection and evaluation tools for such models. Formal tests to address the

goodness-of-fit of Markov-switching models are proposed based on model consistency criteria, as well as parametric and non-parametric encompassing. A combination is made with informal tests using simulation and non-parametric density and conditional mean estimation. Several examples demonstrate the usefulness of these model evaluation methods on real data.

Dick van Dijk and Philip Hans Franses elaborate further on the evaluation of (general) non-linear models based on their forecasting performance. In their paper, *Selecting a Nonlinear Time Series Model using Weighted Tests of Equal Forecast Accuracy*, they suggest a forecast evaluation methodology where different weights are given to different forecasts. Their idea is motivated by the fact that in many situations, specific observations are more important to predict accurately than others. This applies particularly for extreme observations, and to capture this feature they suggest a weighted Diebold and Mariano (1995) type test which zooms in on the tails of the unconditional distribution of the variable of interest when evaluating forecasts from competing models. Their methodology is examined both by simulation and by a real data application. It turns out that by using a weighted evaluation criterion of the forecasts, standard criteria will be outperformed in the sense that the true non-linear models are more often found to perform better in out-of-sample forecasting than a benchmark linear model.

Graph-theoretic methods in the search for causal relationships have been used in other disciplines over the past 20 years but are relatively unfamiliar to economists, with Swanson and Granger (1997) being an exception. In their paper, *Searching for the Causal Structure of a Vector Autoregression*, Selva Demiralp and Kevin Hoover extend these methods to VAR models. Traditionally, the identification of causal orderings of structural VAR models relies on *a priori* knowledge, in the sense that the contemporaneous causations do not have a statistical foundation. The advantage of using graph-theoretic procedures is that these help identify the contemporaneous causal orderings of the data. Demiralp and Hoover demonstrate via a simulation study how a particular algorithm can be used to select – or at least to reduce to a narrower class – amongst a range of admissible causal orders.

Recently, much research has been initiated in the use of automated model selection procedures, taking advantage of our access to computer power. One such example is the *PcGets* programme developed by David Hendry and Hans-Martin Krolzig. In *General-to-Specific Model Selection Procedures for Structural Vector Autoregressions*, Hans-Martin Krolzig demonstrates how recent advances in *Gets* modelling can be used in model selection amongst structural VAR models, where traditionally a vast number of parameters and potential model specifications are employed. Using Monte Carlo experiments, it is argued that *Gets* modelling is generally a successful tool in the selection of SVAR models. The impulse responses generated from the selected SVAR

are found to be more precise and accurate than those of the unrestricted VAR. The procedure is demonstrated in an empirical application to the US monetary system. The paper further outlines future avenues for research including the extension to non-stationary and cointegrated VAR models.

The paper by Julia Campos, David Hendry, and Hans-Martin Krolzig, *Consistent Model Selection by an Automatic Gets Approach*, further addresses the properties of the *PcGets* model selection procedure. In particular, the authors establish the consistency of the selection procedures embodied in *PcGets* and compare them with model selection based on information criteria such as the Hannan–Quinn (HQ), Akaike (AIC), and Schwartz (SIC) criteria by calculating the implicit significance levels of the latter. It turns out that the *PcGets* liberal and conservative algorithms coincide in large samples with those implicit in the consistent HQ and SIC criteria, respectively. Hence, both *PcGets* rules are consistent under the same conditions as HQ and SIC. In small samples, however, *PcGets* has a rather different behaviour. One way of enhancing the performance of selection approaches is by pre-selection to remove the least significant candidate variables.

In *A Flexible Tool for Model Building: the Relevant Transformation of the Inputs Network Approach (RETINA)*, Teodosio Perez-Amaral, Giampiero Gallo, and Halbert White suggest a new computer-intensive method for model building. In particular, their procedure turns out to be strong in capturing a parsimonious representation of the mean of a variable conditional on a potentially large set of variables of interest, in situations where one does not have strong priors concerning the form of a suitable function linking the available information. Although the procedure has certain limitations with respect to the types of models that can be scrutinized, it is shown by simulation how computerized selective search within a wide range of possible models can be successful in retrieving the DGP that generated the data. Horse races between RETINA and other automated model selection procedures such as *PcGets* remain for future research.

Peter Hansen, Asger Lunde, and James Nason contribute to this special issue with their paper *Choosing the Best Volatility Models: The Model Confidence Set Approach*. This tool for model selection, MCS in short, is based on the idea of selecting the 'best' forecasting model amongst a larger set of models, where 'best' is in the sense of the model producing the minimum expected loss, with the loss specified by the researcher. The MCS obtains the best model with a given level of confidence, and the set of models in an MCS can be interpreted in the same way that a confidence interval covers the part of the real line in which the true value of a parameter resides with a certain significance level. In their paper, the MCS approach is applied to a range of volatility models for US stock returns. The MCS is able to separate superior volatility models under MSE and MAD loss functions even at a 10%

significance level. The properties of the MCS approach is further examined through simulations and the tool is found to be rather powerful in selecting the best set of forecasting models within a broader class.

Rodney Strachan and Herman van Dijk address some problems and possible solutions on the topic of Bayesian model selection and evaluation in their paper *Bayesian Model Selection with an Uninformative Prior*. In a Bayesian approach to model selection and evaluation, equal prior probabilities are often assigned to alternative models in order to compare the weight of the evidence of their respective likelihoods. Flat prior densities on the parameters are often used for the same reason. However, resulting posterior probabilities for model comparison are, in such a case, not well defined. This result is known as the Bartlett paradox. Using concepts from vector and matrix spaces, in particular Stiefel manifolds, Strachan and Van Dijk, building on previous work, argue that a model specification can be found whereby a uniform prior is proper, and in the context of cointegration models, a more sensible representation of 'ignorance' can be provided in the specification of prior beliefs. Their approach may also be useful for factor models or other multivariate models where linear combinations of variables are used.

In *An Introduction to Best Empirical Models when the Parameter Space is Infinite Dimensional*, Werner Ploberger and Peter Phillips extend to the infinite dimensional parameter space case some of their earlier work, demonstrating a bound on how close a fitted empirical model can get to the true model when the model is represented by a parametrized probability measure. Their results have implications for model choice in infinite dimensional problems and telescopes on some difficulties presented by models of infinite dimension. Some implications for forecasting are considered and applications are provided, including the empirically relevant case of VAR models of infinite order. Their paper derives results for the stationary case, and serves as an interesting step in the analysis of model choice with infinite dimensional systems, raising questions worthy of future research.

Evaluation of model assumptions is critical in model building. In the paper *Exact Skewness–Kurtosis Tests for Multivariate Normality and Goodness-of-fit in Multivariate Regressions with Application to Asset Pricing Models*, Jean-Marie Dufour, Lynda Khalaf and Marie-Claude Beaulieu address the issue stated in the title of their paper. They propose a class of exact procedures for testing goodness-of-fit of the error distribution in possibly non-Gaussian multivariate linear regression models. For the Gaussian case, their procedures include finite-sample versions of the standard Jarque and Bera (1980) and Mardia and Foster (1983)-type multivariate skewness and kurtosis tests, in addition to some new ways of combining skewness and kurtosis measures. Because the distribution of the latter is analytically intractable, implementation via Monte Carlo methods is suggested. For non-Gaussian error distributions,

nuisance parameter problems are solved through test 'inversion'. The suggested tests for model evaluation are applied to monthly returns of NYSE portfolios.

To conclude, we would like to thank all contributors and participants of the 13th $(EC)^2$ conference who made the Bologna meeting a great success. We appreciate the help from almost 40 referees who assisted in organizing this volume with tight deadlines. Thanks are due to the local organizer Renzo Orsi and the Department of Economics, University of Bologna, for hosting the conference, and to *Journal of Applied Econometrics*, CIDE, Alma Master Studiorum Universitá de Bologna, and UniCredit Banca for providing financial support. We are also grateful to the *Bulletin* Board of Editors for inviting us to act as guest editors of this volume.

NIELS HALDRUP
Department of Economics, University of Aarhus,
Denmark

DAVID F. HENDRY
Nuffield College, Oxford University, UK

HERMAN K. VAN DIJK
Econometric Institute, Erasmus University Rotterdam,
The Netherlands

References

Akaike, H. (1969). 'Fitting autoregressive models, for prediction', *Annals of the Institute of Statistical Mathematics*, Vol. 21, pp. 243–247.

Akaike, H. (1973). 'Information theory and an extension of the maximum likelihood principle', in Petrov B. N. and Csaki F. (eds), *Second International Symposon on Information Theory*, Akademia Kiado, Budapest, pp. 267–281.

Anderson, T. W. (1962). 'The choice of the degree of a polynomial regression as a multiple-decision problem', *Annals of Mathematical Statistics*, Vol. 33, pp. 255–265.

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

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

Coen, P. G., Gomme, E. D. and Kendall, M. G. (1969). 'Lagged relationships in economic forecasting', *Journal of the Royal Statistical Society A*, Vol. 132, pp. 133–163. Reprinted in Mills T.C. (ed.), *Economic Forecasting*. Edward Elgar, 1999.

Diebold, F. X. and Mariano, R. S. (1995). 'Comparing predictive accuracy', *Journal of Business and Economic Statistics*, Vol. 13, pp. 253–263. Reprinted in Mills T. C. (ed.), *Economic Forecasting*. Edward Elgar, 1999.

Durbin, J. and Watson, G. S. (1950). 'Testing for serial correlation in least squares regression I', *Biometrika*, Vol. 37, pp. 409–428.

Durbin, J. and Watson, G. S. (1951). 'Testing for serial correlation in least squares regression II', *Biometrika*, Vol. 38, pp. 159–178.

Hamilton, J. D. (1989). 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica*, Vol. 57, pp. 357–384.

Hannan, E. J. and Quinn, B. G. (1979). 'The determination of the order of an autoregression', *Journal of the Royal Statistical Society, B*, Vol. 41, pp. 190–195.

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

Hendry, D. F. and Krolzig, H.-M. (2003). 'New developments in automatic general-to-specific modelling', in Stigum B. P. (ed.), *Econometrics and the Philosophy of Economics*, pp. 379–419. Princeton University Press, Princeton.

Hjort, N. L. and Claeskens, G. (2003). *Frequentist Model Average Estimators*. Working paper, Department of Statistics, University of Oslo, Norway.

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

Jarque, C. M. and Bera, A. K. (1980). 'Efficient tests for normality, homoscedasticity and serial independence of regression residuals', *Economics Letters*, Vol. 6, pp. 255–259.

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

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

Mardia, K. V. and Foster, K. (1983). 'Omnibus tests of multinormality based on skewness and kurtosis', *Communications in Statistics B (Simulation and Computation)*, Vol. 12, pp. 207–221.

Phillips, P. C. B. (1994). 'Bayes models and forecasts of Australian macroeconomic time series', in Hargreaves C. (ed.), *Non-stationary Time-series Analysis and Cointegration*, pp. 53–86. Oxford University Press, Oxford.

Phillips, P. C. B. (1996). 'Econometric model determination', *Econometrica*, Vol. 64, pp. 763–812.

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

Swanson, N. and Granger, C. (1997). 'Impulse response functions based on causal approach to residual orthogonalization in vector autoregressions', *Journal of the American Statistical Association*, Vol. 92, pp. 357–367.

List of Contributors to *Oxford Bulletin of Economics and Statistics Special Issue on Model Selection and Evaluation*

Dick van Dijk, Erasmus University, Rotterdam
Philip Hans Franses, Erasmus University, Rotterdam
Selecting a Nonlinear Time Series Model using Weighted Tests of Equal Forecast Accuracy

Selva Demiralp, Federal Reserve Bank, Washington
Kevin Hoover, University of California, Davis
Searching for the Causal Structure of a Vector Autoregression

Hans-Martin Krolzig, Nuffield College, Oxford University
General-to-Specific Model Selection Procedures for Structural Vector Auto-regressions

Julia Campos, University of Salamanca
David Hendry, Nuffield College, Oxford University
Hans-Martin Krolzig, Nuffield College, Oxford University
Consistent Model Selection by an Automatic Gets Approach

Teodosio Perez-Amaral, Universidad Complutense de Madrid
Giampiero Gallo, Università di Firenze
Halbert White, University of California, San Diego
A Flexible Tool for Model Building: The Relevant Transformation of the Inputs Network Approach (RETINA)

Peter R. Hansen, Brown University
Asger Lunde, Aarhus School of Business
James Nason, Federal Reserve Bank, Atlanta
Choosing the Best Volatility Models: The Model Confidence Set Approach

Rodney Strachan, University of Liverpool
Herman K. van-Dijk, Erasmus University, Rotterdam
Bayesian Model Selection with an Uninformative Prior

Werner Ploberger, University of Rochester
Peter C. B. Phillips, Cowles Foundation, Yale University, University of Auckland & University of York
An Introduction to Best Empirical Models when the Parameter Space is Infinite Dimensional

Jean-Marie Dufour, CIRANO, CIREQ, and Université de Montréal
Lynda Khalaf, CIREQ, and GREEN, Université Laval

Marie-Claude Beaulieu, CIRPÉE, CIRANO, and Université Laval
Exact Skewness–Kurtosis Tests for Multivariate Normality and Goodness-of-fit in Multivariate Regressions with Application to Asset Pricing Models