Real Time Econometrics

Hashem Pesaran*
Cambridge University and USC

Allan Timmermann
University of California, San Diego

March 28, 2004

Abstract

This paper considers the problems facing decision makers using econometric models in real time. It identifies the key stages involved and highlights the role of automated systems in reducing the effect of data snooping. It sets out many choices that researchers face in construction of automated systems and discusses some of the possible ways advanced in the literature for dealing with them. The role of feedbacks from the decision maker's actions to the data generating process is also discussed and highlighted through an example.

JEL Classification: C51, C52, C53.
Key Words: Specification Search, Data Snooping, Recursive/Sequential Modelling, Automated Model Selection.

1 Introduction

The focus of econometric analysis has predominantly been on retrospectively estimating and interpreting economic relationships. However, in many uses of econometric models by businesses, governments, central banks and traders in financial markets the focus is on making decisions in real time and hence there is an urgent need to develop robust interactive systems that use econometric models to guide decisions in real time. This use of econometric methods is of course not without precursors and is closely related to the development of sequential analysis pioneered by Wald (1947), and recursive estimation methods proposed by Brown, Durbin and Evans (1975) for use in real time quality control.[1]

*We are grateful to Peter Phillips, Clive Granger, Ron Smith, Steven Durlauf, Mardi Dungey and an anonymous referee for their helpful comments. Hashem Pesaran is particularly grateful to Bahram Pesaran for his help with practicing some of the ideas discussed in this paper.

[1] Sequential or recursive techniques have also been used extensively in biomedical research and engineering. Dawid (1984) has been advocating the use of probability forecasting in sequential statistical analysis, known as the "prequential approach".

Automated systems are bound to become an important tool for the development of real-time econometrics and the dialog between Granger and Hendry (Granger and Hendry (2004)) is a pertinent reminder of many of the challenges facing designers and users of such systems. By setting out in advance a set of rules for observation windows and variable selection, estimation and modification of the econometric model, automation provides a way to reduce the effects of data-snooping and facilitates learning from the performance of a given model when applied to a historical data set.

In this paper we set out what we see as the three key stages in real-time automated econometric modeling, namely model construction, monitoring, modification and the model innovation process. The first stage, model construction, sets out the objective function, chooses the estimation method, selects variables and functional forms and seeks to cross-validate the "best" or the "average" model. The second stage sets out procedures for monitoring and evaluating the performance of the econometric model in real time and, if deemed necessary, undertakes the necessary modifications. This involves deciding on statistical decision-based evaluation criteria used to determine whether the performance of the econometric model is satisfactory. Finally, the third stage involves deciding which alternative new modelling strategies to consider in case of unsatisfactory performance and even a criterion for dispensing with the econometric modelling approach altogether if there is systematic evidence of a breakdown.

The plan of the paper is as follows. As a way of furnishing a concrete illustration of the issues involved in real time econometric modeling, Section 2 considers a relatively simple, one-period portfolio decision problem from finance. Section 3 sets out the forecasting and decision problem that underlie automated systems. Sections 4 and 5 consider model construction, monitoring, modification and innovation. Section 6 outlines some future areas of research.

2 An Illustration from Finance

Consider a portfolio of long and short (dollar) positions at time T, represented by the $s \times 1$ vector $\mathbf{d}_T$, and denote the $s \times 1$ vector of returns associated with the underlying assets over the subsequent period by $\mathbf{r}_{T+1}$. Abstracting from transaction costs, the value of this portfolio at the end of $T+1$ is given by

$$V_{T+1} = \mathbf{d}_T' \mathbf{r}_{T+1}. \tag{1}$$

Suppose the loss (the negative of utility) function of some trader is of the constant absolute risk aversion type:

$$\ell(\mathbf{d}_T, \mathbf{r}_{T+1}, \phi) = \exp(-\phi \ \mathbf{d}_T' \mathbf{r}_{T+1}), \ \ \phi > 0, \tag{2}$$

with the expected loss computed conditional on the information set, $\mathcal{F}_T$, that contains $\mathbf{r}_T$, $\mathbf{d}_T$ and their lagged values as well as observations on a number of other state variables (besides $\mathbf{r}_T$) which we denote by $\mathbf{z}_T$. Further, for the time being suppose that the risk aversion coefficient, ϕ, is sufficiently large so that one can safely assume that changes in a trader's short and long positions will not influence the probability distribution of asset returns, $\mathbf{r}_{T+1}$, conditional on $\mathcal{F}_T$, namely there are no market impacts from the trades: $f(\mathbf{r}_{T+1}|\mathbf{r}_T, \mathbf{d}_T, \mathcal{F}_{T-1}) = f(\mathbf{r}_{T+1}|\mathbf{r}_T, \mathcal{F}_{T-1})$.

In the presence of model and parameter uncertainty a trader faces a number of important choices:

- Which probability distributions to consider for returns? For example whether to assume the returns have a multi-variate Gaussian distribution, a mixture of multi-variate normal densities, or opt for multi-variate Student t. The form of the probability distribution of asset returns (particularly in the tails) has important implications for risk management, although it is true that many estimation and model selection procedures are robust (in large samples) to the choice of the distribution of asset returns.

- How to model mean returns? There are many factors that could be considered, say as linear or non-linear functions of $\mathbf{x}_T = \{\mathbf{r}_T, \mathbf{z}_T\}$ and possibly their lagged values; and it seems unlikely that the same set of candidate factors would have been relevant historically. The sheer size of the possible factors that could be considered, the relatively short time series that are typically available,[2] and the time constraint that often exists between the collection and compilation of data at close of one market and the issuance of trade orders for execution at the start of another market, places important restrictions on the search process across the possible factors and how it is implemented. This might introduce a certain degree of randomization in model selection (model averaging) procedures, and search algorithms such as the recursive modelling method, to be reviewed below.

- Which specifications to consider for modelling of asset return volatilities? In addition to the search problems discussed above, one is also faced with the problem of dealing with large correlations typically encountered across asset returns and the possibility that such correlations may vary across time, rising at times of crisis and falling in calmer periods.

Risk management considerations also dictate that the loss function (2) is augmented with the following Value at Risk (VaR) constraint:

$$\Pr\left(\mathbf{d}_T'\mathbf{r}_{T+1} < -V_\alpha \left| \mathcal{F}_T \right.\right) \leq \alpha, \tag{3}$$

where V_α is the maximum loss tolerated over any one trading day with probability α (often taken to be 1%). It is clear that the minimization of the expected loss function in (2) subject to the VaR constraint in (3) requires a complete

[2] At daily frequencies large data sets are available for some assets, such as Yen and Euro dollar rates, but even in the case of these assets one might not wish to use all the data available in forecasting due to the possibility of structural change in asset markets; a topic which we shall return to under the choice of "Data Window".

specification of the probability distribution of asset returns. To simplify the exposition we suppose that in the search the trader focuses on the class of multi-variate Gaussian specifications drawn from the set $\mathcal{M}_T$:

$$\mathbf{r}_{T+1}|\mathcal{F}_T, M_i \sim N(\boldsymbol{\mu}_{iT}, \boldsymbol{\Sigma}_{iT}),\ M_i \in \mathcal{M}_T,$$

as well as the Bayesian model average and other combinations. $\boldsymbol{\mu}_{iT}$ and $\boldsymbol{\Sigma}_{iT}$ denote the time-T conditional mean and conditional variance-covariances of asset returns, respectively, assuming model M_i holds. Note that $\boldsymbol{\mu}_{iT}$ and $\boldsymbol{\Sigma}_{iT}$ would typically depend on one or more elements of $\mathbf{x}_T$ and their lagged values with a number of unknown parameters that need to be estimated or integrated out of the decision process. Abstracting from parameter uncertainty we first note that, under M_i,

$$E[\ell(\mathbf{d}_T, \mathbf{r}_{T+1}, \phi)|\mathcal{F}_T, M_i] = \exp(-\phi\mathbf{d}_T'\boldsymbol{\mu}_{iT} + \frac{1}{2}\phi^2\mathbf{d}_T'\boldsymbol{\Sigma}_{iT}\mathbf{d}_T), \tag{4}$$

and

$$\Pr\left(\mathbf{d}_T'\mathbf{r}_{T+1} < -V_\alpha \,|\mathcal{F}_T, M_i\,\right) = \Phi\left(\frac{-V_\alpha - \mathbf{d}_T'\boldsymbol{\mu}_{iT}}{\sqrt{\mathbf{d}_T'\boldsymbol{\Sigma}_{iT}\mathbf{d}_T}}\right) \leq \alpha,$$

where $\Phi(.)$ stands for the cumulative distribution function of a standard normal variate. Denoting the $\alpha\%$ left tail of the standard normal distribution by $-c_\alpha$, the VaR constraint can also be written as

$$V_\alpha + \mathbf{d}_T'\boldsymbol{\mu}_{iT} - c_\alpha\sqrt{\mathbf{d}_T'\boldsymbol{\Sigma}_{iT}\mathbf{d}_T} \geq 0. \tag{5}$$

The Lagrangian for minimizing $-\phi\mathbf{d}_T'\boldsymbol{\mu}_{iT} + \frac{1}{2}\phi^2\mathbf{d}_T'\boldsymbol{\Sigma}_{iT}\mathbf{d}_T$ subject to (5) is

$$\mathcal{L}_i\left(\mathbf{d}_T\right) = -\phi\mathbf{d}_T'\boldsymbol{\mu}_{iT} + \frac{1}{2}\phi^2\mathbf{d}_T'\boldsymbol{\Sigma}_{iT}\mathbf{d}_T - \lambda\left(V_\alpha + \mathbf{d}_T'\boldsymbol{\mu}_{iT} - c_\alpha\sqrt{\mathbf{d}_T'\boldsymbol{\Sigma}_{iT}\mathbf{d}_T}\right). \tag{6}$$

Since there are s positions to be determined with one VaR constraint we proceed initially by ignoring the VaR constraint (setting λ, the Lagrange multiplier, to zero). In this case the optimal positions under model M_i are given by

$$\mathbf{d}_{iT}^* = (1/\phi)\boldsymbol{\Sigma}_{iT}^{-1}\boldsymbol{\mu}_{iT}. \tag{7}$$

We now check to see under which conditions these positions also satisfy the VaR constraint (3). Substituting (7) in (5) we have

$$s_{iT}^2 - c_\alpha s_{iT} + \phi V_\alpha \geq 0, \tag{8}$$

where $s_{iT} = \left(\boldsymbol{\mu}_{iT}'\boldsymbol{\Sigma}_{iT}^{-1}\boldsymbol{\mu}_{iT}\right)^{1/2}$ is the multi-variate Sharpe ratio (in absolute value) under model M_i. Therefore, the positions given in (7) satisfy the VaR constraint if the risk aversion coefficient is sufficiently large such that

$$\phi \geq \frac{c_\alpha^2}{4V_\alpha},$$

otherwise the optimal solution that satisfies the VaR would be

$$\mathbf{d}_{iT}^{*} = \left(\frac{4V_{\alpha}}{c_{\alpha}^{2}}\right)\boldsymbol{\Sigma}_{iT}^{-1}\boldsymbol{\mu}_{iT}.$$

Combining the two solutions, we have

$$\mathbf{d}_{iT}^{*} = \gamma\boldsymbol{\Sigma}_{iT}^{-1}\boldsymbol{\mu}_{iT}, \tag{9}$$

where

$$\gamma = Min\left(\frac{4V_{\alpha}}{c_{\alpha}^{2}}, \frac{1}{\phi}\right).$$

The realized value of the loss function under model M_i is now given by

$$\ell(\hat{\mathbf{d}}_{iT}, \mathbf{r}_{T+1}, \phi, V_{\alpha}, c_{\alpha}) = \exp(-\phi\gamma\ \hat{\boldsymbol{\mu}}_{iT}'\hat{\boldsymbol{\Sigma}}_{iT}^{-1}\mathbf{r}_{T+1}), \tag{10}$$

where we have replaced $(\boldsymbol{\mu}_{iT}, \boldsymbol{\Sigma}_{iT})$ by their estimates based on a given data or observation window. A decision-based evaluation exercise can now be carried out whereby the performance of the alternative models are compared in terms of their associated realized losses over a given evaluation period, $(R+1, T)$:

$$\bar{\ell}_{iR,T} = (T-R)^{-1}\sum_{t=R+1}^{T}\exp(-\phi\gamma\ \hat{\boldsymbol{\mu}}_{i,t-1}'\hat{\boldsymbol{\Sigma}}_{i,t-1}^{-1}\mathbf{r}_{t}), \tag{11}$$

where $\hat{\boldsymbol{\mu}}_{i,t-1}$ and $\hat{\boldsymbol{\Sigma}}_{i,t-1}$ are the recursive estimates of $\boldsymbol{\mu}_{i,t-1}$ and $\boldsymbol{\Sigma}_{i,t-1}$ based on observations up to and including time $t-1$.[3]

2.1 Feedback Effects

Although feedback effects are likely to be less important in trading than in macroeconomic decision making, i.e., when considering the effect of monetary policy, they can and do exist in trading by large financial corporations, and the above application can be readily adapted to show the additional issues that could arise when the possibility of feedbacks from $\mathbf{d}_t$ to $\mathbf{x}_t$ is recognized by the decision maker. For simplicity of exposition let $s=1$, and suppose that such feedback only takes place through the effects of changes in trade positions on returns. Assume also that the change in the trade position decided at the end of period T is filled some time during the interval $(T; T+1]$, $\Delta d_{T+1} = d_{T+1} - d_T = d_T^{*} - d_T$, or more generally $d_{t+1} = d_t^{*}$ for all t. Under this set up a simple model of returns that allows for feedback is given by

$$M_{i,T}: r_{T+1} = \psi_i\left(d_{T+1} - d_T\right) + \boldsymbol{\beta}_i'\mathbf{x}_{iT} + \varepsilon_{T+1,i},\ \varepsilon_{T+1,i} \sim N(0, \sigma_{iT}^{2}) \tag{12}$$

where $M_{i,T}$ denotes the return regression model assumed by the trader to hold at the end of period T. The coefficient $\psi_i \geq 0$, measures the trader's perception

[3] Decision-based evaluation techniques are discussed in Granger and Pesaran (2000a,b) and reviewed and developed further in Pesaran and Skouras (2002).

of his market impact under $M_{i,T}$.[4] For simplicity we shall assume that ψ_i is fixed, although allowing for time variations in ψ_i is likely to strengthen our arguments.

Abstracting from VaR considerations, in this case we have

$$d_{T+1} = \left(\frac{-\psi_i}{\phi \sigma_{iT}^2 - 2\psi_i} \right) d_T + \frac{\boldsymbol{\beta}_i^{'} \mathbf{x}_{iT}}{\phi \sigma_{iT}^2 - 2\psi_i}, \tag{13}$$

Under zero market impact, $\psi_i = 0$, this solution reduces to the standard mean-variance result, $\boldsymbol{\beta}_i^{'} \mathbf{x}_{iT} / \phi \sigma_{iT}^2$. To implement d_{T+1}, the trader would need to estimate the parameters $\{\psi_i, \boldsymbol{\beta}_i\}$, and the return volatility, σ_{iT}^2. Since these parameters are specific to the model used by the trader they can, at best, be estimated for the particular model that had been used by the trader during the estimation period prior to date T. In reality, the trader might have used many different models in the past, and in that case historical observations on d_t, r_t and $\mathbf{x}_t$, $t = T, T-1, ...$, would not be relevant to the estimation of the parameters of model $M_{i,T}$. The problem arises more clearly if the trader in fact wishes to consider a counter-factual exercise with respect to any model, say $M_{j,T}$, which has been used by him/her in the past. It is not clear how historical data could be used by the trader to estimate the market impact coefficient, ψ_j, under $M_{j,T}$. One way to get around this problem is by generating a sample that is consistent with the counterfactual policy. Such experiments can be costly, though, and are typically not available for econometric estimation purposes.

This application shows many choices that a decision maker will be faced with, even if the nature of the decision problem and the space of assets to be traded are taken as given. In what follows we consider some of these issues more generally and the type of solutions suggested in the literature for dealing with them.

3 The Decision Problem

We shall focus on a single-period forecasting and decision problem but many of our comments also apply to multi-period decisions. The decision problem, as it is usually set up, pre-assumes the existence of a unique time invariant loss function, $\ell(\mathbf{d}_T, \mathbf{x}_{T+1}, \boldsymbol{\phi})$, where $\mathbf{d}_T$ is the $s \times 1$ vector of decision variables set at the end of period T, $\mathbf{x}_{T+1}$ the $k \times 1$ vector of state variables realized over the period from T to $T+1$ and $\boldsymbol{\phi}$ is a vector of fixed coefficients. Decisions are implemented during the period $(T, T+1)$.[5] It is assumed that decisions are made with respect to information available at time T, $\mathcal{F}_T$, which at least contains $\{\mathbf{d}_T, \mathbf{x}_T\}$ and their past values. The decision variables are then derived by solving the following optimization problem

$$\mathbf{d}_T^* = Arg \min_{\mathbf{d}_T \in \mathcal{D}_T} \left\{ \int_{\mathcal{R}_{T+1}} \ell(\mathbf{d}_T, \mathbf{x}_{T+1}, \boldsymbol{\phi}) \hat{f}_t \left(\mathbf{x}_{T+1} \left| \mathcal{F}_T \right. \right) d\mathbf{x}_{T+1} \right\}, \tag{14}$$

[4] It is assumed that the trader does not observe the trade positions of others in the market.

[5] To simplify the exposition we assume that the decision can be exactly implemented.

where $\mathcal{D}_T \subseteq \mathcal{R}^s$ is the set of feasible actions at time T, $\hat{f}_T(\mathbf{x}_{T+1}|\mathcal{F}_T)$ is the decision maker's time-T estimate of the probability forecast density function of $\mathbf{x}_{T+1}$ conditional on $\mathcal{F}_T$, and $\mathcal{R}_{T+1} \subseteq \mathcal{R}^k$ is the domain of variations of $\mathbf{x}_{T+1}$ assumed to be known at time T. In cases where the loss function is quadratic in $\mathbf{d}_T$ only a conditional point forecast of $\mathbf{x}_{T+1}$, namely $E_{\hat{f}_T}(\mathbf{x}_{T+1}|\mathcal{F}_T)$, will be needed, but in general the whole probability density is required.

At each decision point, T, the decision maker faces considerable uncertainty regarding the choice of $\hat{f}_T(\mathbf{x}_{T+1}|\mathcal{F}_T)$ and might also be uncertain about the choice of the loss function and/or its parameter values, as well as the measurement of some of the state variables. But in what follows we suppose that the loss function is known to the decision maker and abstract from data uncertainty. Neither complication is germane to the issues that we wish to raise here.

The decision process is further complicated when the implementation of the decision (or 'action' by the decision maker) influences the "true" conditional probability distribution of the state variables, $\mathbf{x}_{T+1}$, which we denote by $\Pr(\mathbf{x}_{T+1}|\mathcal{F}_T)$. The possibility that changes to $\mathbf{d}_T$ might affect $\Pr(\mathbf{x}_{T+1}|\mathcal{F}_T)$ has further implications for econometric model construction and evaluation, which we highlight in Section 4.

3.1 Formulation and Selection of Models

For a given loss function and a time series of past measurements, $\mathbf{Z}_T(1) = (\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_T) \subseteq \mathcal{F}_T$ on the decision and state variables $\mathbf{z}_t = (\mathbf{d}_t, \mathbf{x}_t)$, the single most important task facing the decision maker at time T is the choice of the forecasting model, $\hat{f}_T(\mathbf{x}_{T+1}|\mathcal{F}_T)$. The real time nature of the decision making process recognizes that the forecasting model and its parameters might need updating at the start of each decision period (say prior to opening of markets). This could involve simple updating of parameters of a given model (keeping the specification of the model fixed), updating the model by searching over a pre-specified set of models, or might even involve searching over new models including new variables/factors, functional forms or dynamic specifications not considered as feasible or of potential usefulness prior to date T. These three levels of model updates can be viewed as "recursive estimation", "recursive modelling" and "innovative modelling", respectively. Clearly, recursive modeling involves recursive estimation, but not *vice versa*, and innovative modelling could encompass both recursive modelling and recursive estimation.

Selection of a model for use in decision making involves providing an answer to a *counter factual exercise*, namely a comparative analysis of the losses that would have been realized in the past under alternative specifications of the forecasting model (and hence the decision rules). Assume that $\Pr(\mathbf{x}_{T+1}|\mathcal{F}_T)$, the data generating process (DGP), is unknown and suppose that at time T the decision maker is faced with a set of forecasting models, $M_i \in \mathcal{M}_T$. Each model, M_i, is defined by the conditional probability density function of $\mathbf{x}_t$ defined over the estimation period $t = T_0, T_0 + 1..., T$, $(T_0 \geq 1)$, and the forecasting period, $t = T + 1$, in terms of a $k_i \times 1$ vector of unknown parameters, $\boldsymbol{\theta}_i$, assumed to lie in the compact parameter space, $\boldsymbol{\Theta}_i$:

$$M_i: \quad \{f_i(\mathbf{x}_{t+1};\boldsymbol{\theta}_i \,|\mathcal{F}_t), \quad \boldsymbol{\theta}_i\in \boldsymbol{\Theta}_i\}. \tag{15}$$

Conditional on each model, M_i, being true it will be assumed that the true value of $\boldsymbol{\theta}_i$, which we denote by $\boldsymbol{\theta}_{i0}$, is fixed and remains constant across the estimation and prediction periods and lies in the interior of $\boldsymbol{\Theta}_i$.[6] Under M_i the solution to the decision problem (14) can then be written as

$$\mathbf{d}_{iT}^*(\boldsymbol{\theta}_i,\boldsymbol{\phi}) = \underset{\mathbf{d}_T\in\mathcal{D}_T}{\arg\min}\left\{\int_{\mathcal{R}_{T+1}} \ell(\mathbf{d}_T,\mathbf{x}_{T+1},\boldsymbol{\phi}) f_i(\mathbf{x}_{T+1};\boldsymbol{\theta}_i \,|\mathcal{F}_T)\, d\mathbf{x}_{T+1}\right\}, \tag{16}$$

which depends on the unknown parameters of the selected forecasting model, $f_i(\mathbf{x}_{t+1};\boldsymbol{\theta}_i \,|\mathcal{F}_t)$. To derive an operationally feasible decision rule, one would need to replace $\boldsymbol{\theta}_i$ with a suitable estimate, or eliminate such parameters by integrating them out with respect to their posterior distributions using Bayesian techniques. Under both approaches we have

$$\mathbf{d}_{iT}^* = \boldsymbol{\psi}_i(\mathcal{F}_T,\boldsymbol{\phi}), \tag{17}$$

where $\boldsymbol{\psi}_i(.)$ represents the vector of decision rules which depends on the loss function and its parameters, $\boldsymbol{\phi}$, the choice of the forecasting model, M_i, and the particular procedure used to deal with the unknown coefficients, $\boldsymbol{\theta}_i$.

Under M_i, the estimation of $\boldsymbol{\theta}_i$, or derivation of its posterior distribution, can be based on the joint likelihood function of $\mathbf{z}_t$ defined over the "observation window": $t = T_0, T_0+1..., T$:

$$L_i^z(\mathbf{z}_{T_0},\mathbf{z}_{T_0+1},...,\mathbf{z}_T;\boldsymbol{\phi},\boldsymbol{\theta}_i \,|\mathcal{F}_{T-1}) = \prod_{t=T_0}^{T} h_i(\mathbf{d}_t;\boldsymbol{\phi}\,|\mathcal{F}_{t-1})\, f_i(\mathbf{x}_t;\boldsymbol{\theta}_i \,|\mathbf{d}_t,\mathcal{F}_{t-1}), \tag{18}$$

where $h_i(\mathbf{d}_t;\boldsymbol{\phi}\,|\mathcal{F}_{t-1})$ is the conditional density of the decision variables assuming model M_i holds. In general, there are complicated restrictions linking the solutions to the decision and estimation problems. However, in the case of atomistic agents where the feedback effects of $\mathbf{d}_t$ on $\mathbf{x}_t$ can be ignored so that

$$f_i(\mathbf{x}_t;\boldsymbol{\theta}_i \,|\mathbf{d}_t,\mathcal{F}_{t-1}) = f_i(\mathbf{x}_t;\boldsymbol{\theta}_i \,|\mathcal{F}_{t-1}), \tag{19}$$

the estimation and decision problems are de-coupled and estimation can be based on the likelihood of the state variables alone:

$$L_i^x(\mathbf{x}_{T_0},\mathbf{x}_{T_0+1},...,\mathbf{x}_T;\boldsymbol{\theta}_i \,|\mathcal{F}_{T-1}) = \prod_{t=T_0}^{T} f_i(\mathbf{x}_t;\boldsymbol{\theta}_i \,|\mathcal{F}_{t-1}). \tag{20}$$

This in turn may require that the decision variables are constrained to lie in a restricted set so that their simulated values computed conditional on model

[6] Phillips (1996) discusses the possibility that none of the models in $\mathcal{M}_T$ represents the true DGP in which case the best (local) approximation is being sought.

M_i do not influence the state variables. In evaluation of trading strategies this requirement is usually met by restricting the trading positions to be sufficiently small or by inclusion of appropriate penalty terms in the loss function so that $\mathbf{d}_t \in \bar{\mathcal{D}}_t \subseteq \mathcal{D}_t$, where $\bar{\mathcal{D}}_t$ is a restricted subset of $\mathcal{D}_t$.

4 Specification Search, Data Snooping, and Use of Automated Techniques

The above over-view of the decision and estimation problems clearly shows a number of important choices that face the decision maker such as the choice of the forecasting model; the choice of the observation (estimation) window, the treatment of possible feedbacks from decisions onto the state variables in cases where the decisions are expected to influence the probability distribution of the state variables and comparison of the performance of alternative models, estimation procedures and observation windows. In this section we discuss the role of automated model selection for each of these points.

4.1 Choice of Forecasting Model

Econometric modelling is often guided by existing economic theory although this will typically only be broadly suggestive of which state variables or functional form to use in the econometric model. To learn about these, the econometrician will therefore in practice undertake substantial specification searches across the set of potential models, $\mathcal{M}_t$. The search for the 'best' or an 'average' model involves using the *same* set of realizations (historical data) many times over. The outcome could also depend crucially on the choice of the criterion function (likelihood or utility), the extent to which model complexity is penalized, and the particular model space over which the search is carried out. As a result the in-sample penalized performance of the model selected is not an unbiased estimator of the model's subsequent (out-of-sample) performance, and can severely underestimate the true loss.[7] This is due to the pre-test bias arising from having to rely on one set of realizations from which to choose the model specification.

One way to deal with this problem is to assess the statistical significance of the best model in the context of the specification search that preceded it. This is the approach taken, e.g., by Sullivan, Timmermann and White (2001) who bootstrap performance statistics to evaluate the significance of the best model drawn from a larger universe. This method provides a correction to the standard p-value that conditions on one model (as if it were pre-specified) and therefore ignores the search that preceded it.

Automation can assist in reducing, but not eliminating, the effects of data-snooping and pre-test bias. Subjective opinions about specific variables and functional forms are likely to benefit significantly from hindsight as such opinions will be formed after gaining experience with different modelling strategies on

[7] Leamer (1978) was one of the first in the econometrics literature to highlight this problem.

an existing data set. Automated data-driven model selection procedures do, to some extent, take the choice of model out of the hand of the modeler, although it is more difficult to reduce hindsight in the choice of the model set, $\mathcal{M}_t$, or in the innovative modelling stage, i.e. when choosing how to revise and modify $\mathcal{M}_t$ over time.[8]

4.2 Variable Selection

Let $\mathcal{X}$ denote the time-invariant universe of all possible prediction variables that could be considered in the econometric model, while N_{xt} is the number of regressors available at time t so $\mathbf{X}^t = (x_{1t}, ..., x_{N_{xt}}) \subseteq \mathcal{X}$. N_{xt} is likely to grow at a faster rate than the sample size, T. At some point there will therefore be more regressors than time-series observations. However, most new variables will represent different measurements of a finite number of underlying economic factors such as output/activity, inflation and interest rates.

Rather than searching over all possible combinations of predictor variables at random, a sensible approach is to first categorize variables and then choose one or a few variables from each category. Alternatively, regressors can be clustered in advance according to a simple algorithm that measures their correlation or by grouping them according to what they measure (e.g. interest rates, inflation measures). One possibility, considered, for example, by Stock and Watson (2002) is to extract common factors and use the most important of these in forecasting.

4.3 Recursive Modelling

In Pesaran and Timmermann (1995, 2000) we proposed a recursive modelling approach that exemplifies many of the points mentioned thus far including how to estimate $\boldsymbol{\mu}_{it}$ in Section 2. Consider an $N_{xt} \times 1$ column vector $\mathbf{v}_i$ with a string of N_{xt} ones or zeros, where a one in the j'th row means that the j'th regressor is included in the model whereas a zero in the j'th row means that this regressor is excluded from the model. Then each possible model at time t can be identified by the N_{xt}-digit string of zeros and ones corresponding to the binary code of its number. Without loss of generality, consider forecasting the first element of $\mathbf{X}_{t+1}, Y_{t+1} = \mathbf{e}'_{1t}\mathbf{X}_{t+1}$ (where $\mathbf{e}'_{1t}$ is an $N_{xt}-$vector with unity in the first position and zeros elsewhere) at time $t+1$ by means of linear regressions

$$M_i(\tau): \; Y_{\tau+1} = \boldsymbol{\beta}_i^{'}\mathbf{X}_{\tau,i} + \epsilon_{\tau+1,i}, \quad \tau = 1, 2, ..., T, \tag{21}$$

where $\mathbf{X}_{\tau,i}$ is a vector of regressors obtained as a subset of the regressors in contention, $\mathbf{X}_\tau$, and $M_i(\tau)$ denotes the i'th regression model at time τ. Notice that this gives a total of $2^{N_{xT}}$ possible models. When N_{xT} is large, a comprehensive

[8] Clearly, the extent to which biases in subjective choices can be reduced depends on the dimensionality of $\mathcal{M}_t$. If $\mathcal{M}_t$ only contains a few explanatory variables and functional forms that the modeler (with the benefit of hindsight) knows work well, then data snooping biases will not be reduced by much. Only if the data and model set is sufficiently large will the hindsight in the choice of a particular model be reduced.

(global) search is therefore not feasible and the approach could be modified by using Monte Carlo chain, simulated annealing, genetic algorithms or a method such as PcGets.

Conditional on $M_i(\tau)$, and given the observations $Y_{\tau+1}$, $\mathbf{X}_{\tau,i}$, $\tau = 1, 2, ..., T-1$, the parameters of model M_i can be estimated by least squares. Denoting these estimates by $\widehat{\boldsymbol{\beta}}_{T,i}$, we have

$$\widehat{\boldsymbol{\beta}}_{T,i} = \left(\sum_{\tau=0}^{T-1} \mathbf{X}_{\tau,i} \mathbf{X}'_{\tau,i} \right)^{-1} \sum_{\tau=0}^{T-1} \mathbf{X}_{\tau,i} Y_{\tau+1}. \tag{22}$$

These OLS estimates are fairly simple to compute and $\widehat{\boldsymbol{\beta}}'_{T,i}\mathbf{x}_{T,i}$ could be used as an estimate of $\boldsymbol{\mu}_{iT}$ in our finance example in Section 2. The choice of $\mathbf{X}_{T,i}$ to be used in forecasting of Y_{T+1} can be based on a number of likelihood based model selection criteria suggested in the literature, such as Akaike's Information Criterion (AIC), Schwarz's Bayesian Information Criterion (BIC) or Phillips and Ploberger (1994)'s Posterior Information Criterion (PIC) which applies to both stationary and non-stationary data.

This approach can of course readily be generalized to allow for non-linear effects through indicator type variables as in threshold autoregressive models, Markov switching or more complicated non-linear indicators. As argued by Phillips (1996) it can also be used to choose design parameters such as whether or not to impose a unit root, choice of cointegration rank and deterministic trend degree etc in broad classes of models such as reduced rank regressions, VARs or Bayesian VARs. Whether only linear models or more general models are considered depends on the relative cost of searching across and estimating nonlinear models and the evidence of misspecification among linear models.

4.4 Model Averaging

So far we have focused on approaches that aim to select a single 'best' model. More generally, the model determination process can be considered as selecting a set of weights on each of the models under consideration, $\mathcal{M}_T$. If a single model is used only one of these weights will be nonzero, but it is possible to assign non-zero weights to several models simultaneously, to constrain the weights to add up to one (which makes sense if the individual forecasts are unbiased) and to impose non-negativity of the weights. For example, the forecasting model under "Bayesian model averaging" (BMA) is given by[9]

$$f_b\left(\mathbf{x}_{T+1} \left| \mathcal{F}_T, \mathcal{M}_T \right.\right) = \sum_{M_i \in \mathcal{M}_T} \Pr\left(M_i \left| \mathcal{F}_T \right.\right) f_i\left(\mathbf{x}_{T+1}, \boldsymbol{\theta}_i \left| \mathcal{F}_T \right.\right), \tag{23}$$

where $\Pr\left(M_i \left| \mathcal{F}_T \right.\right)$ is the posterior probability of model M_i which is obtained from the prior distributions, the model priors $\Pr\left(M_i\right)$, and the priors for the unknown parameters, $\Pr\left(\boldsymbol{\theta}_i \left| M_i \right.\right)$, and the likelihood functions of the models under

[9] See, for example, Leamer (1978), Draper (1995) and Hoeting et al. (1999).

consideration. BMA treats the underlying models as random variables. Markov Chain Monte Carlo Model Composition (MC^3) algorithms have been used to compute posterior probabilities for these models. Alternatively, Bayesian model selection methods based on, e.g., the model with the highest posterior probability can be used. For a discussion of related points and references to recent studies, see section 11.3 in Koop (2003).

Note that the BMA density forecast depends on the particular set of models under consideration at time T. Model innovation concerns the evolution of the model space, $\mathcal{M}_t$, over time, t. Conditional on $\mathcal{M}_t$, model specification uncertainty can be taken into account by integrating the predictive density across the full set of models, i.e. by computing a weighted average of the predictive densities of each model. The weights can either reflect the (relative) likelihood values, the posterior probabilities as in (23) or the relative penalized likelihood values using a penalty term such as in the Akaike or Schwarz information criteria. Even with such approaches, it is still an open question which penalty function will yield the best results when judged according to the metric established by the economic loss function, ℓ.

Model averaging is particularly attractive in situations such as that reported by Phillips (1995) where the forecasts from the individual models are found to be very different, and hence subject to a high degree of uncertainty. In the context of the asset allocation problem reviewed in Section 2, averaging could also proceed over probability density forecasts with different specifications of the conditional volatility, Σ_{it}.

A second possibility is to apply (Bayesian) shrinkage methods to the parameters of an individual forecasting model (e.g. a VAR) or to the combination weights, c.f.. Litterman (1986). This is particularly attractive in the case of models with many parameters.

A third possibility is to apply the 'thick modelling' method developed by Granger and Yeon (2004) which takes averages over the full set of models, suitably truncated to avoid using models with measurably poor performance.[10] This is similar to the Occam's window approach of Kass and Raftery (1995) which first disposes of models of high dimensionality and with weak data support as well as models with low posterior probability prior to averaging across the remaining set of models. Numerically efficient Bayesian MCMC methods have also been developed to deal with situations where the number of models to be combined is very large, c.f. Koop (2003).

4.5 Data Window

Model instability is a real possibility due to technological or institutional changes and policy shifts. How much data to use in the modelling stage depends on the nature of possible model instability and the timing of possible breaks. To test for model instability, the full information set, $\mathcal{F}_T$, should always be used to the

[10] Recent applications of model averaging techniques in econometrics include Fernandez, Ley and Steel (2001a,b), Aiolfi and Favero (2002), Garratt, Lee, Pesaran and Shin (2003).

extent possible. Once more is known about the nature of any model instability, an informed decision can be taken on how much data to discard.

Many classes of estimators take the form of weighted least squares. Let $\mathbf{X}_{\tau,i}$ be the τth observation of the ith subset of $\mathbf{X}-$variables. Further, let $\omega_{\tau T} \in [0; 1]$ be the weight on observation τ at time T. Then

$$\widehat{\boldsymbol{\beta}}_{T,i} = \left(\sum_{\tau=0}^{T-1} \omega_{\tau T} \mathbf{X}_{\tau,i} \mathbf{X}_{\tau,i}^{'} \right)^{-1} \sum_{\tau=0}^{T-1} \omega_{\tau T} \mathbf{X}_{\tau,i}^{\prime} y_{\tau+1}. \qquad (24)$$

This class of estimators encompasses as special cases

$$\begin{array}{rcl} \text{expanding window} & : & \omega_{\tau T} = \dfrac{1}{T} \\ \text{rolling window} & : & \omega_{\tau T} = \left\{ \begin{array}{cc} 1/win & \text{if } \tau \geq T - win + 1 \\ 0 & \text{otherwise} \end{array} \right. \\ \text{discounted least squares} & : & \omega_{\tau T} = (1-\lambda)\lambda^{T-1-\tau}/(1-\lambda^{T}) \\ \text{post-break window} & : & \omega_{\tau T} = \left\{ \begin{array}{cc} 1/(T-\hat{T}_b) & \text{if } \tau \geq \hat{T}_b + 1 \\ 0 & \text{otherwise} \end{array} \right. , \end{array}$$

where win is some fixed, predetermined window length and $\hat{T}_b$ is the most recent breakpoint estimate obtained using data up to period T. Pesaran and Timmermann (2002) propose a reversed ordered Cusum (ROC) method that reverses the sequence of the data and tries to identify the most recent break in the data. This method is designed to answer how much historical data to use to estimate the parameters of a forecasting model. For example, when computing the conditional mean of stock returns, $\boldsymbol{\mu}_{it}$, one may choose only to use data after the oil price shocks of 1974 or after the stock market crash of 1987.

The expanding window method is efficient if the DGP is stationary and the model represents the true DGP. Discounted least squares is optimal in some settings with time-varying second moments, while the post-break data window is optimal provided that a sufficiently large, discrete break has occurred. This latter method is related to the approach of Phillips (1996) for discarding data prior to a certain time period based on the relative likelihood ratio of pairs of competing forecasting models.

4.6 Data Revisions and Measurement Errors

Macroeconomic variables in particular are often subject to important measurement errors and data revisions, c.f. Croushore and Stark (2001) and Egginton, Pick and Vahey (2002). Where feasible, any real-time econometric model should make use of such data in all stages so as not to overstate the degree of predictability.

Financial variables such as interest rates and stock prices are typically not similarly affected by data revisions but may be subject to measurement errors - particularly when such data is fed into a computer in real time. To detect outliers that could lead to unreliable results, a filter could be designed that alerts users to aberrations more than a certain distance away from the standard range observed for a particular variable.

Other operational issues are likely to further constrain the econometric model. For a start, the time required to update a model forecast is constrained by real-time considerations. For example, if minute-by-minute trading is desired and the computation of a one-period forecast takes more than one minute, the forecast will be useless by the time it becomes available. The speed of the algorithm could therefore be an important consideration.

4.7 Cross-validation

In view of the very large set of potential regressors in $\mathcal{X}$, spurious relationships could well result from an extensive model specification search and it is important to reduce the effects of data-snooping. One strategy is to apply cross-validation techniques. This can be implemented by splitting the data set into a training sample $[1; R]$ used to select an econometric model $M_i(\boldsymbol{\theta}_{i,R}, \mathcal{F}_R)$ and estimate its parameters to get a predictive density, $\hat{f}_{i_R}$, and an evaluation sample $[R+1; T]$ used to evaluate the model. Since over-parameterized models are likely to perform poorly in the cross-validation sample, this method can be viewed as an alternative way of penalizing for model complexity.

Cross-validation need not be based on the same metric as that used in the loss function. A range of alternative performance indicators - such as the percentage of correctly predicted signs or the Sharpe ratio in case of a trading system - could also be used. This is important since the sampling distribution of the economic loss measure, $\ell(.)$, may be poorly behaved. Suppose for example that the decision maker cares about the wealth generated by an automated trading system. Cumulated wealth, being the product of a sequence of returns over the trading period, is likely to have a sampling distribution with very large standard errors, so even a relatively long cross-validation sample could well lead to inconclusive results.

Cross-validation techniques have other limitations. The modeler is bound to have some knowledge of the basic characteristics of the historical data from $R+1$ to T. This means that the cross-validation sample is not virgin data and could potentially bias the performance. As a case in point, an econometrician may have a data set on stock returns from 1985 to 2000 and use the first part of the data for model construction, while the latter half is used for cross-validation. Knowing that the last half included a sustained bull market could well bias the modeler to include 'momentum' variables that would work very well in the cross-validation period although they would have little if any predictive power in the subsequent sample.

Furthermore, since the rules for the construction of the econometric system has to be 'written in stone' and be in place at the outset of the cross-validation period, inevitably decision rules will be based on old data. This introduces a 'locality problem' since it is likely that the econometric model would have worked better if more recent data could have been used to change the rules of the econometric system.

A final problem associated with cross-validation is that the statistical tests underlying the evaluation period are likely to have weak power and thus may not be very informative. One way to evaluate this loss in power is by quantifying the impact of the specification search leading to the 'best' model using methods such as those adopted by Sullivan, Timmermann and White (2001). Another strategy is to use stress-testing methods that either simulate or bootstrap from a model that captures salient features of the underlying data set and examines the performance of the model under a range of plausible scenarios.

4.8 Applications

Automated model selection is still a relatively recent tool and not that much experience has yet been gained with it. Pesaran and Timmermann (1995, 2000) and Phillips (1995), Schiff and Phillips (2000) report the outcome of applications of automated model selection experiments for stock returns (portfolio allocation) and forecasts of Asia-Pacific (New Zealand) economic activity variables, respectively. Aiolfi and Favero (2002) use thick modeling to obtain recursive forecasts of asset returns. Coe, Pesaran and Vahey (2003) provide an application of recursive modeling to the term structure of U.K. interest rates to see if econometric techniques could have been used to reduce the interest costs of managing public debt. These studies suggest that the automated procedures generally select parsimonious models whose simulated forecasting performance tend to be better than that provided by reasonable benchmarks such as forecasts based on random walk models or official forecasts (Schiff and Phillips (2000)).

These studies also point out that the best approximating model is time-varying. Phillips (1996) introduces the idea of the 'lifetime' of an econometric model. Pesaran and Timmermann (2000) plot out a sequence of indicator variables tracking the inclusion or exclusion of a particular regressor in the forecasting model and find that some variables drop out after an initial inclusion period or get included in blocks in time.

Monte Carlo evidence reported by Hendry and Krolzig (2002) is also suggestive of the ability of automated modeling approaches to select key regressors at least in the context of linear models in a stationary environment.

Phillips (2003) proposes a powerful future application that would make econometric model determination software available to non-expert users via the web. Users choose a set of potentially relevant forecasting variables and may also provide their own data. The software then produces outputs such as graphs of forecasts surrounded by standard error bands. An extension of this idea is to link the forecasts with an objective function representing the user's attitude towards risk and an optimizer and solve for variables such as savings and portfolio choices.

5 Monitoring, Modification and Innovation

5.1 Performance Monitoring

Once a model has been selected and cross-validated using data up to time T, it is reasonable to monitor its real-time performance at regular intervals, e.g. every h periods at time $T+h, T+2h, T+3h, ...T+nh$, where n is a positive integer. The monitoring frequency, h, should be a function of the degree of model instability observed in prior periods as well as the cost of the monitoring.

The initial model is naturally maintained provided that the economic or statistical loss does not exceed some pre-determined value, $\bar{\ell}_{\max}$:

$$(nh)^{-1} \sum_{t=T+1}^{T+nh} \ell(\mathbf{d}_t^*(\hat{f}_i), \mathbf{x}_{t+1}, \boldsymbol{\phi}) \leq \bar{\ell}_{\max}, \tag{25}$$

where we recall that $\hat{f}_i$ is the predictive density conditional on model M_i and

$$\begin{aligned} d_t^*(\hat{f}_i) &= \underset{\mathbf{d}_t \in \mathcal{D}_t}{\arg\min} \left\{ \int_{\mathcal{R}_{t+1}} \ell(\mathbf{d}_t, \mathbf{x}_{t+1}, \boldsymbol{\phi}) \hat{f}_i \left(\mathbf{x}_{t+1} \left| \mathcal{F}_t \right. \right) d\mathbf{x}_{t+1} \right\}, \\ M_i &= \underset{M_j \in \mathcal{M}_T}{\arg\min} \left\{ (nh)^{-1} \sum_{t=T}^{T+nh} \ell(\mathbf{d}_t^*(\hat{f}_i), \mathbf{x}_{t+1}, \boldsymbol{\phi}) \right\}. \end{aligned} \tag{26}$$

If (25) is satisfied, then forecasts and decisions at time $T+nh$ can be conditioned on $M_i(T+nh)$. Alternatively, model rejection could be based on diagnostic tests that reflect the behavior of the forecast errors, e_t, an approach taken by PcGets. For example, one could monitor whether the forecast error exceeds a certain number (κ) of standard deviations of the forecasting model, $|e_t| \geq \kappa\hat{\sigma}(e_t)$, although account also needs to be taken of possible fat tails in the distribution of the forecast errors.

5.2 Model Modification

If the criterion (25) fails to be satisfied, the model space could be expanded from $\mathcal{M}_{T+nh}$ to $\tilde{\mathcal{M}}_{T+nh}$, say, where $\mathcal{M}_{T+nh} \subset \tilde{\mathcal{M}}_{T+nh}$ so the new model and decision rule solve

$$\begin{aligned} M_i &= \underset{\mathcal{M}_j \in \tilde{\mathcal{M}}_{T+nh}}{\arg\min} \; (nh)^{-1} \sum_{t=T}^{T+nh} \ell(\mathbf{d}_t^*(\hat{f}_i), \mathbf{x}_{t+1}, \boldsymbol{\phi}) \\ d_t^*(\hat{f}_i) &= \underset{\mathbf{d}_t \in \mathcal{D}_t}{\arg\min} \left\{ \int_{\mathcal{R}_{t+1}} \ell(\mathbf{d}_t, \mathbf{x}_{t+1}, \boldsymbol{\phi}) \hat{f}_i \left(\mathbf{x}_{t+1} \left| \mathcal{F}_t \right. \right) d\mathbf{x}_{t+1} \right\}. \end{aligned} \tag{27}$$

Model modification may lead to the inclusion of an omitted factor deemed previously to be unimportant. One example of a dormant factor that came to life is oil prices during the seventies. Since oil prices varied far less prior to the 1970s, few econometric models included them, but this clearly changed subsequently.

Such modification could also be triggered automatically by some real-time monitoring system if there is evidence of a model breakdown. It is closely related to the framework for model determination proposed by Phillips (1996) which includes the possibility that the true probability measure underlying the DGP itself evolves over time due to technological or institutional shifts.

Quite independently of whether or not model performance is being considered, the parameter estimates, $\hat{\boldsymbol{\theta}}_{t,i}$, can be updated either every period ($h = 1$) or at regular intervals, again depending on the relative cost-benefit of such updates.

This process is very similar to the model evaluation stage used to cross-validate the econometric model over the in-sample period. It involves specifying a range of statistical performance measures that can be the basis for decisions on whether and how to modify the econometric model. As part of the monitoring, tests for structural breaks could also be undertaken. Unfortunately, existing breakpoint tests lack power against plausible (local) alternatives and in many cases do not provide precise information about the time of the break. Some breaks may, however, be large enough that they can be detected. Even when such tests indicate some form of instability they typically do not reveal the form of the instability, i.e. a break-down in model parameters or the inclusion of new predictor variables.

5.3 Model Innovation

To avoid some of the same pitfalls involved in the initial model construction phase, the process of monitoring and modifying the econometric model in real time should be set out in advance prior to the beginning of the experiment. This amounts to specifying in advance how the model space is extended, i.e. setting out rules for constructing the sequence $\{\mathcal{M}_{T+nh}\}_{n=1}^{\infty}$. For example, one could specify rules for considering new families of models for the conditional mean, $g^*(\mathbf{X}_{\tau,i};\boldsymbol{\beta}_i)$, new predictive density models $f^*(\mathbf{X}_{\tau,i};\boldsymbol{\theta}_i)$ or new classes of estimators, $\hat{\boldsymbol{\theta}}_i$. Thus, if the purpose is to study the efficient market hypothesis, one could introduce new forecasting models (e.g., cointegrating regressions, neural networks or wavelets) only at the point when they were introduced. Finding that a particular econometric forecasting method could have been used to generate abnormal profits 10 years prior to its invention would not be of much interest. If there are changes to the set of feasible decision rules, one also needs to set out the sequence $\{\mathcal{D}_{T+nh}\}_{n=1}^{\infty}$.

Inevitably innovative modelling relies on the modelers judgement and so it is typically more difficult to automate this stage than e.g. the selection of regressors from a pre-specified universe. The role of automation in this stage is therefore still an open question.

6 Future Research

Much work remains to be done before the scope for success as well as the limitations of automated econometric models can be assessed. Little is known about the performance of different econometric approaches when models are viewed as local approximations to an evolving data generating process as proposed by Phillips (1996). In general, however, parsimonious models tend to produce better out-of-sample forecasting performance than models with a large number of variables and parameters. In economics, non-stationarities take many different forms: at times the underlying DGP may be subject to sudden discrete shocks, at other times changes may evolve more gradually. The challenge is to design robust systems that can handle both types of change. One idea is to attempt to design a real-time break monitoring procedure that detects the speed of change and then conditionally chooses either fast adapting or slowly adapting forecasting models.

Another possibility is to group forecasting models by types (e.g., error correction models, stationary VARs, Bayesian VARs, time-varying parameter models), select the best model within each class and then average across different classes of models using Bayesian type model averaging techniques. A conjecture is that averaging across very different types of models (as opposed to models that only differ by, say, their lag order) may provide robustness against larger types of change. At the same time, averaging across too many types of models may slow down the adaptation of the combined forecast following a large shift in the DGP. Much research is still needed to answer which of these effects dominates in practice.

Automated systems reduce, but do not eliminate the need for discretion in real time decision making. There are many ways that automated systems can be designed and implemented. The space of models over which to search is huge and is likely to expand over time. Different approximation techniques such as genetic algorithms, simulated annealing and MCMC algorithms can be used. There are also many theoretically valid model selection or model averaging procedures. The challenge facing real time econometrics is to provide insight into many of these choices that researchers face in the development of automated systems.

References

[1] Aiolfi, M. and C. Favero, 2002, Model Uncertainty, Thick Modelling and the Predictability of Stock Returns. IGIER Working paper n.221, (July 2002).

[2] Brown, R.L., J. Durbin and J.M. Evans, 1975, Techniques for Testing the Constancy of Regression Relationships over Time. Journal of the Royal Statistical Society, Series B 37, 149-192.

[3] Coe P.J., M.H. Pesaran and S.P. Vahey, 2003, Scope for Cost Minimization in Public Debt Management: The Case of the UK. Cambridge University, www.econ.cam.ac.uk/faculty/pesaran/cpvaug03.pdf.

[4] Croushore, D. and T. Stark, 2001, A Real-Time Data Set for Macroeconomists. Journal of Econometrics 105, 111-130.

[5] Dawid, A.P., 1984, Present Position and Potential Developments: Some Personal Views - Statistical Theory, the Prequential Approach. Journal of the Royal Statistical Society Series A, 147, 278-292.

[6] Draper, D., 1995, Assessment and Propagation of Model Uncertainty. Journal of Royal Statistical Society Series B, 58, 45-97.

[7] Egginton, D.M., A. Pick, and S.P. Vahey, 2002, Keep it Real!: A Real Time UK Macro Data Set. Economic Letters, 77, 15-20.

[8] Fernandez, C., E. Ley, and M.F.J. Steel, 2001a, Benchmark Priors for Bayesian Model Averaging. Journal of Econometrics, 100, 381-427.

[9] Fernandez, C., E. Ley and M.F.J. Steel, 2001b, Model Uncertainty in Cross-Country Growth Regressions. Journal of Applied Econometrics, 16, 563-576.

[10] Garratt, A, K. Lee, M. H. Pesaran and Y. Shin, 2003, Forecast uncertainties in macroeconometric Modelling: An Application to the UK Economy. Journal of the American Statistical Association, 98, 829-838.

[11] Granger, C.W.J. and D. F. Hendry, 2004, A Dialog Concerning a New Instrument for Econometric Modeling. Econometric Theory (this issue).

[12] Granger, C.W.J. and Y. Jeon, 2004, Thick Modeling. Economic Modeling 21, 323-343.

[13] Granger, C. W. J. and M. H. Pesaran, 2000a, A Decision Theoretic Approach to Forecast Evaluation. In W.S. Chan, W.K. Li and Howell Tong (eds), Statistics and Finance: An Interface, London, Imperial College Press, Ch.15, 261-278.

[14] Granger, C. W. J. and M. H. Pesaran, 2000b, Economic and Statistical Measures of Forecast Accuracy. Journal of Forecasting, 19, 537-560.

[15] Hendry, D.F. and H.M. Krolzig, 2002, New Developments in Automatic General-to-specific Modeling. In (B.P. Stigum ed.) Econometrics and the Philosophy of Economics. MIT Press. forthcoming.

[16] Hoeting, J. A., D. Madigan, A.E. Raftery and C.T. Volinsky, 1999, Bayesian Model Averaging: A Tutorial. Statistical Science, 14, 382-417.

[17] Kass, R. and A. E. Raftery, 1995. Bayes Factors. Journal of American Statistical Association 90, 773-795.

[18] Koop, G., 2003, Bayesian Econometrics. Wiley and Sons.

[19] Leamer, E.E., 1978, Specification Searches: Ad Hoc Inference with Nonexperimental Data. Wiley, New York.

[20] Litterman, R.B., 1986, Forecasting with Bayesian Vector Autoregressions: Five Years of Experience. Journal of Business and Economic Statistics 4, 25-38.

[21] Pesaran, M.H. and S. Skouras, 2002, Decision-based Methods for Forecast Evaluation, in M.P. Clements and D.F. Hendry (Eds.). A Companion to Economic Forecasting. Oxford: Basil Blackwell, Chapter 11, pp.241-267.

[22] Pesaran, M.H. and A. Timmermann, 1995, The Robustness and Economic Significance of Predictability of Stock Returns. Journal of Finance, 50, 1201-1228.

[23] Pesaran, M.H. and A. Timmermann, 2000, A Recursive Modelling Approach to Predicting U.K. Stock Returns. The Economic Journal, 110, 159-191.

[24] Pesaran, M.H. and A. Timmermann, 2002, Market Timing and Return Prediction under Model Instability. Journal of Empirical Finance, 9, 495-510.

[25] Phillips, P.C.B., 1995, Automated Forecasts of Asia-Pacific Economic Activity. Asia-Pacific Economic Review 1 (1), 92-102.

[26] Phillips, P.C.B., 1996, Econometric Model Determination, Econometrica, 64, 763-812.

[27] Phillips, P.C.B., 2003, Laws and Limits of Econometrics. The Economic Journal, 113, C26-C52.

[28] Phillips, P.C.B. and W. Ploberger, 1994, Posterior Odds Testing for a Unit Root with Data-based Model Selection. Econometric Theory, 10, 774-808.

[29] Schiff, A.F. and Phillips, P.C.B., 2000, Forecasting New Zealand's Real GDP. New Zealand Economic Papers, 34 (2), 159-182.

[30] Stock, J.H. and M. Watson, 2002, Macroeconomic Forecasting Using Diffusion Indexes. Journal of Business and Economic Statistics, 20, 147-162.

[31] Sullivan, R., A. Timmermann and H. White, 2001, Dangers of Data-driven Inference: The case of calendar effects in stock returns. Journal of Econometrics, 105, 249-286.

[32] Wald, A., 1947. Sequential Analysis, John Wiley, New York.