Can Econometrics Improve Economic Forecasting?

DAVID F. HENDRY and MICHAEL P. CLEMENTS*

1. INTRODUCTION

Economic forecasting has rarely had a good press, but currently, in the UK at least, has a singularly poor record. Here, we consider a number of potential avenues by which econometrics may be able to help forecasting. We analyze some positive suggestions and present some critical appraisals, although most of our conclusions await empirical evaluation prior to their operational use. We use the word prediction to denote any statement about an unobserved outcome; a forecast is the special case of a prediction about a future event, where the future may be relative to the starting point rather than an absolute time still to come.

Econometric theory comprises a body of tools and techniques for analyzing the properties of prospective methods under hypothetical states of nature. It can deliver relevant conclusions about forecasting only when those states adequately capture the appropriate aspects of the real world to be forecast. Given the difficulties of econometric modelling, our analysis allows the forecasting model to differ from the data generation process (DGP); the DGP to be non-stationary and susceptible to structural breaks; and the specification of the 'best' model to be unknown *a priori* and hence selected using data evidence and theoretical considerations. All of these aspects matter, and so call in to question the relevance of 'optimal' forecasting devices based on assuming constant, stationary DGPs which coincide with a unique model under analysis. This work builds on CLEMENTS and HENDRY (1994a), who offer a taxonomy of sources of forecast errors.

On the modelling side, we argue for congruent, encompassing, and invariant models. These can be developed in-sample, but the last cannot be achieved for all equations in a system which is subject to regime changes. We note the possibility of using intercept corrections to robustify forecasts against some types of shocks, based on HENDRY and CLEMENTS (1994), and contrast that proposal with differencing to convert step changes to 'blips'. An artificial data example illustrates the latter.

On the critical side, we discuss the role of forecast evaluation criteria and their invariance to admissible transforms. The limitations of mean square forecast error (MSFE) criteria discussed in CLEMENTS and HENDRY (1993a) add to the complications of evaluation, but the illustration highlights the dangers of only taking a few scalar

* Institute of Economics and Statistics, and Nuffield College, Oxford. Financial support from the U.K. Economic and Social Research Council under grant R000233447 is gratefully acknowledged by both authors. We are indebted to JURGEN DOORNIK, NEIL ERICSSON, AGUSTIN MARAVALL and GRAYHAM MIZON for helpful comments.

measures. Forecast standard errors can reveal some of the uncertainties inherent in forecasts and so should be reported, but unfortunately need to be computed for all relevant transformations. Comparisons of conditional with unconditional predictors can be informative and the complexities of the arguments for and against pooling in worlds with regime shifts are described in the light of the encompassing critique.

Thus, we see several ways in which econometrics may be able to help in forecasting. Economic forecasting itself has had a checkered history, ranging from a denial of the possibility of economic forecasting in principle (MORGENSTERN, 1928), through a period of increasing optimism in the value of model-based forecasting in the late 1960s and early 1970s (see, for example, BURNS, 1986), to the current, rather downbeat, assessment of the value of such models. We begin in section 2 with a summary of some of the salient episodes in the history of economic forecasting, and of the role played by models in particular, to set the scene. We end that section by noting some of the recent contributions which are relatively critical of the standard approach to producing and evaluating forecasts. We reference our own work in section 2 to inform the reader of where we have attempted to deal with the issues that arise. Section 3 begins by noting a number of methods of forecasting, and argues that a major benefit of forecasts based on time-series and econometric models is that they allow the derivation of measures of forecast uncertainty. Then, we discuss the issue of when forecasts are informative, before closing that section by setting up the DGP as a cointegrated system of I(1) variables. In section 4 we establish a framework that will enable us to analyze the properties of forecasts when the DGP is susceptible to structural breaks, incorrect model specifications are used, and parameters are estimated from small data samples where recent observations may be badly measured. In section 5, we examine each of the resulting sources of forecast error and how econometrics can help. Section 6 discusses forecast evaluation criteria, and the case for pooling forecasts. Structural breaks induce model mis-specification: to the extent that models are differentially susceptible to such breaks, there may be scope for pooling (although pooling runs counter to the encompassing principle and has little to commend it in an unchanging world, see CLEMENTS and HENDRY, 1993a). Our preferred course of action involves a form of intercept correction. We illustrate the analysis with an example in section 7, and section 8 concludes.

This paper mainly stresses the positive aspects of econometrics in economic forecasting. However, section 6 criticises the use of MSFEs in evaluating forecasts, and section 2 notes some criticisms of the way in which econometrics has been applied to forecasting.

2. A HISTORY OF THE THEORY OF ECONOMIC FORECASTING

There are comments about forecasting and forecasting methods throughout the statistics and economics literature in the first quarter of the 20th century, especially related to 'business barometers' (see MORGAN, 1990). However, the first comprehensive treatise seems to be MORGENSTERN (1928), who analyzed the methodology of economic forecasting. We know of no translation into English, but the book was extensively reviewed by MARGET (1929). The following is based on, and extends, HENDRY and MORGAN (1994).

MORGENSTERN argued against the possibility of economic and business forecasting in principle. His first argument adopted the (then widely-held) belief that economic forecasting could not rely on probability reasoning because economic data are neither homogeneous nor independently distributed 'as required by statistical theory'. Samples of the size which satisfied the assumptions would be too small to be usable (see KRÜGER, GIGERENZER and MORGAN (1987) on the probability revolution in the sciences). PERSONS (1925) held a similar view. Despite the almost universal distrust of the probability approach prior to HAAVELMO (1944) (as documented, for example, by MORGAN, 1990), economists at the time had been making increasing use of statistical methods and statistical inference to 'test' economic theories against the data. Standard errors and multiple correlation coefficients were sometimes given as a gauge of the validity of the relationship. HAAVELMO argued forcibly that the use of such tools along side the eschewal of formal probability models was insupportable: 'For *no tool developed in the theory of statistics has any meaning* – except, perhaps, for descriptive purposes – *without being referred to some stochastic scheme*' (HAAVELMO, 1944, from the Preface, original italics).

Secondly, MORGENSTERN argued that forecasts would be invalidated by agents' reactions to them. This is an early variant of the now named 'Lucas critique' (LUCAS, 1976), and was an issue which concerned many business-cycle analysts of the period. Because of the inherent impossibility of economic forecasting and the adverse impact it would have on agents' decisions, MORGENSTERN foresaw dangers in the use of forecasting for purposes of stabilization and social control. Thus, he made an all-out assault on economic forecasting and its applications.

MARGET agreed with MORGENSTERN's first point, but not the conclusion since most economic forecasting, while based on statistical data, depended on extrapolating previous patterns rather than on probability-based forecasting techniques. However, to the extent that statistical tools played a part in identifying the 'patterns' to be extrapolated, MARGET's defence of forecasting is suspect. Further, MARGET argued that if a causal explanation was possible, even if not exact or entirely regular, forecasting should be possible. This presupposes that causation implies predictability, which is not obviously the case: a potential counter-example is the theory of evolution. Section 3.2 discusses when forecasts are informative.

On a more mundane level, MARGET's counter to MORGENSTERN requires regularity between causes and their effects. As noted below in discussing HAAVELMO's views on prediction, this was explicitly recognised as a necessary ingredient for success. Nihilistic critiques, such as that of ROBBINS (1932), denied there were good reasons to believe that this would be the case: on the claim that average elasticities calculated for certain epochs could be described as 'laws' he wrote: 'there is no reason to suppose that there having been so in the past is the result of the operation of homogeneous causes, nor that their changes in the future will be due to the causes which have operated in the past.' (ROBBINS, 1932, p.101).

MORGENSTERN was one of the first to explicitly raise the problem of 'bandwagon feedback effects', but MARGET denied that the problem was either specific to economics or that it precluded accurate forecasting, and constructed a counter-example based on continuous responses. Finally, MARGET argued for using forecasting to test economic theories. This seems to have been a new proposal, and by the middle 1930s, economists had begun to use economic forecasting for testing their models: TINBERGEN (1939) was one of those responsible for developing forecasting tests of econometric models, despite the criticisms of KEYNES (1939) and FRIEDMAN (1940). At the conclusion of the debate between MARGET and MORGENSTERN, some of the limitations to, and possibilities for, economic forecasting had been clarified, but no formal theory was established.

HAAVELMO (1944) treated forecasting as a straightforward application of his general methodology, namely a probability statement about the location of a sample point not yet observed. Once a structural representation of the joint data density function was available, there were no additional problems specifically associated with prediction. Since HAAVELMO's treatment of forecasting (or rather prediction, in his terminology) in a probabilistic framework was in many ways the progenitor of the classical, textbook approach to forecasting, we now sketch his analysis.

Suppose there are N observable values $(x_1,..., x_N)$ on the random variable X, from which to predict the H future values $(x_{N+1},..., x_{N+H})$. Let the joint probability of the observed and future x's be $\mathrm{D}(\cdot)$, which is assumed known. Denote the probability distribution of the future x's conditional on past x's by $\mathrm{D}_2\,(x_{N+1},..., x_{N+H} \mid x_1,..., x_N)$, and the probability distribution of the observed x's as $\mathrm{D}_1(\cdot)$, then, factorizing into conditional and marginal probabilities:

$$\mathrm{D}\,(x_1,..., x_{N+H}) = \mathrm{D}_2\,(x_{N+1},..., x_{N+H} \mid x_1,..., x_N) \times \mathrm{D}_1\,(x_1,..., x_N)$$

Thus, for any realisation of the observed x's, denoted by the set $E_1 \subset R^N$, we can calculate from $\mathrm{D}_2(\cdot)$ the probability that the future x's will lie in any E_2, where $E_2 \subset R^H$. Conversely, for a given probability we can calculate 'regions of prediction'. Thus, 'the problem of prediction ... (is) ... merely ... a problem of probability calculus' (p.107).

In practice $\mathrm{D}_2\,(\cdot)$ is not known and must be estimated on the basis of the realisation E_1, which requires the 'basic assumption' that: 'The probability law $\mathrm{D}\,(\cdot)$ of the $N + H$ variables $(x_1,..., x_{N+H})$ is of such a type that the specification of $\mathrm{D}_1(\cdot)$ implies the complete specification of $\mathrm{D}(\cdot)$ and, therefore, of $\mathrm{D}_2(\cdot)$.' (p.107: our notation). Loosely, this requires the sort of regularity denied by ROBBINS, or again quoting from HAAVELMO, 'a certain persistence in the type of mechanism that produces the series to be predicted.' (p.107). We will consider the extent to which the failure of this type of persistence (e.g. structural breaks) can be countered by appropriate forecasting technique (see section 5).

Prior to HAAVELMO's 1944 paper, the first macro-economic model had been built by TINBERGEN in the Netherlands during the 1930s. The pioneering work of TINBERGEN on model construction and policy analysis received prominence in two reports commissioned by the League of Nations. MORGAN (1990, chapter 4) details these developments and the critical reaction from KEYNES (1939) and others. Crudely, KEYNES' position was that if the data analysis failed to confirm the theory, then blame the data and statistical methods employed. On the contrary, 'For TINBERGEN, econometrics was concerned both with discovery and criticism' (MORGAN, 1990, p.124).[1]

The influence of TINBERGEN ensured that official Dutch economic forecasting was more model-orientated than anywhere else in the world, and the role of econometric modelling and forecasting was further strengthened by THEIL (1958, 1966). For example, MARRIS (1954) compares the Dutch and British approaches to forecasting, and concludes that the British makes use of 'informal models' (cf. what THEIL (1961, p.51) describes as a 'trial and error' method) where, as already mentioned, the Dutch approach is based on a formal model. Although MARRIS sides with the Dutch approach, he has some reservations which include 'the danger of its operators becoming wedded to belief in the stability of relationships (parameters) which are not in fact stable.' (MARRIS, 1954, p.772). In other words, MARRIS fears the 'mechanistic' adherence to models in the generation of forecasts when the economic system changes. More generally, a recognition of the scope for, and importance of, adjusting purely model-based forecasts has a long lineage (see, inter alia, THEIL (1961, p.57), KLEIN (1971), KLEIN, HOWREY and MACCARTHY (1974), the sequence of reviews by the UK ESRC Macroeconomic Modelling Bureau, WALLIS et al., 1984-7, TURNER (1990), WALLIS and WHITLEY (1991), HENDRY and CLEMENTS (1994) and CLEMENTS, 1994). HENDRY and CLEMENTS (1994) attempt to establish a general framework for the analysis of adjustments to model-based forecasts. It is based on the relationships between the data generating process (DGP), the estimated econometric model, the mechanics of the forecasting technique, data accuracy, and any information about future events held at the beginning of the forecast period. The key motivating factors are the recognition that typically the DGP will not be constant over time and that in practice, the econometric model and the DGP will not coincide (see CLEMENTS and HENDRY, 1994a,b).

HAAVELMO's probability approach appears to have been generally accepted by the end of the 1940s, especially in the US, where it underlay the macroeconometric model built for the Cowles Commission by KLEIN, who was one of the prime movers. KLEIN (1950) explicitly recognises that, when the goal is forecasting, econometric modelling practice may differ from when explanation or description are the aim. KLEIN describes a mathematical model consisting of the structural relationships in the economy (production functions, demand equations, etc.) and argues that for forecasting it may not be necessary to uncover the structural parameters but that attention can usefully focus on

1. The elements of this debate survive to the present day: see HENDRY (1993).

the reduced form. However, KLEIN believes that this will result in worse forecasts unless the system is exactly identified, since otherwise there will be a loss of information. Some of our recent results suggest such fears may be exaggerated. CLEMENTS and HENDRY (1993b) show that it may be possible to improve forecast accuracy by imposing restrictions even when the restrictions are false, once proper account is taken of the forecast error variance due to parameter estimation uncertainty.

Apart from KLEIN (1950), BROWN (1954) (who derived formulae for standard errors of forecasts from systems), and THEIL (1958, 1966), time-series analysts played the dominant role in developing theoretical methods for forecasting in the postwar period (see WIENER (1949), KALMAN (1960), BOX and JENKINS (1970) and HARVEY (1981); WHITTLE (1963) provides an elegant summary of least-squares methods). More recently, numerous texts and several specialist journals have come into existence; empirical macro-econometric forecasting models abound and are often subject to forecast evaluation tests (for a partial survey, see FAIR, 1986). The sequence of appraisals of U.K. macro-models by WALLIS et al. (1984-7) provides detailed analyses of forecasting performance and the sources of forecast errors, and KLEIN and BURMEISTER (1976) and KLEIN (1991) provide similar analyses of the US industry of macro-forecasting. WALLIS (1989) references some of the important contributions of the 1970s and 1980s.

Nevertheless, there remain difficulties with existing approaches. GRANGER and NEWBOLD (1973) survey a number of ad hoc forecast evaluation criteria which they find to be either inadequate or misleading. CHONG and HENDRY (1986) argue that three commonly used procedures for assessing overall model adequacy are problematical: namely, dyamic simulation performance, the historical record of genuine forecast outcomes, and the economic plausibility of the estimated system. They propose a test of forecast encompassing to evaluate the forecasts from large-scale macroeconomic models when data limitations render standard tests infeasible (see also HENDRY, 1986). As an application of the encompassing principle, forecast encompassing is more soundly based than some of the criteria criticised by GRANGER and NEWBOLD (1973). ERICSSON (1992) and CLEMENTS and HENDRY (1993a) consider the traditional use of MSFEs in forecast evaluation (see section 6), and CLEMENTS and HENDRY (1994a,b) focus on the wider implications of forecasting in a changing world where the forecast model and the mechanism generating the data do not coincide.

3. A FRAMEWORK FOR ECONOMIC FORECASTING

3.1. *Forecasting methods*

There are undoubtedly dozens of ways of making economic forecasts but any model based (or systematic and repeatable method) appears to require the following three ingredients:

(*i*) that there are regularities on which to base models;

(*ii*) that such regularities are informative about the future; and:
(*iii*) that they are encapsulated in the selected forecasting model.

The first two items relate to properties of the economy under analysis, and so are common to all forecasting models for that economy, while the third relates to the model selected. OECD economies exhibit some constant features but are also subject to regime shifts (the formation of the European Community, 'oil shocks', moving from fixed to floating exchange rates etc.). When the latter are recurrent, it may be possible to model them *ex post*, even if their next occurrence remains uncertain, thereby increasing the constancy of the system.

Based on CLEMENTS and HENDRY (1994a), we enumerate seven distinct forecasting methods in common use, namely: guessing (which 'just' relies on luck); extrapolation (which relies on tendencies persisting); leading indicators (which rely on continuing to lead in a systematic manner); surveys (which rely on the implementation accuracy of plans, expectations and anticipations); analysis 'in the context of an implicit, perhaps informal model, not necessarily written down' (WALLIS, 1989, p.31, and THEIL, 1961, p.51); time-series models such as the ARIMA class (see BOX and JENKINS, 1970) and VARs (see DOAN et al., 1984), (which rely on 'continuity' of the time series representation); and econometric systems (which rely on capturing the structural aspects of agents' decision taking). In practice, all seven methods may contribute: for example, macro-econometric forecasts use intercept corrections which may invoke luck, extrapolation, etc.

Formal econometric systems of national economies fulfill many useful roles other than just being devices for generating forecasts; for example, such models consolidate existing empirical and theoretical knowledge of how economies function, provide a framework for a progressive research strategy, and help explain their own failures. They are open to adversarial scrutiny, are replicable, and hence offer a scientific basis for research: compare, in particular, guessing and the use of informal models.

Perhaps at least as importantly, time-series and econometric methods are based on statistical models, and therefore allow derivations of measures of forecast uncertainty, and associated tests of forecast adequacy.[2]

It is not clear how to interpret point forecasts in the absence of any guidance as to their accuracy, and we contend that this factor alone constitutes a major way in which econometrics can improve the *value* of economic forecasts. In his review article, CHATFIELD (1993) argues strongly in favour of interval forecasts and assesses the ways

2. HARVEY (1989), for example, terms forecasting methods not based on statistical models *ad hoc*. This description would appear to apply to many popular techniques, such as exponentially weighted moving averages (EWMA) and the schemes of HOLT (1957) and WINTERS (1960), which effectively fit linear functions of time to the series but place greater weight on the more recent observations. The intuitive appeal of such schemes stems from the belief that the future is more likely to resemble the recent rather than the remote past. HARVEY (1989) shows that such procedures can be given a sound statistical basis since they are derivable from the class of 'structural' time series models.

in which these can be calculated. A prediction interval is a range within which predictions will fall with a certain probability (commonly taken at 95%). The calculation of prediction intervals is relatively straightforward for scalar or vector processes, particularly in the absence of parameter uncertainty, but otherwise requires recourse to the formulae in SCHMIDT (1974,1977) and CALZOLARI (1981). However, they are more difficult to calculate for large, non-linear macroeconometric models, so analogous expressions for prediction intervals are not available, although a range of simulation methods can be employed. Non-modelled variables can in principle be handled, but their treatment in practice may have a significant influence on the results: treating variables which are hard to model as 'exogenous', then supposing that such variables are known with certainty, may seriously understate the uncertainty of the forecasts. Furthermore, it is not clear what the impact of forecasters' judgements or intercept corrections will be on prediction intervals.

When a macro-econometric model is claimed to capture the constant features of the economic system in an undominated way, it must satisfy the following criteria (see HENDRY and RICHARD, 1982):

(a) be congruent, so that it fully embodies the available data information;
(b) encompass or account for the results obtained by rival explanations;
(c) be invariant to structural change;
(d) accurately and precisely estimate all unknown coefficients.

We discuss below the importance, or otherwise, of these requirements for forecasting.

3.2. *When are forecasts informative?*

The informativeness of a forecast is a relative concept, dependent on the pre-existing state of information. If one knew nothing about a dice-rolling experiment, then it is informative to be told that the only possible outcomes are the integers 1 through 6. Once the equal probability of such outcomes is known, then the forecast that the next outcome has a uniform distribution over [1,6] is uninformative. When there exist transforms of data such that the unconditional distribution is well defined, a forecast could be defined as informative when the forecast error has a more concentrated probability distribution than the unconditional distribution of the variable being forecast. We illustrate these ideas with the weakly stationary first-order autoregressive process (so $|\psi| < 1$):

$$w_t = \psi\, w_{t-1} + v_t \text{ where } w_0 \sim \mathrm{N}\left(0, \frac{\sigma_v^2}{1-\psi^2}\right) \tag{1}$$

and $\{v_t\}$ is an independent, identically distributed normal random variable with zero mean and constant variance σ_v^2, denoted by $v_t \sim \mathrm{IN}(0, \sigma_v^2)$, then w_t is strictly stationary. Thus, w_t meets the requirement that the unconditional distribution is well defined:

$$w_t \sim \mathrm{N}\left(0, \frac{\sigma_v^2}{1-\psi^2}\right) \text{ for all } t.$$

Consider forecasting the value of the process in period $T+1$ conditional on information in period T, which we denote by $w_{T+1 \mid T}$ when ψ is known. The 1-step ahead forecast error is $e_{T+1 \mid T} = w_{T+1} - w_{T+1 \mid T} = v_{T+1}$, so that $\mathrm{V}[e_{T+1 \mid T}] = \sigma_v^2 < \sigma_w^2 \equiv \mathrm{V}[w_t]$, where $\mathrm{V}[\cdot]$ denotes a variance. Thus, 1-step forecasts are informative. CLEMENTS and HENDRY (1994a) show that as h increases, $\mathrm{V}[e_{T+h \mid T}] \to \mathrm{V}[w_t]$ (monotonically from below for known parameters) so that informativeness disappears.

From the forecast error variance, we can define a $100(1-\alpha)\%$ prediction interval as a range that should include that % of future realisations of forecasts, by:

$$w_{T+1 \mid T} \pm z_{\alpha/2} \sqrt{\mathrm{V}\left[e_{T+1 \mid T}\right]}$$

where $z_{\alpha/2}$ denotes the percentage point of the standard normal distribution. The generalisation to h-steps ahead follows immediately.

Consider applying this notion of informativeness to predicting the change in the log of real equity prices (w_t in (1)), which is believed to be nearly white noise (so that ψ is close to zero):

a) the log-difference transform assumes a more standardized behaviour of percentage changes over absolute changes;

b) the direction of change from day to day is difficult to predict so the conventional comment that 'real equity price changes are unpredictable' probably refers to this aspect (when $\psi \cong 0$, then $\sigma_v^2 \cong \sigma_w^2$ and forecasts are close to being uninfomative);

c) however, to a novice, many helpful features may need highlighting, such as the asymmetry of changes, the high probability of 'outliers' (thick tails), the non-normality etc. of $\mathrm{D}(\cdot)$;

d) even so, a stockbroker may be able to make money from a prediction where the forecast error is only a little more concentrated than the outcome; and:

e) a trader in derivatives, such as options, may also make money from predicting the conditional variance more accurately than σ_v^2.

If the prediction of the change in the log of real equity prices is uninformative, is that of the level? Again the issue turns on the state of existing information. Non-stationary, evolutionary processes (such as those with unit roots) do not have well-behaved unconditional distributions, so the conditional prediction of the level next day is more concentrated than the unconditional. Nevertheless, when the optimal prediction of the change in the log of real equity prices is zero, we doubt that investors would attach any value to the prediction that next day's level will be the same as today's on average.

Consider predicting income, consumption and saving when $I \equiv C + S$. Then the linear combination $I - C - S \equiv 0$ is an uninformative prediction as it is always true, yet the components should be more or less accurately predicted.[3] Thus, informativeness is not an invariant construct, nor is uninformativeness, by the obverse argument.

3.3. *The Data Generation Process*

The analysis assumes a linear, closed system for tractability, where all non-deterministic variables are forecast within the system. The vector of all n variables is denoted by w_t and the system is represented by a first-order VAR:

$$\mathrm{w}_t = \tau + \Upsilon \mathrm{w}_{t-1} + \nu_t \tag{2}$$

where $\nu_t \sim \mathrm{IN}(0, \Omega)$. The parameter estimates are $\left(\hat{\tau} : \hat{\Upsilon} : \hat{\Omega}\right)$ where:

$$\operatorname*{plim}_{T\to\infty} \hat{\tau} = \tau_p \text{ and } \operatorname*{plim}_{T\to\infty} \hat{\Upsilon} = \Upsilon_p ,$$

and the potential inconsistencies are due to model mis-specification. Although we allow $\tau_p \neq \tau$ and $\Upsilon_p \neq \Upsilon$, when the $\{\nu_t\}$ process is independently sampled, or is even just a martingale difference sequence relative to the history W_{t-1} of $\{\mathrm{w}_t\}$, then from (2):

$$\mathrm{E}\left[\mathrm{w}_t \mid \mathrm{W}_{t-1}\right] = \tau + \Upsilon \mathrm{w}_{t-1} \tag{3}$$

so no inconsistency arises even though ν_t need not be an innovation against a broader information set (i.e. there are possibly important omitted variables). This notion of a congruent empirical model is discussed in section 5 and more extensively in HENDRY (1994).

The system is assumed to be integrated and to satisfy $r < n$ cointegration relations such that:

3. THEIL (1961, pp.14,15) discusses whether forecasts should obey 'internal consistency', in that they obey the same identities as the outcomes, and suggests that other considerations may contradict this requirement.

$$\Upsilon = I + \alpha\beta'$$

where α and β are $n \times r$ matrices of rank r. Then (2) can be reparameterized as a vector error correction (VECM: see JOHANSEN, 1988):

$$\Delta w_t = \tau + \alpha\beta' w_{t-1} + v_t \tag{4}$$

Under stationarity of both Δw_t and $\beta' w_t$, when the system grows at the (vector) rate $E[\Delta w_t] = \gamma$, then the long-run (or unconditional) solution is:

$$\alpha E\left[\beta' w_t\right] = \gamma - \tau \tag{5}$$

and since $\Delta(\beta' w_t)$ is the difference of a stationary variable, $\beta'\gamma = 0$. When $\beta'\alpha$ is non-singular:

$$E\left[\beta' w_t\right] = -(\beta'\alpha)^{-1}\beta'\tau \tag{6}$$

Consequently, when $\tau = \alpha\mu$ (say), $E[\beta' w_t + \mu] = 0$, so from (5), $\gamma = 0$, and the system has no growth. This matches the fact that (4) can then be written as:

$$\Delta w_t = \alpha(\beta' w_{t-1} + \mu) + v_t \tag{7}$$

More generally, when $\tau = \alpha\mu + \gamma$ we have:

$$\Delta w_t = \gamma + \alpha(\beta' w_{t-1} + \mu) + v_t \tag{8}$$

The intercept and its treatment will prove to be important below. Throughout, we treat β as if it were known, since sampling errors in it are an order of magnitude (in the sample size) less important than in the coefficients of stationary variables.

Finally, a VAR in differences (DVAR) may be used, which within sample is mis-specified relative to the VECM unless $r = 0$:

$$\Delta w_t = \gamma + e_t \tag{9}$$

From (5) and (6), $\gamma = (I - \alpha(\beta'\alpha)^{-1}\beta')\tau$ (cf. CLEMENTS and HENDRY, 1994a) so that (4) and (9) coincide when $\alpha = 0$ ($\gamma = \tau$ and $e_t = v_t$). More generally, $\Upsilon^i\tau \to \gamma$ as $i \to \infty$.

4. SOURCES OF FORECAST ERROR

We now consider dynamic forecasts from a complete and closed econometric system where parameters are subject to change in the forecast period. We draw on the analysis in CLEMENTS and HENDRY (1994a) but focus on potential biases arising from various problems. The forecast commences from the initial conditions denoted by $\hat{w}_T$, which may differ from the 'true' value ($\hat{w}_T \neq w_T$) due to poor measurement of provisional statistics (which in practice are subject to revision). Then, j-step ahead forecasts for the levels under assumed constant parameters are given by:

$$\hat{w}_{T+j} = \hat{\tau} + \hat{\Upsilon}\hat{w}_{T+j-1} = \left(\sum_{i=0}^{j-1} \hat{\Upsilon}^i\right)\hat{\tau} + \hat{\Upsilon}^j\hat{w}_T \text{ for } j = 1,\ldots, h. \tag{10}$$

Denote the forecast errors by $\hat{v}_{T+j} = w_{T+j} - \hat{w}_{T+j}$.

We consider the situation where the system experiences a step change between the estimation and forecast periods, such that $(\tau : \Upsilon)$ changes to $(\tau^* : \Upsilon^*)$ over $j = 1,\ldots, h$ and the variance, autocorrelation and distribution of the error change to $v_{T+j} \sim \mathsf{D}(0, \Omega^*)$. Thus, the data generated by the process for the next h periods is given by:

$$\begin{aligned} w_{T+j} &= \tau^* + \Upsilon^* w_{T+j-1} + v_{T+j} \\ &= (I + \Upsilon^*)\tau^* + (\Upsilon^*)^2 w_{T+j-2} + \Upsilon^* v_{T+j-1} + v_{T+j} \\ &= \sum_{i=0}^{j-1} (\Upsilon^*)^i \tau^* + \sum_{i=0}^{j-1} (\Upsilon^*)^i v_{T+j-i} + (\Upsilon^*)^j w_T. \end{aligned} \tag{11}$$

Then, the j-step ahead forecast error is decomposed into the six components:

$$\begin{aligned} \hat{v}_{T+j} &= \sum_{i=0}^{j-1} (\Upsilon^*)^i \tau^* + \sum_{i=0}^{j-1} (\Upsilon^*)^i v_{T+j-1} + (\Upsilon^*)^j w_T - \sum_{i=0}^{j-1} \hat{\Upsilon}^i \hat{\tau} - \hat{\Upsilon}^j \hat{w}_T \\ &\cong \left(\sum_{i=0}^{j-1} (\Upsilon^*)^i \tau^* - \sum_{i=0}^{j-1} \hat{\Upsilon}^i \hat{\tau}\right) + \sum_{i=0}^{j-1} (\Upsilon^*)^i v_{T+j-i} + \Upsilon_p^j (w_T - \hat{w}_T) \\ &\quad + \left[(\Upsilon^*)^j - \Upsilon^j\right] w_T + (\Upsilon^j - \Upsilon_p^j)\, w_T + (\Upsilon_p^j - \hat{\Upsilon}^j)\, w_T, \end{aligned} \tag{12}$$

where the estimated parameter multiplying the error in the initial conditions has been replaced by the sample period plim [dropping a term of $O_p(T^{-1})$]. The six components of forecast error are due to the changed intercepts; error accumulation; mis-measured initial conditions; parameter non-constancy; model mis-specification (i.e. inconsistent parameter estimates); and sampling variability. The last three terms could be written as $\left[(Y^*)^j - \hat{Y}^j\right] w_T$, but (12) clarifies the different effects of mistakes when econometric modelling.

To a first approximation, using $E[\hat{Y}] \cong Y_p$ and neglecting higher-order terms due to non-linearity, so the bias in the jth power is assumed negligible relative to other sources of forecast bias, from (12) the expectation of the j-step forecast error conditional on the correct initial condition w_T (which reflects what will transpire on average) is:

$$E\left[\hat{v}_{T+j} \mid w_T\right] \cong \left[\sum_{i=0}^{j-1} (Y^*)^i \tau^* - \sum_{i=0}^{j-1} Y_p^i \tau_p\right] + (Y^*)^j w_T - Y_p^j E\left[\hat{w}_T \mid w_T\right] \quad (13)$$

To understand the implications of (13), we note several special cases. First, even when a model is correctly specified in-sample (so that $\tau_p = \tau$ and $Y_p = Y$), forecasts are biased in general when the DGP is non-constant unless considerable cancellation occurs in (14):

$$E\left[\hat{v}_{T+j} \mid w_T\right] \cong \left[\sum_{i=0}^{j-1} (Y^*)^i \tau^* - \sum_{i=0}^{j-1} Y^i \tau\right] + (Y^*)^j w_T - Y^j E\left[\hat{w}_T \mid w_T\right] \quad (14)$$

Below, we consider the special case where only τ changes, and the initial conditions are unbiased (so that $E[\hat{w}_T \mid w_T] = w_T$), which yields:

$$E\left[\hat{v}_{T+j} \mid w_T\right] \cong \sum_{i=0}^{j-1} Y^i (\tau^* - \tau) \quad (15)$$

This more tractable expression highlights the main effects of a parameter change as we will see.

When the DGP remains constant and a congruent, if incomplete, model is used, the only bias arises from:

$$E[\hat{v}_{T+j} \mid w_T] \cong Y^j \left(w_T - E[\hat{w}_T \mid w_T]\right) \quad (16)$$

More accurate (i.e. unbiased) initial conditions would lead to unbiased forecasts in this setting. Again, the form of (16) captures the important implications of our approach.

Finally, the outcome for an inconsistent but constant model with correct initial conditions is:

$$\mathrm{E}\,[\hat{\mathrm{v}}_{T+j} \mid \mathrm{w}_T] \cong \left(\sum_{i=0}^{j-1} \Upsilon^i \tau - \sum_{i=0}^{j-1} \Upsilon_p^i \tau_p \right) + [\Upsilon^j - \Upsilon_p^j]\mathrm{w}_T \tag{17}$$

This remains complex to analyze, so we consider the special case when the intercept is zero and known to be such, so that:

$$\mathrm{E}\,[\hat{\mathrm{v}}_{T+j} \mid \mathrm{w}_T] \cong [\Upsilon^j - \Upsilon_p^j]\,\mathrm{w}_T \tag{18}$$

The next section analyses these cases in more detail.

When the system is a VECM, j-step ahead forecasts for the differences, $\Delta \mathrm{w}_{T+j}$ over $j = 1,\ldots, h$ are the changes in the levels:

$$\Delta \hat{\mathrm{w}}_{T+j} = \hat{\mathrm{w}}_{T+j} - \hat{\mathrm{w}}_{T+j-1}\,, \tag{19}$$

with forecast errors of $\hat{\mathrm{u}}_{T+j} = \Delta \mathrm{w}_{T+j} - \Delta \hat{\mathrm{w}}_{T+j} = \hat{\mathrm{v}}_{T+j} - \hat{\mathrm{v}}_{T+j-1}$. Thus, the forecast error in the j-step ahead difference is just the difference between the j and j - 1 step ahead errors in forecasting the levels of the series.

We now consider the forecasts from the DVAR. From (9), forecasts from the DVAR for $\Delta \mathrm{w}_t$ are defined by setting $\Delta \mathrm{w}_{T+j}$ equal to the (estimated) population growth rate $\hat{\gamma}$:

$$\Delta \tilde{\mathrm{w}}_{T+j} = \hat{\gamma} \tag{20}$$

so that j-step ahead forecasts of the level of the process are obtained by integrating (20) from the initial condition $\hat{\mathrm{w}}_T$:

$$\tilde{\mathrm{w}}_{T+j} = \tilde{\mathrm{w}}_{T+j-1} + \hat{\gamma} = \hat{\mathrm{w}}_T + j\,\hat{\gamma} \quad \text{for } j = 1, \ldots, h \tag{21}$$

From (11) for the actual values of the data generated by the (changed) process, and (21) for the forecasts from the DVAR, we obtain the j-step ahead forecast error decomposed as in (12) into the five components:

$$\begin{aligned}\tilde{v}_{T+j} &= \sum_{i=0}^{j-1} (\Upsilon^*)^i \tau^* + \sum_{i=0}^{j-1} (\Upsilon^*)^i v_{T+j-i} + (\Upsilon^*)^j w_T - j\hat{\gamma} - \hat{w}_T \\ &= \left(\sum_{i=0}^{j-1} (\Upsilon^*)^i \tau^* - j\hat{\gamma}\right) + \sum_{i=0}^{j-1} (\Upsilon^*)^i v_{T+j-i} + (w_T - \hat{w}_T) \\ &\quad + \left[(\Upsilon^*)^j - \Upsilon^j\right] w_T + (\Upsilon^j - I)\, w_T. \end{aligned} \tag{22}$$

Then the expectation of the j-step forecast error conditional on the correct initial condition w_T (cf. (13)) is given by:

$$E\left[\tilde{v}_{T+j} \mid w_T\right] \cong \left[\sum_{i=0}^{j-1} (\Upsilon^*)^i \tau^* - j\gamma\right] + (\Upsilon^*)^j w_T - E\left[\hat{w}_T \mid w_T\right] \tag{23}$$

Since $\hat{\gamma}$ is a consistent estimator of γ, we replaced γ_p by the population value, γ (see CLEMENTS and HENDRY, 1993c). As in the case of the correctly-specified model, we can obtain a better understanding of (23) by assuming that only τ changes, and the initial conditions are unbiased. Then compare the following with (15):

$$E\left[\tilde{v}_{T+j} \mid w_T\right] \cong \left(\sum_{i=0}^{j-1} \Upsilon^i \tau^* - j\gamma\right) + (\Upsilon^j - I)\, w_T \tag{24}$$

which is biased even when $\tau^* = \tau$. However, under such circumstances, CLEMENTS and HENDRY (1992) show that unconditionally it is approximately unbiased $E[\tilde{v}_{T+j}] \simeq 0$.

5. HOW CAN ECONOMETRICS HELP?

The six-fold partition of sources of forecast error can be condensed to five by treating τ, Υ and Ω as subsets of a generic parameter vector θ:

(*i*) parameter change: $\theta^* \neq \theta$;

(*ii*) model mis-specification: $\theta_p \neq \theta$;

(*iii*) estimation uncertainty: $V[\theta_p - \hat{\theta}] \neq 0$;

(*iv*) initial condition mis-measurement: $(w_T - \hat{w}_T) \neq 0$;

(*v*) error accumulation: $V\left[\sum_{i=0}^{j-1} (\Upsilon^*)^i v_{T+j-i}\right] \neq 0$.

All these sources of uncertainty can be reduced by appropriate techniques, though none can be eliminated.

(*i*) We note that parameter change may derive from lack of invariance to regime changes and begin with the special case in (15) where only τ changes. A number of general implications follow. First, after a step change, all the forecast errors have the same sign. Next, they are increasing in the horizon for the levels of the variables. Third, successive errors are related by:

$$E\left[\hat{u}_{T+j} \mid w_T\right] = E\left[\hat{v}_{T+j} \mid w_T\right] - E\left[\hat{v}_{T+j-1} \mid w_T\right] \cong \Upsilon^{j-1}(\tau^* - \tau) \tag{25}$$

so that forecasts of the change in w_t have biases which have the same sign but do not accumulate over steps ahead (cf. (15)). Thus, successive differences stabilize. Consider forecasting the acceleration in w_t, namely $\Delta^2 w_t$. From (25), the bias is given by:

$$E\left[\hat{u}_{T+j} \mid w_T\right] - E\left[\hat{u}_{T+j-1} \mid w_T\right] = \left(\Upsilon^{j-1} - \Upsilon^{j-2}\right)(\tau^* - \tau) \tag{26}$$

so that further differencing does remove most of the systematic bias that remains. However, the cost of doing so is to lose the long-run information embodied in the cointegrating vectors. More usefully, models can be reset 'on track', such that mis-prediction in the past leads to an offsetting correction when forecasting. HENDRY and CLEMENTS (1994) show that adding on the residual of the current period to the next period's forecast differences the original forecast errors for *1*-step ahead forecasts. This improves forecast performance (using expected squared forecast error) when $\rho > \frac{1}{2}$ where ρ is the autocorrelation coefficient between the original forecast errors. Thus, an appropriate intercept correction will remove any fixed component.

Does the DVAR offer any protection against the change in the intercept over the forecast period? From (24) we obtain:

$$E\left[\tilde{u}_{T+j} \mid w_T\right] = E\left[\tilde{v}_{T+j} \mid w_T\right] - E\left[\tilde{v}_{T+j-1} \mid w_T\right] \cong \Upsilon^{j-1}\tau^* - \gamma + (\Upsilon^j - \Upsilon^{j-1})\, w_T.$$

When $\Upsilon^{j-1}\tau^* - \gamma > 0$ and $w_T > 0$ (i.e. a positive shock when the initial condition is positive), the DVAR has smaller biases than the VECM since $\Upsilon^j < \Upsilon^{j-1}$. Below, we consider $\gamma = \tau = 0$ with $\tau^* > 0$ and $w_T > 0$ by commencing forecasts 2 periods after a break, but there is no clear outcome in general.

It is also worth considering the sequence of *h*-step forecasts at increasing *T*, since this more accurately reflects the calculations in the Monte Carlo illustration in section 7. That is, we analyse the sequence $E[\hat{v}_{T+j+s} \mid w_{T+s}]$ for fixed values of j (j = 1,4 in the illustration). For such a sequence, (12) is amended to:

$$\hat{v}_{T+j+s} = \sum_{i=0}^{j-1}(\Upsilon^*)^i \tau^* + \sum_{i=0}^{j-1}(\Upsilon^*)^i v_{T+j+s-i} + (\Upsilon^*)^j w_{T+s} - \sum_{i=0}^{j-1} \hat{\Upsilon}^i \hat{\tau} - \hat{\Upsilon}^j \hat{w}_{T+s} \tag{27}$$

so that when the model is correctly specified, only τ changes, and the initial conditions are unbiased (so that $E\left[\hat{w}_{T+s} \mid w_{T+s}\right] = w_{T+s}$), then the bias of (27) is just that of (15):

$$E\left[\hat{v}_{T+j+s} \mid w_{T+s}\right] \cong \sum_{i=0}^{j-1} \Upsilon^i (\tau^* - \tau) \tag{28}$$

However, (28) is independent of s, so that $E\,[\hat{v}_{T+j+s} \mid w_{T+s}] - E\,[\hat{v}_{T+j+s-1} \mid w_{T+s-1}] = 0$. This result rests on the particular form of parameter change, namely that the new value of τ^* is time invariant over the forecast period, so that the first term in (27) does not depend on when the forecast is made (the value of s). The general case of parameter change as in (14) is similar but there is less constancy to the forecast errors between periods.

For forecasts from the DVAR, the equivalent of (27) is given by:

$$\tilde{v}_{T+j+s} = \sum_{i=0}^{j-1} (\Upsilon^*)^i \tau^* + \sum_{i=0}^{j-1} (\Upsilon^*)^i v_{T+j+s-i} + (\Upsilon^*)^j w_{T+s} - j\,\hat{\gamma} - \hat{w}_{T+s} \tag{29}$$

so that again assuming that only τ changes and the initial conditions are unbiased, we obtain:

$$E\left[\tilde{v}_{T+j+s} \mid w_{T+s}\right] \cong \left(\sum_{i=0}^{j-1} \Upsilon^i \tau^* - j\gamma\right) + (\Upsilon^j - I)\, w_{T+s} \tag{30}$$

which resembles (24), but is not independent of s. Thus differencing yields:

$$E\left[\tilde{v}_{T+j+s} \mid w_{T+s}\right] - E\left[\tilde{v}_{T+j+s-1} \mid w_{T+s-1}\right] = (\Upsilon^j - I)\, \Delta w_{T+s} \tag{31}$$

For large j after the change in τ, then $(\Upsilon^j - I)\, E\,[\Delta w_{T+s}] \cong -\alpha(\beta'\alpha)^{-1}\beta'\gamma^* \neq 0$, but is likely to be small. A comparison of the bias in predicting the levels (28) for the correctly-specified model against (30) for the DVAR, again appears to be indeterminate for small j.

Finally, notice that the last three terms in the general decomposition (12) vanish when $w_T = 0$. For stationary transformations of the variables in the model (e.g. differences of

I (1) variables and cointegrating combinations), this can always be achieved by subtracting sample means. In that case, the only remaining term in (13) (assuming the initial conditions do not differ systematically from their true values) is:

$$E\left[\hat{v}_{T+j} \mid w_T = 0\right] \cong \left(\sum_{i=0}^{j-1} (\Upsilon^*)^i \gamma^* - \sum_{i=0}^{j-1} \Upsilon^i \gamma \right)$$

which highlights the role played by the (changed) growth rate, and motivates the attention we pay to the generation of forecasts that are to some extent robust to such changes.

(*ii*) When parameter estimates are inconsistent, the econometrician's advice must be to develop congruent empirical models: cf. HENDRY (1994). Econometrics would not be viable if it required the assumption that the model was the DGP, and it is possible to develop a theory of congruent models which do not need to be facsimiles of the DGP. An empirical model is the data population counterpart of the postulated theory specification, and arises as a reduction of the DGP (see HENDRY, 1987, 1993). Such a model is congruent if it matches the data evidence in all measurable respects – which does not require that it is either complete or correct. Key elements of congruency include that the empirical model's error is a homoscedastic innovation against the available information; that the conditioning variables are weakly exogenous for the parameters of the model; and that those parameters are constant. In addition, an encompassing model will be undominated in its class. It is possible to design models with all of these characteristics which thereby exhaust the available information and in essence represent the DGP up to an innovation error which cannot be distinguished from the actual DGP error on the given data. In the context of forecasting, therefore, Υ_p and Υ can be conflated in congruent models, and Ω is the 'smallest' that can be obtained given available information. Invalid specification could still induce inconsistency.

(*iii*) CLEMENTS and HENDRY (1994a) present the associated forecast-error variance formulae, but we focus on the central tendency here. When parameter values are uncertain, h-step forecast confidence intervals do not always increase monotonically as the horizon of the forecast increases, but may first expand then contract. Since the forecast variance must eventually equal the unconditional data variance for weakly stationary processes, conditional dynamic forecasts can have larger variances than unconditional. The calculations in HENDRY (1986) suggest that this can occur within the policy relevant forecast horizon which would allow scope for pooling forecasts (e.g. between unconditional and conditional predictors). Alternatively, there is scope for imposing parameter values rather than estimating them, since forecast error variances will fall when restrictions are nearly satisfied by the data: parsimonious yet valid parameterizations can pay dividends here (see CLEMENTS and HENDRY, 1993b).

(*iv*) Next, consider incorrect initial conditions when the DGP remains constant so the only bias is (16):

$$\mathrm{E}\left[\hat{\mathrm{v}}_{T+j} \mid \mathrm{w}_T\right] - \mathrm{E}\left[\hat{\mathrm{v}}_{T+j-1} \mid \mathrm{w}_T\right] \cong (\mathrm{Y}^j - \mathrm{Y}^{j-1}) \left(\mathrm{w}_T - \mathrm{E}\left[\mathrm{w}_T \mid \hat{\mathrm{w}}_T\right]\right).$$

This produces the same form of effect as in (i): changes between successive forecasts may be small, in which case an intercept correction can fix the problem. Badly measured initial values can distort short-term forecasts and WALLIS (1989, p.32) views measurement-errors as one explanation for intercept corrections. CLEMENTS and HENDRY (1994a) suggest several possibilities for improving current practice, using the econometric model to predict the initial conditions, intercept corrections, and extraneous information such as surveys.

(*v*) Error accumulation is pervasive as the horizon expands, and any predictibility in economic time series quickly fades. This may be exacerbated by $\{\mathrm{v}_t\}$ not being an innovation against the available information [which would also affect (ii)]. Further extensions of the information set will lower Ω, but may lead to less precisely estimated coefficients.

6. FORECAST EVALUATION

We can illustrate the limitations of comparisons based on MSFEs by considering an example based on BAILLIE (1993) and NEWBOLD (1993). They consider the comparison of forecasts from two mis-specified models of a first-order moving-average DGP, namely a first-order autoregression (denoted M_1) and white noise (M_2). They intended to establish that rankings by forecast accuracy could depend on the forecast horizon and hence warn against focusing exclusively on *1*-step ahead forecasts. They measure the costs of forecast errors at horizon h by MSFE_h in levels, defined for forecasting y_{T+h} by $\hat{y}_{T+h}$ as:

$$\mathrm{MSFE}_h = \mathrm{E}\left[(y_{T+h} - \hat{y}_{T+h})^2 \mid y_T\right].$$

The autoregression does best on *1*-step, whereas the white-noise prediction of zero 'wins' at any longer horizon, using MSFE_h as the criterion. A necessary condition for this outcome is that the two models are mis-specified for the DGP, since otherwise the correctly specified model would dominate at all horizons. This follows from the proposition that the conditional expectation is the minimum MSFE predictor.

Unfortunately, the result is not invariant to linear transformations, and e.g. for *2*-step forecasts using MSFE_2 for differences, we overturn the previous result relative to the *2*-step ahead ranking in levels: M_1 'beats' M_2 in both periods. Thus, the ranking depends crucially on the particular transformation of the variables for which forecast accuracy is assessed. Switches of rankings between linear transforms can be ruled out by using invariant criteria, such as that discussed in CLEMENTS and HENDRY (1993a), denoted below by G. This finding also serves to show that mis-specification is not sufficient for a rank reversal to occur.

Interestingly, if G is used as the evaluation criterion, in the BAILLIE-NEWBOLD example, as the forecast horizon increases, M_1 'beats' M_2 on average because of the gain on period 1. The G measure implicitly weights the levels and differences forecasts and so correctly highlights that on average, M_1 is the better model. Only if forecasters are certain that levels alone matter should they continue to use MSFE comparisons of predictions of the levels of the series.

Moreover, when models are mis-specified, it may be the case that the 'pooling' of forecasts produces combined forecasts that are superior to any individual forecast. When each forecast is model-based and derives from a common pool of information, then our advice is that the combination of the resulting forecasts is rarely a good idea. It is better to sort out the individual models and to derive a preferred model that contains the useful features of the original models, in line with the principle of encompassing. Briefly, encompassing is the requirement that an empirical model be able to account for the findings of rival models (see, HENDRY and RICHARD, 1982, MIZON, 1984, MIZON and RICHARD, 1986, and HENDRY and RICHARD, 1989).

Structural change over the forecast period will in general induce model mis-specification (over that period). Since the models will typically be differentially susceptible to the structural break, 'pooling' may produce better forecasts (see section 7). However our preferred strategy would be to attempt to 'robustify' the VECM via appropriate intercept corrections.

7. FORECASTING WITH A STRUCTURAL BREAK

In this section, we illustrate the analysis with an example based on a single realization of artificial data. Using PC-NAIVE (see HENDRY, NEALE and ERICSSON, 1991), 120 observations were generated for a bivariate cointegrated system (denoted y and z), over 1960:1 to 1989:4 with a break in 1984:4 (observation 100) when the growth of y was increased from 0 to 0.05 by doubling the intercept. The original value of τ ensured zero growth, whereas after the regime shift, substantial growth occurred. The data were generated to resemble that for real US M1 as shown in BABA, HENDRY and STARR (1992). Using PcFiml (DOORNIK and HENDRY, 1994), 1-step and 4-step ahead forecast errors were computed from four estimated models of the DGP: an unrestricted VAR in levels, a VAR with cointegration imposed (VECM), a DVAR and a DDVAR (i.e. a VAR in second differences). These were all estimated over the sample 1960:4 to 1985:2 (denoted T_0), and used to predict over the next 18 quarters (T_1).

Figure 1 shows the time series for various transformations of the data, and figures 2 to 9 record the forecast-period outcomes with *1*-step and the *4*-step forecasts bordered by ± 2 standard errors on either side (these ignore the asymptotically negligible parameter uncertainty, which in practice contributes relatively little here). The figures show the results for the four specifications of VAR, VECM, DVAR, and DDVAR, and the 8 blocks in each figure are for the transformations $\Delta^2 y_t$, $\Delta^2 z_t$, $\Delta(y-z)_t$, Δy_t, Δz_t, $(y-z)_t$, y_t and z_t, where the cointegration vector is $(y-z)_t$. In figure 1, the ninth graph is the time series of real money from BABA et al. (1992).

The VECM is the correct specification, the VAR has one cointegration vector too many, the DVAR one too few, and the DDVAR differences the data too often under the null of no break. It can be seen that in *1*-step comparisons, the DDVAR does well and the DVAR not too badly, but both VAR and VECM go off course for prolonged periods. This establishes that differencing offers robustness against step breaks in intercepts of the form described above, but at the cost of poorer fit and forecasting when no breaks occur. A better solution would be to retain the VECM and use the 'classical' intercept correction of setting the model back on track.

At 4-steps ahead, there is less to choose between the alternatives: all persist in long deviations for the levels because the model coefficients are not re-estimated as the sample increases, so again, intercept correction seems useful. Table 1 shows the components of the MSFEs, but we noted several caveats on that measure above. The means and standard deviations of the data over the two sub-periods show that noticeable changes occurred in the observations. The DDVAR specification is the most robust for the mean forecast error overall, followed by the DVAR; however, the DDVAR model does least well on average in terms of forecast-error standard deviations.

In section 6, we argued that model mis-specification paved the way for averaging forecasts over models. All the models considered in this section are mis-specified, if only over the forecast period (as in the case of the VECM). Since the models are differentially susceptible to the structural break, we considered an average of the forecasts from the VECM (standing for the structural econometric model here) and the DVAR (standing for the time-series method). The results show that the 'pooled' method does as well as any single forecasting device, particularly on the average forecast error for the levels and ECM transforms.

Table 1

Data means

	$\Delta^2 y$	$\Delta^2 z$	$\Delta(y-z)$	Δy	Δz	$(y-z)$	y	z
T_0	0.00	-0.00	0.00	-0.00	-0.00	1.0	0.38	-0.62
T_1	-0.01	-0.00	-0.02	0.05	0.07	1.3	1.13	-0.14

1-step forecast error means

	$\Delta^2 y$	$\Delta^2 z$
VAR	0.09	-0.02
ECM	0.09	-0.03
DVAR	0.01	0.08
DDVAR	-0.02	-0.01

4-step forecast error means

	$\Delta^2 y$	$\Delta^2 z$	$\Delta(y-z)$	Δy	Δz	$(y-z)$	y	z
VAR	0.01	-0.01	-0.02	-0.08	-0.07	-0.22	-0.36	-0.13
VECM	0.02	-0.01	-0.02	-0.07	-0.05	-0.22	-0.37	-0.14
DVAR	0.00	0.00	0.03	-0.03	-0.06	0.17	-0.10	-0.27
DDVAR	0.01	0.00	0.05	0.05	-0.00	0.14	0.12	-0.02
Pooled	0.02	-0.00	0.02	-0.01	-0.03	-0.04	-0.12	-0.08

Data standard deviations

	$\Delta^2 y$	$\Delta^2 z$	$\Delta(y-z)$	Δy	Δz	$(y-z)$	y	z
T_0	0.08	0.20	0.14	0.06	0.12	0.12	0.26	0.25
T_1	0.08	0.19	0.14	0.08	0.12	0.24	0.25	0.44

1-step forecast error standard deviations

	$\Delta^2 y$	$\Delta^2 z$
VAR	0.09	0.11
ECM	0.10	0.10
DVAR	0.07	0.11
DDVAR	0.07	0.14

4-step forecast error standard deviations

	$\Delta^2 y$	$\Delta^2 z$	$\Delta(y-z)$	Δy	Δz	$(y-z)$	y	z
VAR	0.09	0.19	0.16	0.07	0.12	0.24	0.19	0.17
VECM	0.10	0.19	0.15	0.09	0.12	0.22	0.26	0.14
DVAR	0.09	0.18	0.14	0.07	0.12	0.22	0.17	0.20
DDVAR	0.08	0.19	0.19	0.08	0.14	0.44	0.25	0.34
Pooled	0.09	0.18	0.14	0.07	0.12	0.20	0.17	0.20

Figures

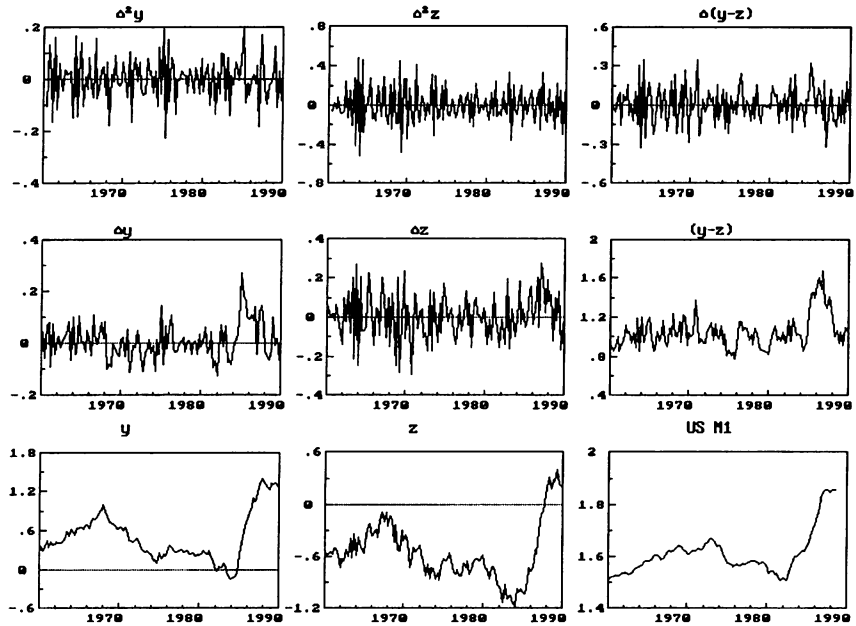

Figure 1: the artificial data time series for alternative transforms

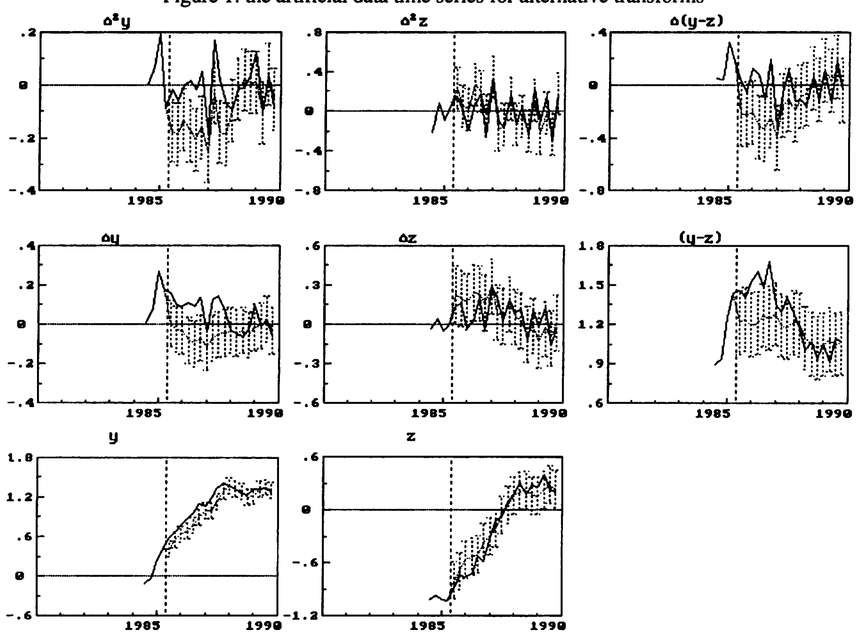

Figure 2: 1–step forecasts from the VAR

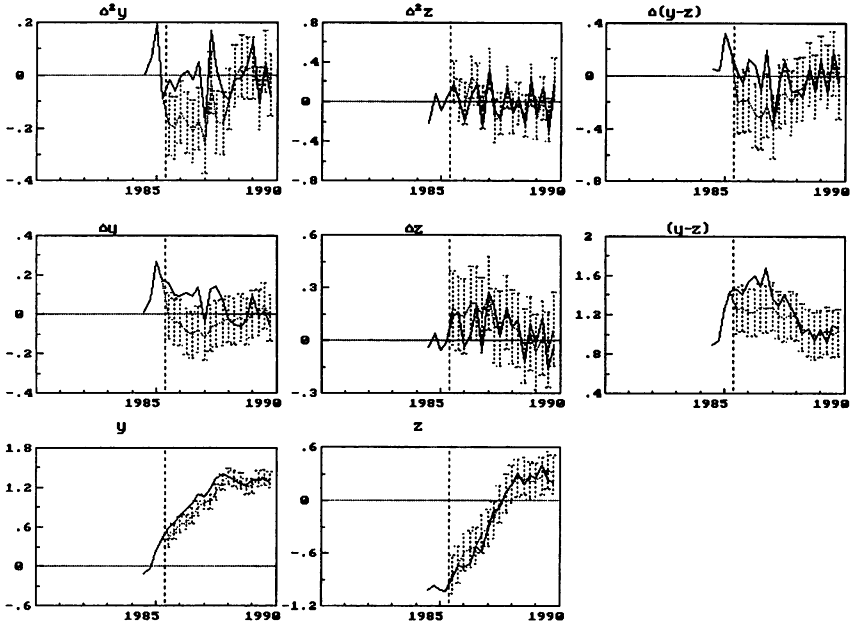

Figure 3: 1–step forecasts from the VECM

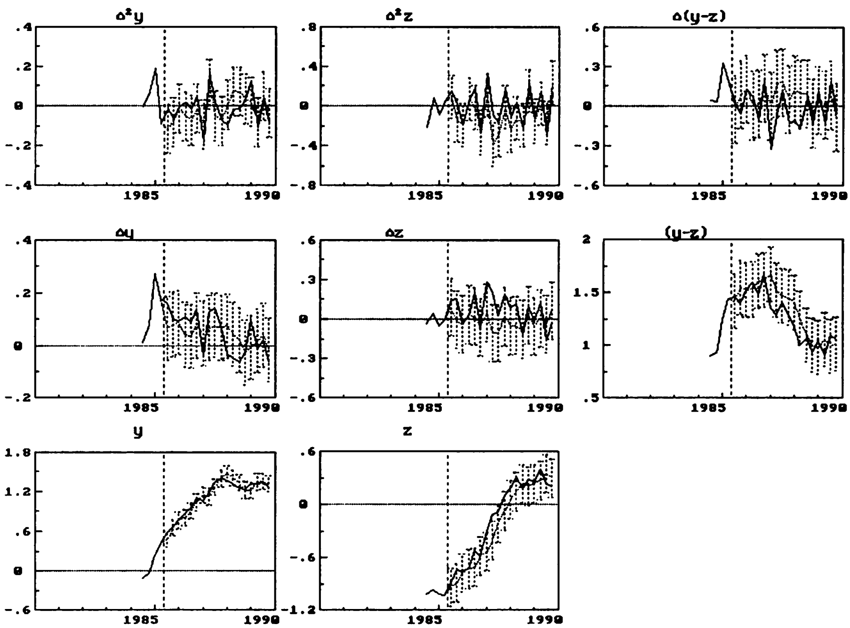

Figure 4: 1–step forecasts from the DVAR

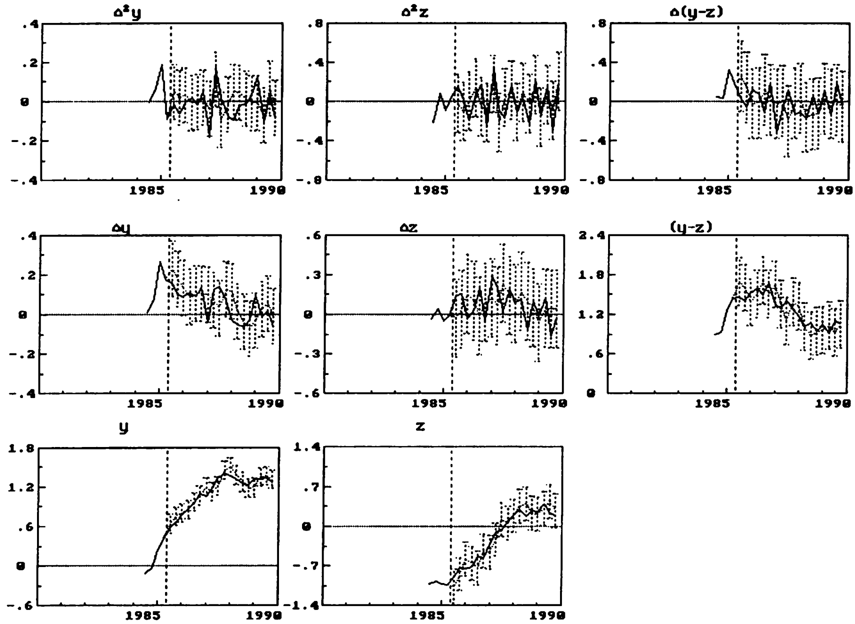

Figure 5: 1–step forecasts from the DDVAR

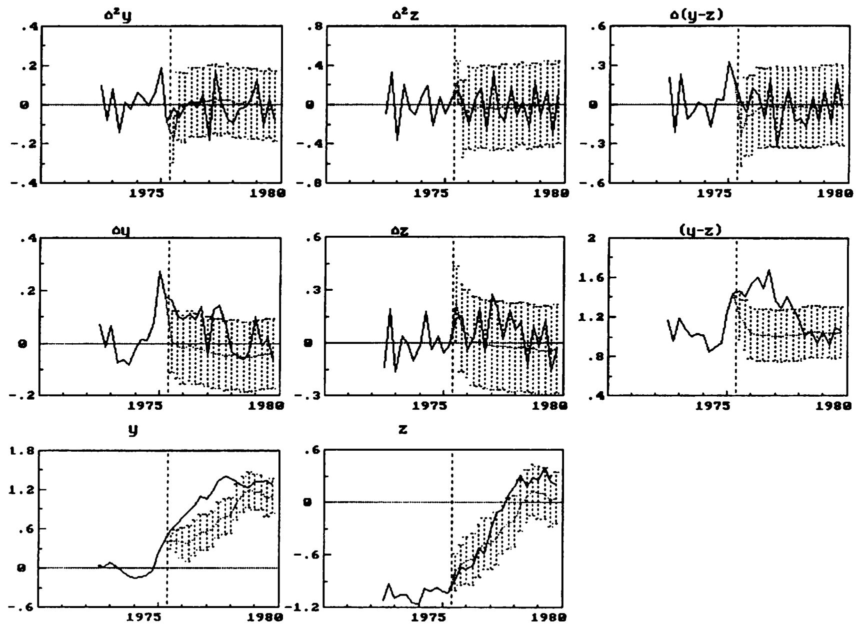

Figure 6: 4–step forecasts from the VAR

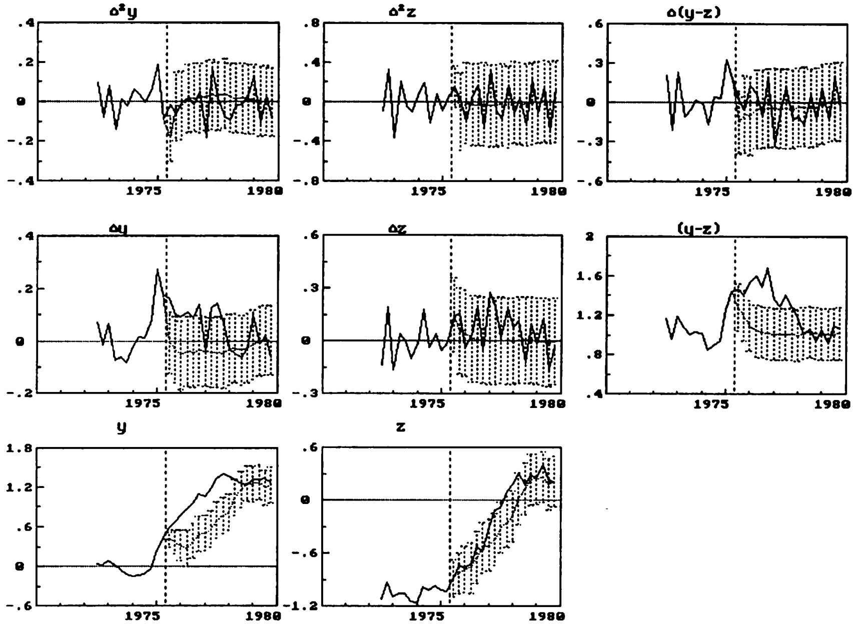

Figure 7: 4–step forecasts from the VECM

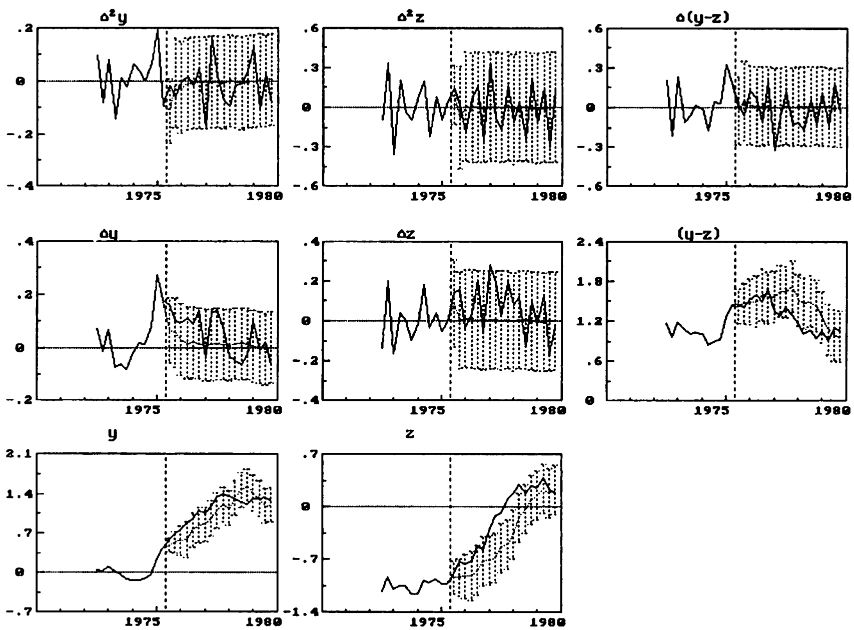

Figure 8: 4–step forecasts from the DVAR

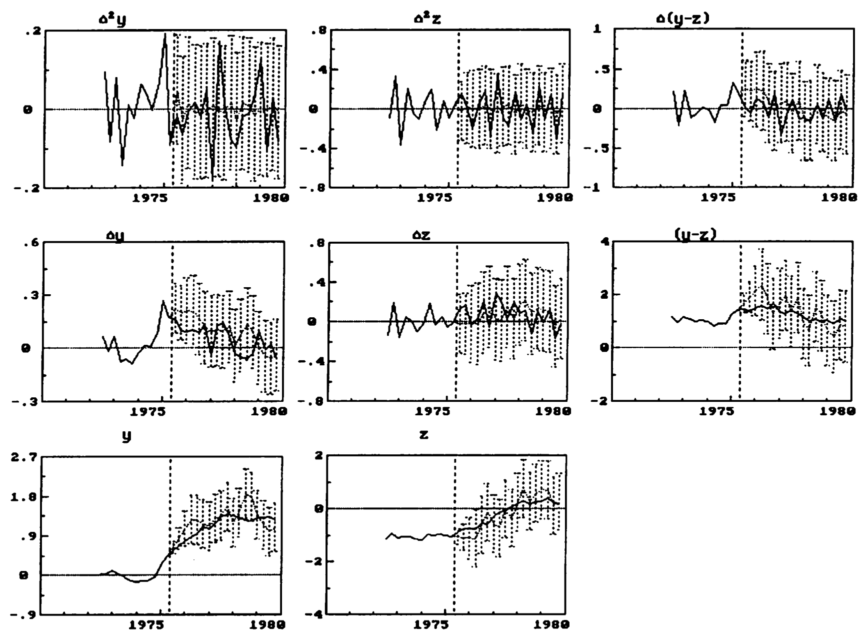

Figure 9: 4–step forecasts from the DDVAR

8. CONCLUSIONS

Econometrics can help economic forecasting. Econometric models provide a potentially scientific basis for economic forecasts and allow an assessment of the degree of confidence with which they may be held, an ingredient missing with some of the other methods of producing forecasts. We have focused on the positive aspects, in particular, on how good practice can partially mitigate errors in forecasting that are generated by the evolving nature of the economy and the impossibility of building models that are exact facsimiles of it. However, we have also noted certain pitfalls.

We have largely ignored the problem of model selection, which is dealt with elsewhere (see CLEMENTS and HENDRY, 1993c), but caution that in some circumstances, members of the ARIMA class of predictors can beat the estimated DGP. Parsimony appears more relevant in the context of forecasting than for fit or explanation, and imposing parameter values rather than estimating them may result in smaller overall forecast error variances. Thus *ex ante* forecast performance is not necessarily a reliable guide to model validity: the econometric model could be rejected in comparisons of *ex ante* forecast errors, even though it encompasses the time-series model. Such evidence would clarify the status of the rival predictors, and may ensure the future forecast dominance of the econometric model as corroborative evidence accrues, but cautions against drawing hasty conclusions from *ex ante* forecasting competitions between rival models.

Finally, this suggests a 'forecasting *versus* policy' dilemma: the DGP clearly provides the best guide to policy, but, due to the various modelling uncertainties, may not forecast as well as an extrapolative predictor despite the former being correctly specified and more informative. This is an uncomfortable tentative conclusion for those involved in forecasting and policy analysis, but points to the value of further research on this topic.

BIBLIOGRAPHY

BAILLIE, R.T. (1993), Comment on 'On the limitations of comparing mean squared forecast errors', by M.P. Clements and D.F. Hendry, Journal of Forecasting, 12, 639-41.

BABA, Y., HENDRY, D.F. and STARR, R.M. (1992), 'The Demand for M1 in the U.S.A., 1960-1988', Review of Economic Studies, 59, 25-61.

BOX, G.E.P. and JENKINS, G.M. (1970), Time Series Analysis Forecasting and Control, San Francisco: Holden-Day.

BROWN, T.M. (1954), 'Standard Errors of Forecast of a Complete Econometric Model', Econometrica, 22, 178-92.

BURNS, T. (1986), 'The Interpretation and Use of Economic Predictions', Proceedings of the Royal Society, A, no.407, 103-25. Reprinted in Predictability in Science and Society (ed.) J. Mason, P. Mathias and J.H. Westcott, London: Royal Society and British Academy.

CALZOLARI, G. (1981), 'A note on the variance of ex post forecasts in econometric models', Econometrica, 49, 1593-6.

CHATFIELD, C. (1993), 'Calculating Interval Forecasts', Journal of Business and Economic Statistics, 11, 2, 121-35.

CHONG, Y.Y. and HENDRY D.F. (1986), 'Econometric Evaluation of Linear Macro-Economic Models', Review of Economic Studies, LIII, 671-90.

CLEMENTS, M.P. (1994), 'Rationality and the Role of Judgement in Macroeconomic Forecasting', (Paper presented at the 1994 RES Conference held at Exeter), Institute of Economics and Statistics, Oxford.

CLEMENTS, M.P. and HENDRY, D.F. (1992), 'Forecasting in Cointegrated Systems', forthcoming, Journal of Applied Econometrics.

CLEMENTS, M.P. and HENDRY, D.F. (1993a), 'On the Limitations of Comparing Mean Square Forecast Errors', Journal of Forecasting, 12, 617-37.

CLEMENTS, M.P. and HENDRY, D.F. (1993b), 'Model Selection when Forecasting' unpublished paper, Institute of Economics and Statistics, Oxford.

CLEMENTS, M.P. and HENDRY, D.F. (1993c), 'Forecast Biases in Cointegrated Systems: analytical results using symmetry arguments', unpublished paper, Institute of Economics and Statistics, Oxford.

CLEMENTS, M.P. and HENDRY, D.F. (1994a), 'Towards a Theory of Economic Forecasting', in C. Hargreaves (ed.), Non-stationary Time-Series Analyses and Cointegration, Oxford: Oxford University Press. Forthcoming.

CLEMENTS, M.P. and HENDRY, D.F. (1994b), Economic Forecasting, Cambridge: Cambridge University Press. Forthcoming.

DOAN, T., LITTERMAN, R. and SIMS, C.A. (1984), 'Forecasting and conditional projection using realistic prior distributions', Econometric Reviews, 3, 1-100.

DOORNIK, J.A. and HENDRY, D.F. (1994), PcFiml version 8: An Interactive Econometric Modelling System, London: International Thomson Publishing.

ERICSSON, N.R. (1992), 'Parameter Constancy, Mean Square Forecast Errors, and Measuring Forecast Performance: An Exposition, Extensions, and Illustration', Journal of Policy Modeling, 14, 4, 465-98.

FAIR, R.C. (1986), 'Evaluating the Predictive Accuracy of Models', 1981-95 in Z. Griliches and M.D. Intriligator (eds), Handbook of Econometrics, Vol.3. Amsterdam: North-Holland.

FRIEDMAN, M. (1940), 'Review of J. Tinbergen: Statistical Testing of Business-cycle Theories, Vol. II: Business Cycles in the United States of America 1919-1932', American Economic Review, 30, 657-60.

GRANGER, C.W.J. and NEWBOLD, P. (1973), 'Some comments on the evaluation of economic forecasts', Applied Economics, 5, 35-47.

HAAVELMO, T. (1944), 'The Probability Approach in Econometrics', Supplement to Econometrica, 12.

HARVEY, A.C. (1981), Time Series Models. London: Philip Allan.

HARVEY, A.C. (1989), Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge, Cambridge University Press.

HENDRY, D.F. (1986), 'The Role of Prediction in Evaluating Econometric Models', Proceedings of the Royal Society, A407, 25-33.

HENDRY, D.F. (1987), 'Econometric Methodology: A Personal Perspective', Chapter 10 in T. Bewley (ed.), Advances in Econometrics, Cambridge: Cambridge University Press.

HENDRY, D.F. (1993), 'The Roles of Economic Theory and Econometrics in Time Series Economics', (Paper presented at the 1993 Meeting of the European Econometric Society), Institute of Economics and Statistics, Oxford.

HENDRY, D.F. (1994), Dynamic Econometrics. Oxford: Oxford University Press. Forthcoming.

HENDRY, D.F. and CLEMENTS, M.P. (1994), 'On a Theory of Intercept Corrections in Macro-Economic Forecasting', in S. Holly (ed.) Money, Inflation and Employment: Essays in Honour of Sir James Ball, Edward Elgar. Forthcoming.

HENDRY, D.F. and MORGAN, M.S. (1994), The Foundations of Econometric Analysis, forthcoming, Cambridge University Press.

HENDRY, D.F. and RICHARD, J-F. (1982), 'On the Formulation of Empirical Models in Dynamic Econometrics', Journal of Econometrics, 20, 3-33.

HENDRY, D.F. and RICHARD, J-F. (1989), 'Recent Developments in the Theory of Encompassing', in Contributions to Operations Research and Econometrics. The Twentieth Anniversary of Core, (ed.) B. Cornet and H. Tulkens. Cambridge, Mass.: MIT Press, 393-440.

HENDRY, D.F., A.J. NEALE and N.R. ERICSSON (1991), PC-NAIVE: An Interactive Program for Monte Carlo Experimentation in Econometrics, Version 6.1, Oxford Institute of Economics and Statistics.

HOLT, C. C. (1957), 'Forecasting seasonals and trends by exponentially weighted moving averages', ONR Research Memorandum 52, Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

JOHANSEN, S. (1988), 'Statistical Analysis of Cointegration Vectors', Journal of Economic Dynamics and Control, 12, 231-54.

KALMAN, R.E. (1960), 'A New Approach to Linear Filtering and Prediction Problems', Journal of Basic Engineering, 82, 35-45.

KEYNES, J.M. (1939), 'Professor Tinbergen's Method', Economic Journal, 49, 558-68.

KLEIN, L. R. (1950), Economic Fluctuations in the United States 1921-1941, Cowles Commission Monograph 11, New York: Wiley.

KLEIN, L.R. (1971), An Essay on the Theory of Economic Prediction, Chicago: Markham.

KLEIN, L.R., HOWREY, E.P. and MACCARTHY, M.D. (1974), 'Notes on Testing the Predictive Performance of Econometric Models', International Econometric Review, 15, 366-83.

KLEIN, L. R. (ed.) (1991) Comparative Performance of U.S. Econometric Models, New York: Oxford University Press.

KLEIN, L.R. and BURMEISTER, E. (eds.) (1976), Econometric Model Performance, Philadelphia: University of Pennsylvania Press.

KRÜGER, L., GIGERENZER, G. and MORGAN, M.S. (1987), The Probabilistic Revolution, Vol. 2, MIT Press.

LUCAS, R.E. (1976), 'Econometric Policy Evaluation: A Critique', in K. Brunner and A.H. Meltzer (eds.), The Phillips Curve and Labor Markets, 19-46, Amsterdam: North-Holland.

MARGET, A. W. (1929), 'Morgenstern on the Methodology of Economic Forecasting', Journal of Political Economy, 37, 312-39.

MARRIS, R.L. (1954), 'The Position of Economics and Economists in the Government Machine: a Comparative Critique of the United Kingdom and the Netherlands', Economic Journal, 64, 759-83.

MIZON, G.E. (1984), 'The Encompassing Approach in Econometrics', in D.F. Hendry and K.F. Wallis (eds) Econometrics and Quantitative Economics, Oxford: Basil Blackwell, 135-72

MIZON, G.E. and RICHARD, J-F. (1986), 'The Encompassing Principle and its Application to Non-Nested Hypothesis Tests', Econometrica, 54, 657-78.

MORGAN, M. S. (1990), The history of econometric ideas, Cambridge University Press.

MORGENSTERN, O. (1928), Wirtschaftsprognose: eine Untersuchung ihrer Voraussetzungen and Möglichkeiten, Vienna, Julius Springer.

NEWBOLD, P. (1993), Comment on 'On the limitations of comparing mean squared forecast errors', by M.P. Clements and D.F. Hendry, Journal of Forecasting, 12, 658-60.

PERSONS, W.M. (1925), 'Statistics and Economic Theory', Review of Economic Statistics, 7, 179-97.

ROBBINS, L. (1932), An Essay on the Nature and Significance of Economic Science, London: Macmillan.

SCHMIDT. P. (1974), 'The Asymptotic Distribution of Forecasts in the Dynamic Simulation of an Econometric Model', Econometrica, 42, 303-9.

SCHMIDT, P. (1977), 'Some small sample evidence on the distribution of dynamic simulation forecasts', Econometrica, 45, 97-1005.

THEIL, H. (1958), Economic Forecasts and Policy, North-Holland. (1961, Second edition).

THEIL, H. (1966), Applied Economic Forecasting, North-Holland, Amsterdam.

TINBERGEN, J. (1939), Statistical Testing of Business-Cycle Theories, Vol. I: A Method and its Application to Investment Activity; Vol. II: Buisiness Cycles in the United States of America, 1919-1932, Geneva: League of Nations.

TURNER, D.S. (1990), 'The Role of Judgement in Macroeconomic Forecasting', Journal of Forecasting, 9, 315-345.

WALLIS, K.F. (1989), 'Macroeconomic forecasting: a survey', Economic Journal, 99, 28-61.

WALLIS, K.F. and WHITLEY, J.D. (1991), 'Sources of Error in Forecasts and Expectations: U.K. Economic Models 1984-8', Journal of Forecasting, 10, 231-53.

WALLIS, K.F., ANDREWS, M.J., BELL, D.N.F, FISHER, P.G., and WHITLEY, J.D. (1984), Models of the UK Economy: A Review by the ESRC Macroeconomic Modelling Bureau, Oxford: Oxford University Press.

WALLIS, K.F., ANDREWS, M.J., BELL, D.N.F, FISHER, P.G., and WHITLEY, J.D. (1985), Models of the UK Economy: A Second Review by the ESRC Macroeconomic Modelling Bureau, Oxford: Oxford University Press.

WALLIS, K.F., ANDREWS, M.J., FISHER, P.G., LONGBOTTOM, J.A, and WHITLEY, J.D. (1986), Models of the UK Economy: A Third Review by the ESRC Macroeconomic Modelling Bureau, Oxford: Oxford University Press.

WALLIS, K.F., FISHER, P.G., LONGBOTTOM, J.A, TURNER, D.S. and WHITLEY, J.D. (1987), Models of the UK Economy: A Fourth Review by the ESRC Macroeconomic Modelling Bureau, Oxford: Oxford University Press.

WHITTLE, P. (1963), Prediction and Regulation. London: English Universities Press.

WIENER, N. (1949), Extrapolation, Interpolation and Smoothing of Stationary Time Series. New York:Wiley.

WINTERS, P. R. (1960), 'Forecasting sales by exponentially weighted moving averages', Management Science, 6, 324-42.

SUMMARY

After reviewing the history of analyses of economic forecasting, the role of econometrics in improving economic forecasting is considered, building on CLEMENTS and HENDRY (1994a). The basis of the analysis is a world where model selection is difficult, no model coincides with the economic mechanism, and that mechanism is both non-stationary and evolves over time. On the constructive side, econometric analysis suggests ways of reducing each of the resulting five sources of forecast uncertainty (parameter non-constancy; estimation uncertainty; variable uncertainty; innovation uncertainty; and model mis-specification). On the critical side, the lack of invariance of forecast evaluation procedures to the representation of the model may camouflage inadequate models. We show that forecasts generated from vector autoregressions in differences may be more robust to certain forms of structural change over the forecast period, and that a similar result can be achieved by suitable forms of intercept corrections in vector error-correction mechanisms.