Forecast Failure, Expectations Formation, and the Lucas Critique

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

March 19, 2002

Abstract

Since forecast failure is due to unanticipated deterministic shifts, 'sensible' agents should adopt 'robust forecasting rules'. In such a non-stationary world, causal variables can dominate non-causal in forecasting, so 'rational expectations' do not have a sound basis: agents cannot know how all relevant information enters the data density at every point in time. Although econometric models 'break down' intermittently, that is not due to the Lucas critique and need not preclude policy analyses.

Contents

*Financial support from the UK Economic and Social Research Council under grant R000237500 is gratefully acknowledged, as are comments from Guillaume Chevillon, Mike Clements, Roger Farmer, Stéphane Gregoir, Søren Johansen, Grayham Mizon and Ken Wallis.

1 Introduction

There are important implications for modelling expectations formation in economics from forecast failure being due to unanticipated large shifts in deterministic factors: see Clements and Hendry (1998a, 1999a) for documentation, and Clements and Hendry (1998b) for a general discussion of the implications of forecast failure. Deterministic variables include intercepts, linear trends etc. (such as $1, 1, 1, \ldots$; and $1, 2, 3, \ldots$), namely variables whose future values are known with certainty when they follow the given sequence. Those authors then define deterministic shifts as any change in the mean component (i.e., unconditional expectation of the non-integrated – $\mathsf{I}(0)$ – transformations) of the variables under analysis, both directly or induced by shifts in parameters of deterministic variables. Consider the simplest linear trend model of a variable y_t with $\alpha \neq 0$ and $\beta \neq 0$:

$$y_t = \alpha \times 1 + \beta \times t + u_t \tag{1}$$

where the precise properties of $\{u_t\}$ need not be specified beyond being $\mathsf{I}(0)$, and the multiplication is to highlight the interpretation in (2) below. A deterministic shift from time T onwards in (1) induces:

$$y_{T+h} = \alpha^* \times 1 + \beta^* \times (T+h) + u_{T+h} \equiv \alpha \times \mu + \beta \times (T+h)^* + u_{T+h} \tag{2}$$

where $\mu = \alpha^*/\alpha \neq 1$ is the shifted intercept, and $(T+h)^* = (T+h)\,\beta^*/\beta$ is an altered trend growth rate. Thus, although intercepts and linear trends are themselves deterministic, they play a problematic role in econometric models since the combined effect of the parameter times the variable is not deterministic: many future changes are unlikely to be foreseeable, let alone known.

Deterministic shifts in this sense transpire to be the main explanation for forecast failure, namely a significant deterioration in forecast performance relative to the anticipated outcome, usually based on the historical performance of a model. Other factors that mimic deterministic shifts, such as mis-estimating or mis-specifying deterministic components in models (e.g., omitting the trend in 1)), are included under that heading. Shifts in other parameters of models appear to have effects that are in practice difficult to detect: see Hendry and Doornik (1997) and Hendry (2000b).

The aim of this paper is to explore the implications for agents' expectations formation of these and related results on the causes of forecast failure. The following analysis is based on the fact that economies are non-stationary, and sufficiently high-dimensional and complicated that it is an impossible requirement for a model to coincide with the data generating process (DGP) in macroeconomics: see Barrell (2001) who presents six examples of endemic structural change since the 1990s, and Clements and Hendry (2001b) who consider the historical prevalence of forecast failure in UK output forecasts and their association with major 'economic events'. Such evidence is buttressed by the outcomes of the major empirical forecasting competitions, such as Makridakis, Andersen, Carbone, Fildes *et al.* (1982), reviewed by Fildes and Makridakis (1995), and more recently, Makridakis and Hibon (2000) discussed by Clements and Hendry (2001a) and Fildes and Ord (2001). Although which model does best in such forecasting competitions depends on how the forecasts are evaluated and what horizons and samples are selected, a class of 'simple' extrapolative methods tends to outperform econometric systems. However, all these findings can be explained by a theory of mis-specified models of changing DGPs: see Clements and Hendry (1999b) and section 6.2. We refer to a theory assuming correctly specified models of constant DGPs as 'optimality theory' following Makridakis and Hibon (2000).

The structure of the paper is as follows. Section 2 outlines the six main implications of observed forecast failure that will form the body of the paper. Then section 3 reviews the sources of such forecast failure, leading to the taxonomy of forecast errors in section 3.1. Section 4 discusses the detectability of deterministic shifts, and the difficulty of detecting other structural breaks. Section 5 then considers robust forecasting rules which seem to 'win' forecasting competitions, investigates why they do so, and proposes that 'sensible' agents will adopt such rules. The potential dominance in forecasting of 'non-causal variables' is shown: a 'causal variable' is one which actually enters the DGP with a non-zero effect; a non-causal variable does not enter the DGP. Section 6 discusses the assumptions needed to sustain 'rational expectations', and highlights the epistemological problems they raise when the data are generated by an inherently non-stationary process of the kind needed to account for the very forecast failure that prompted their rise to prominence in empirical macro-econometrics. Section 7 notes results from Hendry and Mizon (2000a, 2000b) on the inappropriateness of selecting policy models by forecast accuracy, and the converse that forecast failure does not preclude the policy use of econometric models. Section 8 concludes.

2 Six implications of observed forecast failure

The first implication of the theory which accounts for observed forecast failure is that one can disprove the basic 'optimality theorem' that forecasting based on causal variables will dominate that using non-causal: see Hendry (1997) and section 6.1. Since 'rational expectations' rely on causal claims – being the correct conditional expectations – they do not, therefore, have a sound theoretical basis, an issue amplified in section 6.

The second implication is that the methods which should do best in forecasting competitions are those which are relatively immune to unmodelled deterministic shifts: Clements and Hendry (1999b) present evidence supporting that view. Robust forecasting methods are generally different from 'rational expectations', which are not robust in the face of unanticipated deterministic shifts: this aspect is followed up in section 6.4.

Thirdly, in such a non-stationary world, forming 'rational expectations' requires agents not only to know all the relevant information, but also to know how every item of information enters the joint data density at every point in time, when many of the events involved cannot be unanticipated. 'Sensible' agents who learn that they cannot form expectations any better than the best forecasting models may well adopt such 'robust forecasting rules': see section 6.3.

Fourthly, these criticisms apply *a fortiori* to 'model consistent' expectations, which seem to embody the worst of all possible worlds: they are neither rational nor robust unless the model is correctly specified over the future. 'Rational agents' would be wise not to use model-based expectations, when models – including their own – are manifestly mis-specified (for earlier critiques, see *inter alia*, Pesaran, 1987): section 6.5 amplifies.

Fifthly, should agents adopt such robust forecasting rules, they have the property that their form does not usually vary with policy changes (although the resulting forecasts do), rendering the Lucas (1976) critique otiose: see section 6.6.

The final, and closely related, implication is that although econometric models 'break down' intermittently, not only is this unlikely to be due to the 'Lucas critique', but rather suggests its irrelevance, consistent with the increasing empirical evidence of its absence in practice: see Ericsson and Irons (1995) and section 6.6.

The remainder of the paper establishes these six implications.

3 Sources of forecast failure

Forecast failure should rarely occur in a constant-parameter, stationary world, albeit that it is all too common in practice. Many 'conventional' results that can be proved in a constant-parameter, stationary-data world change radically under parameter non-constancy as we have noted: examples include the potential dominance in forecasting of causal variables by non-causal, of well-specified models by badly mis-specified, of 1-step forecasts by multi-step, and of known parameter values by estimated, as well as the value-added of intercept corrections and differencing transforms: see Clements and Hendry (1998a, 1999a).

The first of these examples is central to the implications of expectations formation for econometric modelling. If forecasting devices using no causally-relevant variables can outperform those based on the best-available causal information, there is no 'approximately-correct' concept applicable to 'rational expectations': unless the underlying model held by all relevant agents is perfectly specified, it can be dominated in principle by non-causal devices – and in practice, such devices regularly 'win' forecasting competitions: see Makridakis *et al.* (1982).

Forecasting models have three main components: deterministic terms, whose future values are known; observed stochastic variables with unknown future values; and unobserved errors all of whose values are unknown. Any, or all, of these three components potentially could be the source of forecast failure. Moreover, a model's relationships could be: mis-specified; poorly estimated; based on inaccurate data; incorrectly selected by data-based methods; involve collinear variables with non-parsimonious formulations; or suffer structural breaks. Given the complexity of modern economies, most of these possible causes will be present in any empirical macro-model, and will reduce forecast performance, either from inaccuracy or imprecision. However, some mistakes have more pernicious effects on forecasts than others, and most combinations do not seem to induce systematic forecast failure: theoretical and simulation analyses and empirical evidence implicate structural breaks in deterministic terms as the primary cause of forecast failure.

The potential sources of forecast errors include mistakes that derive from:

(1) formulating a forecasting model from an inadequate theory,
(2) selecting by inappropriate empirical criteria,
(3) mis-specifying the deterministic components,
(4) mis-specifying the stochastic components,
(5) mis-estimating the parameters,
(6) using inaccurate observations,
(7) subject to intermittent structural breaks.

Such a framework closely mimics the apparent empirical world, so any resulting forecast-error taxonomy must include a source for each of the effects in 1–7, partitioned appropriately for deterministic, observed-stochastic, and error influences. Although the decompositions of the resulting forecasting mistakes are not unique, for any given system they can be expressed in nearly-orthogonal effects corresponding to influences on forecast-error means and variances respectively (higher moments undoubtedly alter as well, but that will not be of concern here). The former involve all the deterministic terms; the latter the remainder. We now consider these major categories of error for a first-order vector autoregression (VAR) describing an integrated-cointegrated process that is subject to intermittent structural breaks. A general (i.e., model-free) forecast-error taxonomy can be developed, and delivers the same implications as the simpler exemplar used here: see Hendry (2000a).

3.1 Forecast-error taxonomy

Consider a vector of n stochastic variables $\{\mathbf{x_t}\}$, where the sequential expectation of $\mathbf{x_t}$ at time t is $\mathsf{E}_t[\mathbf{x_t}|\mathbf{X^1_{t-1}}, \mathbf{q_t}]$ (where that exists), conditional on information $\mathbf{X^1_{t-1}} = (\mathbf{x_1} \ldots \mathbf{x_{t-1}})$ available at the time, when $\mathbf{q_t}$ denotes the relevant deterministic factors (such as intercepts, trends, and indicator variables), with $\mathbf{Q^1_t} = (\mathbf{q_1} \ldots \mathbf{q_t})$. Because the underlying densities may be changing at each point in time, all expectations operators must be time dated. It is desired to forecast a function of $\mathbf{x_{T+h}}$, such as itself, or perhaps the non-integrated components, denoted $\{\mathbf{y_{T+h}}\}$ below, over forecast horizons $h = 1, \ldots, H$, from a forecast origin at T. Forecast accuracy is to be judged by a criterion function $\mathbf{C_e}\left(\mathbf{e_{T+1|T}} \ldots \mathbf{e_{T+H|T}}\right)$, which we take to depend only on the forecast errors $\mathbf{e_{T+h|T}} = \mathbf{x_{T+h}} - \widehat{\mathbf{x}}_{\mathbf{T+h|T}}$, where 'smaller' values of $\mathbf{C_e}(\cdot)$ are preferable. Even so, unless the complete joint density is used, evaluation outcomes depend on the specific transformation of $\mathbf{x_{T+h}}$ considered.

To illustrate the sources of forecast error concretely, we focus on a first-order vector autoregressive (VAR) DGP which is stationary in-sample (perhaps after appropriate cointegration and differencing transformations), written as:

$$\mathbf{y_t} = \boldsymbol{\phi} + \boldsymbol{\Pi}\mathbf{y_{t-1}} + \boldsymbol{\epsilon}_\mathbf{t}, \tag{3}$$

with intercept $\boldsymbol{\phi}$ and dynamic matrix $\boldsymbol{\Pi}$ (all of whose eigenvalues lie inside the unit circle) where $\boldsymbol{\epsilon}_\mathbf{t} \sim \mathsf{IN_n}\left[\mathbf{0}, \boldsymbol{\Omega}_\epsilon\right]$. The forecaster uses the corresponding forecasting model:

$$\widehat{\mathbf{y}}_{T+h|T+h-1} = \widehat{\boldsymbol{\phi}} + \widehat{\boldsymbol{\Pi}}\mathbf{y_{T+h-1}} \tag{4}$$

where $\widehat{\ }$ denotes estimates for parameters, and forecasts for random variables. In (3), when $\mathbf{x_t}$ is $\mathsf{I}(1)$ and there are $r < n$ cointegrating vectors $\boldsymbol{\beta}$, the $\mathbf{y_t}$ comprise r elements $\boldsymbol{\beta}'\mathbf{x_t}$, and $n-r$ elements of $\Delta\mathbf{x_t}$, so are $\mathsf{I}(0)$ by construction. Further, the only deterministic term retained from $\mathbf{q_t}$ is the intercept (although this allows a trend in the $\mathbf{x_t}$). The forecast-error transformation of interest is $\widehat{\boldsymbol{\epsilon}}_{T+h|T} = \mathbf{y_{T+h}} - \widehat{\mathbf{y}}_{\mathbf{T+h|T}}$, and $\mathbf{C}_{\widehat{\boldsymbol{\epsilon}}}(\cdot)$ will comprise only the mean and variance of $\{\widehat{\boldsymbol{\epsilon}}_{T+h|T}\}$. It is important to remember that the detailed results are dependent on these choices: 'forecast-accuracy' measures depend on both the data transformation inspected and the loss function (see e.g., Clements and Hendry, 1993).

The unconditional mean of $\mathbf{y_t}$ is:

$$\mathsf{E}\left[\mathbf{y_t}\right] = \left(\mathbf{I_n} - \boldsymbol{\Pi}\right)^{-1}\boldsymbol{\phi} = \boldsymbol{\varphi} \tag{5}$$

and hence, in deviation form, the DGP is:

$$\mathbf{y_t} - \boldsymbol{\varphi} = \boldsymbol{\Pi}\left(\mathbf{y_{t-1}} - \boldsymbol{\varphi}\right) + \boldsymbol{\epsilon}_\mathbf{t}.$$

The h-step ahead forecasts at time T for $h = 1, \ldots, H$ from (4) can be written as:

$$\widehat{\mathbf{y}}_{T+h} - \widehat{\boldsymbol{\varphi}} = \widehat{\boldsymbol{\Pi}}\left(\widehat{\mathbf{y}}_{T+h-1} - \widehat{\boldsymbol{\varphi}}\right) = \widehat{\boldsymbol{\Pi}}^h\left(\widehat{\mathbf{y}}_T - \widehat{\boldsymbol{\varphi}}\right), \tag{6}$$

where $\widehat{\boldsymbol{\varphi}} = (\mathbf{I_n} - \widehat{\boldsymbol{\Pi}})^{-1}\widehat{\boldsymbol{\phi}}$. Although $\mathbf{y_T}$ is unknown in practice, we assume $\mathsf{E}[\widehat{\mathbf{y}}_\mathsf{T}] = \boldsymbol{\varphi}$ here, so that on average no systematic bias results from data inaccuracy – in practice, however, the forecast origin might be unavailable or less well measured than earlier data (see e.g., Wallis, 1986) and such mis-measurement can be an important source of systematic forecast errors.

After the forecasts have been made at time T, the DGP parameter values $(\boldsymbol{\phi} : \mathbf{\Pi})$ change to $(\boldsymbol{\phi}^* : \mathbf{\Pi}^*)$, where $\mathbf{\Pi}^*$ still has all its eigenvalues less than unity in absolute value, so the process remains I(0). The break may well be endogenously induced, perhaps as a proxy for a 'sharp' non-linearity, or may reflect changes elsewhere in the economic system, but the implications of its not being modelled appear unrelated to its source, though not to prospects of eventually developing models which anticipate such breaks. Then, from $T+1$ onwards, the data are generated by:

$$\begin{aligned}
\mathbf{y_{T+h}} &= \boldsymbol{\phi}^* + \mathbf{\Pi}^* \mathbf{y_{T+h-1}} + \boldsymbol{\epsilon}_{\mathbf{T+h}} \\
&= \boldsymbol{\varphi}^* + \mathbf{\Pi}^* \left(\mathbf{y_{T+h-1}} - \boldsymbol{\varphi}^*\right) + \boldsymbol{\epsilon}_{\mathbf{T+h}} \\
&= \boldsymbol{\varphi}^* + \left(\mathbf{\Pi}^*\right)^{\mathbf{h}} \left(\mathbf{y_T} - \boldsymbol{\varphi}^*\right) + \sum_{\mathbf{i=0}}^{\mathbf{h-1}} \left(\mathbf{\Pi}^*\right)^{\mathbf{i}} \boldsymbol{\epsilon}_{\mathbf{T+h-i}},
\end{aligned} \tag{7}$$

so both the slope and the intercept alter, where $\boldsymbol{\varphi}^* = \left(\mathbf{I_n} - \mathbf{\Pi}^*\right)^{-\mathbf{1}} \boldsymbol{\phi}^*$. Let $\widehat{\boldsymbol{\epsilon}}_{T+h|T} = \mathbf{y_{T+h}} - \widehat{\mathbf{y}}_{\mathbf{T+h|T}}$, then the forecast-error taxonomy in (8) decomposes the various sources into eight interpretable, and potentially freely-varying, components (see Clements and Hendry, 1998a):[1]

VAR forecast-error taxonomy

$$\begin{aligned}
\widehat{\boldsymbol{\epsilon}}_{T+h|T} \simeq\; & \left(\mathbf{I_n} - \left(\mathbf{\Pi}^*\right)^{\mathbf{h}}\right)\left(\boldsymbol{\varphi}^* - \boldsymbol{\varphi}\right) && (\textit{ia}) \text{ equilibrium-mean change} \\
& + \left(\left(\mathbf{\Pi}^*\right)^{h} - \mathbf{\Pi}^{\mathbf{h}}\right)\left(\mathbf{y_T} - \boldsymbol{\varphi}\right) && (\textit{ib}) \text{ dynamic change} \\
& + \left(\mathbf{I_n} - \mathbf{\Pi}_{\mathbf{p}}^{\mathbf{h}}\right)\left(\boldsymbol{\varphi} - \boldsymbol{\varphi}_{\mathbf{p}}\right) && (\textit{iia}) \text{ equilibrium-mean mis-specification} \\
& + \left(\mathbf{\Pi}^{\mathbf{h}} - \mathbf{\Pi}_{\mathbf{p}}^{\mathbf{h}}\right)\left(\mathbf{y_T} - \boldsymbol{\varphi}\right) && (\textit{iib}) \text{ dynamic mis-specification} \\
& - \left(\mathbf{I_n} - \mathbf{\Pi}_{\mathbf{p}}^{\mathbf{h}}\right)\left(\widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_{\mathbf{p}}\right) && (\textit{iiia}) \text{ equilibrium-mean estimation} \\
& - \mathbf{F_h}\left(\widehat{\mathbf{\Pi}} - \mathbf{\Pi}\right)^{\boldsymbol{\nu}} && (\textit{iiib}) \text{ dynamic estimation} \\
& - \left(\mathbf{\Pi}_{\mathbf{p}}^{\mathbf{h}} + \mathbf{C_h}\right)\left(\mathbf{y_T} - \widehat{\mathbf{y}}_{\mathbf{T}}\right) && (\textit{iv}) \text{ forecast-origin uncertainty} \\
& + \sum_{i=0}^{h-1} \left(\mathbf{\Pi}^*\right)^{i} \boldsymbol{\epsilon}_{\mathbf{T+h-i}} && (\textit{v}) \text{ error accumulation.}
\end{aligned} \tag{8}$$

The first two terms isolate the impacts of changes in DGP parameter values on the mean and dynamics respectively; the next two isolate (so far as possible), the effects of mis-specification, again separately for the mean and dynamics; the next pair show where estimation variation enters; the seventh term reflects all the uncertainty around the forecast-origin; and the last term is the unavoidable error accumulation as the forecast horizon lengthens. In (8), terms involving $\mathbf{y_T} - \boldsymbol{\varphi}$ have zero expectations even under changed parameters (e.g., (*ib*) and (*iib*)). Moreover, for symmetrically-distributed shocks, biases in $\widehat{\mathbf{\Pi}}$ for $\mathbf{\Pi}$ will not induce biased forecasts (see e.g., Clements and Hendry, 1998a), and the $\boldsymbol{\epsilon}_{\mathbf{T+h}}$ have zero means by construction. Consequently, the primary sources of systematic forecast failure are (*ia*), (*iia*), (*iiia*), and (*iv*). However, on *ex post* evaluation, (*iv*) will be removed, and in congruent models with freely-estimated intercepts, (*iia*) and (*iiia*) will be zero on average. That leaves (*ia*) as the main source of bias: the remaining terms add variance effects, discussed in section 3.3. We now consider these implications in more detail, commencing with mean effects, then turn to variance components, and return to the implications for 'rational expectations' of (8) in section 6.

[1] The matrices $\mathbf{C_h}$ and $\mathbf{F_h}$ are complicated functions of the whole-sample data set, the method of estimation, and the forecast-horizon (see e.g., Calzolari, 1981), $(\cdot)^{\nu}$ denotes column vectoring, and the subscript p denotes a plim.

3.2 Mean effects

The key effect derives from changes to the 'equilibrium mean' (not necessarily the intercept in a model, as seen in (8)), namely the unconditional expectation $\mathsf{E}_t[\mathbf{y_t}]$ (where that exists) of the non-integrated components of the vector of variables under analysis. Because the model will generally be a mis-specified representation of the DGP, a different expectations operator, $\mathcal{E}[\cdot]$, is used to denote expectations based on the estimated model. Consequently, $\mathsf{E}_{\mathsf{T+h}}[\mathbf{y_{T+h}}] - \mathcal{E}_{\mathcal{T}}[\mathbf{y_{T+h}}]$ is the major determinant of systematic forecast failure.[2]

The admissible deductions on observing either the presence or absence of forecast failure are rather stark, particularly for methodologies in which forecasts are deemed the appropriate way to judge empirical models. Its absence is consistent with severe mis-specification, inaccurate data, and inconsistent parameter estimation, so hardly provides much comfort. Conversely, in the present setting of structural change in deterministic components, there may exist non-causal models (i.e., models none of whose 'explanatory' variables enter the DGP) that do not suffer forecast failure, and indeed may forecast absolutely more accurately on reasonable measures, than previously congruent, theory-based, efficiently-estimated econometric models: example are provided in Hendry (1997), Clements and Hendry (1999a), and section 5 below. Consequently, even relative forecasting success or failure is not a reliable basis for selecting between models – other than for forecasting purposes. Moreover, a model that suffers severe forecast failure may nonetheless have constant parameters on *ex post* re-estimation: apparent failure on forecasting need have no implications for the goodness of a model, or its theoretical underpinnings, as it may arise from incorrect data, that are later corrected (as in the concept of extended constancy in Hendry, 1996). To summarize, while forecast failure may reflect an inadequate model in-sample, it is neither necessary nor sufficient; whereas large deterministic shifts are sufficient to induce forecast failure.

Further, mis-specification of zero-mean stochastic components is unlikely to be a major source of forecast failure, but stochastic mis-specification (e.g., omitting key explanatory variables) could entail deterministic mis-specification, which might interact with deterministic breaks elsewhere in the economy, and thereby precipitate failure. Equally, the false inclusion of variables which experience equilibrium-mean shifts could have a marked impact on forecast failure: the model mean shifts although the data mean does not, thereby changing $\mathcal{E}_{\mathcal{T}}[\mathbf{y_{T+h}}]$, when $\mathsf{E}_{\mathsf{T+h}}[\mathbf{y_{T+h}}]$ is unaltered. Finally, forecast-origin mis-measurement can also be pernicious, as an incorrect starting level is 'carried forward' in dynamic models – hence most forecasting agencies carefully appraise the latest observations for consistency with other available information. Economic agents might find such detailed analysis expensive and unrewarding in terms of costs relative to resulting benefits.

3.3 Variance effects

Compared to the problems of the previous section, estimation uncertainty for the parameters of stochastic variables seems a secondary problem, as such errors add variance terms of $\mathsf{O}(1/T)$ for stationary components – and $\mathsf{O}(1/T^2)$ for non-stationary – for samples of size T. However, mis-estimation of coefficients of deterministic terms could be deleterious to forecast accuracy when $\mathsf{E}_{\mathsf{T+h}}[\mathbf{y_{T+h}}] - \mathcal{E}_{\mathcal{T}}\widehat{[\mathbf{y_{T+h}}]}$ is large by chance. For some model classes, such as VARs in levels, estimated intercepts have very large

[2]In non-stationary dynamic processes, where unconditional expectations may not be defined, this expression can be generalized to $\mathsf{E}_{\mathsf{T+h}}[\mathbf{x_{T+h}}|\mathbf{X_T}] - \mathcal{E}_{\mathcal{T}}[\mathbf{x_{T+h}}|\mathbf{X_T}]$ where $\mathbf{X_T}$ denotes observations up to and including the forecast origin.

standard errors, but this just reflects the inappropriateness of that parameterization. Finally, neither collinearity nor a lack of parsimony *per se* seem key culprits, although interacting with breaks occurring elsewhere in the economy could induce serious problems: see Clements and Hendry (1998a).

4 The detectability of structural breaks

The taxonomy in (8) also highlights the relative ease and difficulty with which different kinds of structural breaks can be detected. This is the topic of a separate analysis in Hendry (2000b), so is merely noted here.

Apparently-large shifts in both the VAR intercept, ϕ, and dynamic coefficient matrix, $\mathbf{\Pi}$, need not be detectable, so long as φ remains constant; whereas seemingly-small changes in φ can have a substantial, and easily detected, effect. This can be seen most easily from (7). Consider the set of parameter shifts $(\phi^*,\ \mathbf{\Pi}^*)$ such that:

$$(\mathbf{I_n} - \mathbf{\Pi}^*)^{-1}\,\phi^* = \varphi,$$

is constant. Then:

$$\mathbf{y_{T+h}} = \varphi + (\mathbf{\Pi}^*)^{\mathbf{h}}\,(\mathbf{y_T} - \varphi) + \sum_{\mathbf{i=0}}^{\mathbf{h-1}} (\mathbf{\Pi}^*)^{\mathbf{i}}\,\boldsymbol{\epsilon}_{\mathbf{T+h-i}} \tag{9}$$

where $\mathsf{E}_{\mathsf{T+h}}\,[\mathbf{y_T} - \varphi] = \mathbf{0}\ \forall h$ so the forecasts are unbiased, and the only detectable 'shift' is due to the variance change from $(\mathbf{\Pi}^*)^h\,(\mathbf{y_T} - \varphi)$ rather than $\mathbf{\Pi^h}\,(\mathbf{y_T} - \varphi)$ to be determined against the background noise from $\sum_{i=0}^{h-1} (\mathbf{\Pi}^*)^i\,\boldsymbol{\epsilon}_{\mathbf{T+h-i}}$. If the economy is close to equilibrium at the time of the break, so $\mathbf{y_T} \simeq \varphi$ then the outcomes will not reflect any change, so the forecasts will perforce be close to the actuals. Thus, the changes in terms of $(\phi^*,\ \mathbf{\Pi}^*)$ – which define the complete set of parameters of the VAR here – could be very large relative to the error variances, yet not be reflected in data variation. Hendry and Mizon (2000b) consider the bleak implications for impulse-response analyses, where shifts in dynamics that alter the signs of impulse responses may not be detectable, even when rigorous testing for parameter constancy is undertaken.

With $\mathsf{I}(1)$ data generated from a cointegrated VAR, the detectability of a change is still not well reflected by the original VAR parameterization. The vector equilibrium-correction (VEqCM) parameterization clarifies this outcome. In terms of the $n \times 1$ vector of $\mathsf{I}(1)$ time-series variables $\mathbf{x_t}$, write the first-order dynamic linear system as:

$$\mathbf{x_t} = \mathbf{\Upsilon}\mathbf{x_{t-1}} + \boldsymbol{\tau} + \mathbf{v_t} \tag{10}$$

where $\mathbf{v_t} \sim \mathsf{IN}_n\,[\mathbf{0}, \mathbf{\Omega_v}]$. In (10), the initial value $\mathbf{x_0}$ is fixed, $\mathbf{\Upsilon}$ is an $n \times n$ matrix of coefficients, and $\boldsymbol{\tau}$ an $n \times 1$ vector of intercepts. When $\mathbf{x_t} \sim \mathsf{I}(1)$ and the cointegrating rank is r, (10) can be reparameterized as:

$$\Delta\mathbf{x_t} = \mathbf{\Psi}\mathbf{x_{t-1}} + \boldsymbol{\tau} + \mathbf{v_t} \tag{11}$$

where $\mathbf{\Psi} = \mathbf{\Upsilon} - \mathbf{I_n} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $n \times r$ of rank $r < n$. Hence:

$$\Delta\mathbf{x_t} = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x_{t-1}} + \boldsymbol{\tau} + \mathbf{v_t}. \tag{12}$$

Let $\boldsymbol{\alpha}_\perp$, $\boldsymbol{\beta}_\perp$ be $n \times (n-r)$ matrices orthogonal to $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ respectively, so $\boldsymbol{\alpha}'_\perp\boldsymbol{\alpha} = \mathbf{0}$, $\boldsymbol{\beta}'_\perp\boldsymbol{\beta} = \mathbf{0}$, then (10) is not $\mathsf{I}(2)$ if $\boldsymbol{\alpha}'_\perp\boldsymbol{\beta}_\perp$ has rank $(n-r)$, which we assume, as well as sufficient restrictions to ensure uniqueness in $\boldsymbol{\alpha}_\perp$, $\boldsymbol{\beta}_\perp$, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$. As above, let the expectation of $\Delta\mathbf{x_t}$ be $\boldsymbol{\gamma}$ $(n \times 1)$, which defines the growth in the system, and let $\mathsf{E}\left[\boldsymbol{\beta}'\mathbf{x_t}\right] = \boldsymbol{\mu}$, which is $r \times 1$, then from (12), $\boldsymbol{\tau} = \boldsymbol{\gamma} - \boldsymbol{\alpha\mu}$ (see Johansen and Juselius, 1990), and $\boldsymbol{\beta}'\boldsymbol{\gamma} = \mathbf{0}$. Thus, the DGP becomes:

$$\Delta\mathbf{x_t} = \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x_{t-1}} - \boldsymbol{\mu}\right) + \mathbf{v_t}. \tag{13}$$

In relation to (8), the first r elements of $\mathbf{y_t}$ are the cointegrating combinations $\boldsymbol{\beta}'\mathbf{x_t}$ and the remaining $(n-r)$ are $\boldsymbol{\alpha}'_{\perp}\boldsymbol{\Delta}\mathbf{x_t}$.

The parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\mu}$ in (13) correspond directly to $\boldsymbol{\varphi}$: thus shifts in the equilibrium-mean correspond to shifts in either $\mathsf{E}_{\mathsf{T+h}}\left[\boldsymbol{\beta}'\mathbf{x_{T+h}}\right] = \boldsymbol{\mu}$ or $\mathsf{E}_{\mathsf{T+h}}\left[\Delta\mathbf{x_{T+h}}\right] = \boldsymbol{\gamma}$ or both. Such changes are readily detectable: the former because the model now drives forecasts towards an incorrect equilibrium, and hence in the opposite direction to the shift in the data mean, and the latter as the model has an incorrect growth rate enforced. Conversely, mean-zero shifts correspond to changes in the dynamics $\boldsymbol{\alpha}$, so are not easy to detect as the only manifest signal of the change is an 'above average' fluctuation in the data. The case where $\boldsymbol{\alpha}$ changes to zero is included – loss of cointegration need not induce forecast failure – whereas the introduction of new cointegration relationships between $\mathsf{I}(1)$ variables almost certainly induces a mean shift, and will be detectable. Surprisingly, shifts in the cointegration vectors $\boldsymbol{\beta}$ to $\boldsymbol{\beta}^*$ are not even easy to detect when entailed equilibrium-mean shifts are fully offset, so:

$$\mathsf{E}_{\mathsf{T+h}}\left[\left(\boldsymbol{\beta}^*\right)'\mathbf{x_{T+h}} - \boldsymbol{\mu}^*\right] = \mathbf{0},$$

although it may be that the usual empirical problem involves changes to $\boldsymbol{\varphi}$ when $\boldsymbol{\beta}$ alters.

Thus, the implicit variation-free assumptions about parameters are crucial in a world of structural shifts. The ease of detecting changes in equilibrium-means reflects their pernicious effects in forecasting; whereas the difficulty of perceiving other changes has consequential benefits of robustness in forecasting, but drawbacks of non-detection in modelling.

5 Robustness of forecasts to breaks

Two models with some robustness to deterministic shifts are a VAR in the differences of the variables (DV):

$$\Delta\mathbf{x_t} = \boldsymbol{\gamma} + \boldsymbol{\xi}_\mathbf{t}, \tag{14}$$

which is generally mis-specified by omitting any cointegrating vectors; and a DV in the differences of the variables (DDV), defined by:

$$\Delta^2\mathbf{x_t} = \boldsymbol{\zeta}_\mathbf{t} \;\text{ or }\; \boldsymbol{\Delta}\mathbf{x_t} = \boldsymbol{\Delta}\mathbf{x_{t-1}} + \boldsymbol{\zeta}_\mathbf{t}. \tag{15}$$

These models are convenient for the analytic calculations, but can be generalized in an obvious manner to allow for longer lag structures, trends etc. in empirical work. However, (15) will rarely provide a congruent data characterization due to 'over-differencing'.

To illustrate in the simplest setting, consider 1-step forecasting from T, where the in-sample parameters are constant and of negligible variance, estimated from accurate data for the correctly-specified first-order VEqCM (13) with no additional dynamics, immediately following an unanticipated deterministic shift at time $T-1$, so:

$$\begin{aligned}
\Delta\mathbf{x_{T-1}} &= \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x_{T-2}} - \boldsymbol{\mu}\right) + \mathbf{v_{T-1}} \\
\Delta\mathbf{x_{T}} &= \boldsymbol{\gamma}^* + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x_{T-1}} - \boldsymbol{\mu}^*\right) + \mathbf{v_{T}} \\
\Delta\mathbf{x_{T+1}} &= \boldsymbol{\gamma}^* + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x_{T}} - \boldsymbol{\mu}^*\right) + \mathbf{v_{T+1}}.
\end{aligned} \tag{16}$$

The VEqCM forecasts, however, are:

$$\Delta\widehat{\mathbf{x}}_{T+1|T} = \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{\mathbf{T}} - \boldsymbol{\mu}\right). \tag{17}$$

Then the taxonomy (8) simplifies to

$$\mathbf{e}_{\mathbf{T+1|T}} = \boldsymbol{\Delta}\mathbf{x}_{\mathbf{T+1}} - \boldsymbol{\Delta}\widehat{\mathbf{x}}_{\mathbf{T+1|T}} = \left(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}\right) - \boldsymbol{\alpha}\left(\boldsymbol{\mu}^* - \boldsymbol{\mu}\right) + \mathbf{v}_{\mathbf{T+1}}, \tag{18}$$

with a forecast-error bias of:

$$\mathsf{E}_{\mathsf{T}+1}\left[\mathbf{e}_{\mathbf{T+1|T}}\right] = \left(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}\right) - \boldsymbol{\alpha}\left(\boldsymbol{\mu}^* - \boldsymbol{\mu}\right),$$

and variance of $\mathsf{V}_{\mathsf{T}+1}\left[\mathbf{e}_{\mathbf{T+1|T}}\right] = \boldsymbol{\Omega}_{\mathbf{v}}$.

By way of contrast, consider using $\Delta\widetilde{\mathbf{x}}_{T+1|T} = \Delta\mathbf{x}_{\mathbf{T}}$ from (15). The corresponding taxonomy for (15) from (16) is:

$$\begin{aligned}\Delta\mathbf{x}_{\mathbf{T+1}} - \boldsymbol{\Delta}\widetilde{\mathbf{x}}_{\mathbf{T+1|T}} &= \boldsymbol{\gamma}^* + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{\mathbf{T}} - \boldsymbol{\mu}^*\right) + \mathbf{v}_{\mathbf{T+1}} - \boldsymbol{\Delta}\mathbf{x}_{\mathbf{T}} \\ &= \Delta\mathbf{v}_{\mathbf{T+1}} + \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{\Delta}\mathbf{x}_{\mathbf{T}},\end{aligned} \tag{19}$$

with a forecast-error bias of:

$$\mathsf{E}_{\mathsf{T}+1}\left[\Delta\mathbf{x}_{\mathbf{T+1}} - \boldsymbol{\Delta}\widetilde{\mathbf{x}}_{\mathbf{T+1|T}}\right] = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{\gamma}^*,$$

which is zero since the new mean in the cointegration relationships induced by the structural change has already been incorporated. However:

$$\mathsf{V}_{\mathsf{T}+1}\left[\Delta\mathbf{x}_{\mathbf{T+1}} - \boldsymbol{\Delta}\widetilde{\mathbf{x}}_{\mathbf{T+1|T}}\right] = 2\boldsymbol{\Omega}_{\mathbf{v}},$$

so there is a variance offset to the gain in bias, and that trade-off worsens as the forecast horizon increases. Nevertheless, until the VEqCM switches to having growth of $\boldsymbol{\gamma}^*$ and equilibrium mean of $\boldsymbol{\mu}^*$, (15) will outperform, at least in terms of bias, despite its non-causal basis. As Clements and Hendry (1999a) show, even though the forecast-error variances from (15) are larger and increase faster than those from the VEqCM, MSFEs can still favour the former.

It must be stressed that none of these models is robust to unanticipated deterministic shifts that occur after forecasts are announced – that requires magic, not science – but in real time, later forecasts will be made after any breaks, which is when the advantages of robustness appear. The VEqCM will persistently fail until it is revised to incorporate the relevant breaks, either by ‘intercept shift’ indicators, or inclusion of the forces that induced the shift. Moreover, ‘optimality’ results can no longer be established, as the relative performance of such models, and others, depends on the relative magnitudes and frequencies of deterministic shifts.

Equally, pooling forecasts from VEqCMs and DDVs may be advantageous in some periods, but cannot be useful in others, so any ranking is again event dependent. Intercept corrections based on recent forecast errors improve the robustness of VEqCMs by mimicking differencing (see Clements and Hendry, 1999a), but worsen forecast-error variances; and pre-testing does not uniformly help (see Clements and Hendry, 2001c).

To summarize the argument so far, deterministic shifts are a pernicious, so easily detected, source of forecast failure; their effects can be offset by ‘robust devices’, which thereby do not suffer forecast failure; these are implications of the generalized theory of forecasting; and are consistent with the outcomes empirical forecasting exercises. Thus, we turn to the additional implications of this framework for how agents living in such a world might form their expectations about the future.

6 Implications for 'rational expectations'

We now consider the six main implications for 'rational expectations' of forecast failure, as summarized in section 2.

6.1 Forecasts using causally-irrelevant variables

In the example of the previous section, $\Delta \mathbf{x}_{\mathbf{t-1}}$ does not enter the DGP (13), so is not a causally-relevant variable. Yet it can produce 'better' forecasts than the in-sample DGP, at least for some relevant states of nature. Thus, we have disproved by a counter example any theorem which claims that knowledge of the causally-relevant variables will deliver the 'best' forecasts. In a world with deterministic shifts, even forecasting from the existing DGP will not be useful unless all future breaks over the forecast horizon can be fully anticipated. Such an assumption is not a sound basis for any operational approach to economic policy or forecasting. This establishes the first implication noted in section 2.

When the variables under analysis are a subset of those in the DGP, there exists a well-defined local DGP (LDGP), derivable by reduction, with an innovation error representation (see e.g., Hendry, 1995, chapter 9). Bontemps and Mizon (2001) show that a model is congruent when it encompasses that LDGP. However, it cannot be proved that variables which are causally irrelevant in the DGP remain so in the LDGP, as they may proxy for variables which were eliminated by marginalization operations (see e.g., Hendry and Mizon, 1999). Similarly, causally-relevant variables become those that enter the LDGP with non-zero effects. Nevertheless, with that *caveat*, the result above holds, namely variables that do not enter the DGP or LDGP can deliver forecasts that out-perform on some measures models based on those that are relevant, as in section 5.

6.2 Winning forecasting competitions

Since forecast failure regularly occurs, we infer that deterministic shifts do so as well. Thus, models related to (15), namely ones that are highly adaptive, will perform relatively successfully in forecasting competitions, as seems to be the case empirically: see e.g., Makridakis *et al.* (1982), Fildes (1992), and Allen and Fildes (2001). Moreover, intercept corrections that 'set a model back on track' correspond to imposing an additional unit root over the forecast horizon, helping explain why such devices enhance forecast performance: see Turner (1990) and Clements and Hendry (1998a). Also, Clements and Hendry (1999b) demonstrate both analytically and empirically that in the face of deterministic shifts immediately prior to a forecast horizon, (14) and (15) will outperform (13), with the latter being best over very short horizons. This behaviour is precisely what was observed by Eitrheim, Husebø and Nymoen (1999) in their study of the forecasting performance of the Norges Bank macro-econometric model. Over the longest (12 quarter) evaluation horizon, the Bank's model performed well on a MSFE measure (albeit using known future values of all 'exogenous' variables), followed by a DV modelled to be congruent: the equivalent of the DDV did worst. But over a sequence of three 4-period divisions of the same evaluation data, which both reduces the adverse variance effect for the DDV and creates more forecasts after breaks, the DDV did best more often than any other method. This establishes the second implication.

6.3 Untenable informational requirements

'Rational expectations' requires agents not only to know all the relevant information, but also to know how every component enters the joint data density at each point in time. In an economy made non-stationary by unanticipated deterministic shifts—when many of the events, and their consequences, cannot be anticipated—such assumptions are untenable. These claims are clear from (8): the model error $\widehat{\epsilon}_{T+h|T}$ equals the 'rational expectations' error $\boldsymbol{\epsilon}_{\mathbf{T+h|T}}$ (i.e., term (*v*) in (8)) if and only if every other term is zero.[3] When $\widehat{\mathbf{y}}_{T+h|T} = \mathsf{E}_{\mathsf{T+h}}[\mathbf{y}_{\mathbf{T+h}}|\mathbf{Y}^{\mathbf{1}}_{\mathbf{T}}, \{\mathbf{Q}^{*}\}^{\mathbf{1}}_{\mathbf{T+h}}]$, where $\{\mathbf{Q}^{*}\}^{T+1}_{T+h}$ denotes the actual values of the deterministic factors over the forecast period (thereby incorporating any deterministic shifts), and $(\mathbf{Q}^{\mathbf{1}}_{\mathbf{T}}, \{\mathbf{Q}^{*}\}^{\mathbf{T+1}}_{\mathbf{T+h}}) = \{\mathbf{Q}^{*}\}^{\mathbf{1}}_{\mathbf{T+h}}$, then indeed (8) collapses to leave only $\boldsymbol{\epsilon}_{\mathbf{T+h|T}} = \sum_{\mathbf{i=0}}^{\mathbf{h-1}} (\boldsymbol{\Pi}^{*})^{\mathbf{i}}\, \boldsymbol{\epsilon}_{\mathbf{T+h-i}}$. However, the required implicit knowledge of recent and *future* deterministic shifts is infeasible. Moreover, once $\{\mathbf{Q}^{*}\}^{1}_{T+h}$ is replaced by $\mathbf{Q}^{\mathbf{1}}_{\mathbf{T+h}}$ and $\mathsf{E}_{\mathsf{T+h}}$ by E_{T} (or even worse, $\mathcal{E}_{\mathcal{T}}$), the previous section showed that the resulting forecasting device can be dominated by methods that involve no causally-relevant variables. Thus, an 'as if' argument cannot be used to sustain the claim of 'rational expectations' in modelling economic behaviour. This establishes the third implication.

6.4 Robust forecasts and robust decision rules

'Rational agents' would be wise not to use model-based expectations, when models – including their own – are manifestly mis-specified, and the variables involved are subject to deterministic shifts. Agents will learn that they cannot form expectations any better than the best forecasting models, where the latter are 'robust forecasting rules', so adopt those. But forecasting models such as DDV have the property that they do not usually vary with policy changes, although their forecasts do. Consider a simple model of the form:

$$\begin{aligned} \Delta x_t &= \gamma_0 + \theta \Delta z^e_{t+1} + v_{1,t} \\ \Delta z_t &= \gamma_1 D_{1,t} + \alpha \left(x_{t-1} - \beta z_{t-1} - \mu D_{2,t} \right) + v_{2,t} \end{aligned} \tag{20}$$

where the agent needs to form a forecast (expectation) of the next change in z_t (denoted Δz^e_{t+1}) to reach their decision, but z_t depends on intermittent unpredictable deterministic shifts, represented by the indicator variables D, which shift both the equilibrium mean μ and the growth rate γ_1. The 'rational expectation' of Δz_{t+1} can only be formed on assumptions about the $D_{i,t+1}$, and incorrect assumptions could generate systematically wrong forecasts, whereas a forecast like DDV is robust after any breaks. Here, the DDV entails $\Delta z^e_{t+1} = \Delta z_t$ so the first equation becomes a conditional model, and consequently does not vary with changes in the generating process for z_t. This establishes the fourth implication.

A possibility of interest is that the agent in question is the central government, and so makes the policy changes. In that case, a regime change will be known, so an econometric model which embodied the policy could outperform a DDV that did not (as the regime shift acts like a break after forecasting, to which the DDV is not robust: see Hendry and Mizon, 2000a).

6.5 Model-consistent expectations

In practice, the assertion by several non-nested macro-models to all embody 'rational expectations' involves logical contradictions: at most one could be correct, and at best the rest have 'model consistent'

[3]Except cancellation by fluke.

expectations. But since such models are bound to be mis-specified, 'model consistent' expectations have no formal basis. Indeed, they seem to embody the worst of all possible features: they are not robust to structural change, unlike robust rules, and most cannot coincide with how agents actually form expectations. Forecasting one period after a deterministic shift as in (17), imposing the restriction that expectational variables coincide with the biased forecasts from the model, would seem to compound the problem. The outcomes in the economy are the result of agents' actions, so when outcomes differ systematically from the model, the model must be incorrect: the forecast failure cannot be attributed to agents' expectations errors, even though the latter undoubtedly occur. This establishes the fifth implication.

Worse still, exploiting 'model-consistent expectations' can imply unrealistic 'trade-offs' – precisely what earlier 'Keynesian' models were criticized for – by claiming that agents correctly foresee the consequences of policy changes, and hence (e.g.) 'destabilizing policy rules' can lead agents to behave in such a way that the policy is instantly effective (so never destabilizes) – at least within the model, if not in reality.

6.6 Implications for the Lucas critique

Lucas (1976) criticized the use of estimated econometric models for policy analysis in the following quote:

"...Given that the structure of an econometric model consists of optimal decision rules for economic agents, and that optimal decision rules vary systematically with changes in the structure of series relevant to the decision maker, it follows that any change in policy will systematically alter the structure of econometric models..." Lucas [1976, p.41].

In the intervening period, the critique has been criticized in turn by many authors, including Gordon (1976), Sims (1982, 1986), and Hendry (1988), and most recently by Marcellino and Salmon (2000), who show the internally-inconsistent use of 'rationality' in Lucas's analysis. There is also scant evidence of its operation empirically (see, in particular, Ericsson and Irons, 1995). The present criticisms derive from a different source, namely the implications of observed forecast failure and its explanation by deterministic shifts.

There are several questionable assumptions essential for Lucas's assertion. First, that "optimal decision rules vary systematically with changes in the structure of series relevant to the decision maker" – unanticipated deterministic shifts are among the most drastic changes facing agents, yet the robust forecasting rules described above do not alter with them. Thus, an econometric model in which variables such as $\Delta \mathbf{x_t}$ entered, representing $\Delta \widetilde{\mathbf{x}}_{t+1|t}$, would not necessarily change even when forecast failure occurred ('gear changes' of the kind discussed by Flemming, 1976, shifting variables from $\mathsf{I}(1)$ to $\mathsf{I}(2)$ say, could induce changed forecasting rules). In fact, this is precisely the formulation discussed by Favero and Hendry (1992), albeit on completely different grounds.

Secondly, "that any change in policy will systematically alter the structure of econometric models" conflates regime shifts in policy, when the underlying rules are altered, with different policy-variable values within the same regime. Provided the models embody the policy variables of relevance, changes in their values will be correctly reflected, and changes in rules need not alter them.

The third, less obvious, assumption, is that agents can extract the "structure of series relevant to the(m)", and this seems untenable in precisely the type of processes consistent with intermittent forecast failure. Thus, the critique is far from being an explanation for forecast failure, and does not entail the implication that econometric models have no future in a world where policy rules change. This establishes the sixth implication.

An implicit assumption in Lucas (1976) is that any changes in econometric models will have noticeable empirical consequences. As we have seen, zero-mean changes are difficult to detect, and do not necessarily disrupt forecasts. Thus, the unimportant consequences of such changes, rather than their absence, could account for the lack of empirical evidence that the critique occurs: see Ericsson and Irons (1995). Nevertheless, for policy models, such undetected changes could be hazardous: the estimated parameters would appear to be constant, yet be mixtures across regimes, leading to inappropriate advice – indeed, estimated impulse responses could have the wrong sign (see Hendry and Mizon, 2000b). In a progressive research context (i.e., from the perspective of learning about the economy), such an outcome is unproblematic for econometric systems, since most policy changes involve deterministic shifts (as opposed to mean-preserving spreads), hence earlier incorrect inferences will be detected rapidly – although that is cold comfort to the policy maker, or the economic agents subjected to the wrong policies.

7 Selecting policy models by forecast accuracy

Implicit in the Lucas critique is the notion that forecast failure will result from policy changes when expectations are incorrectly modelled by the econometrician. We have shown that such an implication is false, as is the converse that observing forecast failure implies either policy changes or incorrectly-modelled expectations. Nevertheless, there seems to be an impression that before econometric models are used to guide economic policy they ought to be able to forecast well. This section shows that it is not in general sensible to select policy models by forecast accuracy alone.[4]

A statistical forecasting system is one having no economic-theory basis, in contrast to econometric models which are based on economic theory. Since the former systems rarely have implications for economic-policy analysis – and may not even entail links between target variables and policy instruments – being the 'best' available forecasting device is insufficient to ensure its value for policy analysis. Consequently, the main issue is the converse: does the existence of a dominating forecasting procedure invalidate the use of an econometric model for policy? Since forecast failure often results from factors unrelated to the policy change in question, an econometric model may continue to characterize the responses of the economy to a policy, despite its forecast inaccuracy. Further, when policy changes are implemented, forecasts from a statistical model may be improved by combining them with the predicted policy responses from an econometric model. Thus, forecasting models should remain distinct from policy models, as discussed in Hendry and Mizon (2000a).

The rationale for their analysis follows from the taxonomy of forecast errors in section 3 which recorded that deterministic shifts were the primary source of systematic forecast failure in econometric models. Devices like intercept corrections can robustify forecasting models against breaks which have occurred prior to forecasting (see e.g., Clements and Hendry, 1996, and Hendry and Clements, 2000). While such 'tricks' may mitigate forecast failure, the policy-analysis implications of the resulting models are neither more nor less useful than those of the failed counterparts – although the policy advice might be better when based on viable forecasts. Conversely, as noted in section 6, post-forecasting policy changes will induce breaks in models that did not embody the relevant policy links, whereas

[4]This section draws heavily on Hendry and Mizon (2000a) and is only included for completeness.

econometric systems need not experience that policy-regime shift. Consequently, when both structural breaks and regime shifts occur, neither class of model alone is adequate, which suggests that they should be combined: Hendry and Mizon (2000a) propose, and empirically illustrate, one approach to doing so.

8 Conclusion

Forecast failure is primarily due to unanticipated large shifts in deterministic factors. Consequently, it becomes impossible to prove the most basic theorem that forecasting based on causal variables will dominate that using non-causal, unless the model coincides with the generating mechanism after the break. Of course, such a result does not prove that causal models are inadvisable, merely that they cannot be shown to dominate. In such a non-stationary world, 'rational expectations' are epistemologically unsound, since they require agents to know all the relevant information, and how every item enters the joint data density at every point in time – an impossible requirement in macroeconomics when some of the events involved, and many of their consequences, cannot be anticipated. Methods which are relatively immune to unanticipated structural breaks then do best in forecasting, and need not involve causal variables. 'Sensible' agents will adopt such 'robust forecasting rules', which do not need to vary with regime changes (although the resulting forecasts obviously change).

Because 'deterministic shifts' are pernicious for causally-based forecasts, they are easily detected, although changes to other parameters are not. Thus, forecast failure is not due to changed reaction parameters *per se*, although it might result from induced deterministic shifts. Since impulse-response analyses are sensitive to the signs of parameters whose changes are hard to detect, they are not reliable. Conversely, while econometric models might forecast poorly when deterministic shifts occur, that does not impugn their value for policy analyses: and fixing forecast errors by devices such as intercept corrections cannot improve the policy-analysis properties of the models. The problems in detecting shifts in mean-zero parameters are as hard for econometric models as for other approaches, but since policy experiments generally involve mean shifts, regime changes will reveal previously-hidden mis-specifications, transforming forecast failure from a problem into a learning device.

Thus, the intermittent 'break down' of econometric models actually reveals the unsustainability of 'rational expectations' in empirical modelling, rather than problems with the econometric systems it was intended to criticize. Moreover, forecast failure is unlikely to be due to the 'Lucas critique', and is consistent with the increasing empirical evidence of its irrelevance in practice. In realistic settings, conditional expectations based on *past* information may be dominated in forecasting by other forecasting rules, and hence are not actually rational. Conversely, 'fully rational expectations', by requiring knowledge of unanticipated breaks, have impossible information requirements. Thus, it is essential to move towards more reasonable expectations models in macroeconomics, particularly formulations that reflect the instrumental nature of expectations in agents' decisions – see e.g., Feige and Pearce (1976).

References

Allen, P. G., and Fildes, R. A. (2001). Econometric forecasting strategies and techniques. In Armstrong, J. S. (ed.), *Principles of Forecasting*, pp. 303–362. Boston: Kluwer Academic Publishers.

Barrell, R. (2001). Forecasting the world economy. In Hendry, D. F., and Ericsson, N. R. (eds.), *Understanding Economic Forecasts*, pp. 151–172. Cambridge, Mass.: MIT Press.

Bontemps, C., and Mizon, G. E. (2001). Congruence and encompassing. In Stigum, B. P. (ed.), *Econometrics and the Philosophy of Economics*. Cambridge, Mass.: MIT Press. forthcoming.

Brunner, K., and Meltzer, A. (eds.)(1976). *The Phillips Curve and Labor Markets*, Vol. 1 of *Carnegie-Rochester Conferences on Public Policy*. Amsterdam: North-Holland Publishing Company.

Calzolari, G. (1981). A note on the variance of ex post forecasts in econometric models. *Econometrica*, **49**, 1593–1596.

Clements, M. P., and Hendry, D. F. (1993). On the limitations of comparing mean squared forecast errors. *Journal of Forecasting*, **12**, 617–637. With discussion.

Clements, M. P., and Hendry, D. F. (1996). Intercept corrections and structural change. *Journal of Applied Econometrics*, **11**, 475–494.

Clements, M. P., and Hendry, D. F. (1998a). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.

Clements, M. P., and Hendry, D. F. (1998b). Some methodological implications of forecast failure. mimeo, Institute of Economics and Statistics, University of Oxford.

Clements, M. P., and Hendry, D. F. (1999a). *Forecasting Non-stationary Economic Time Series*. Cambridge, Mass.: MIT Press.

Clements, M. P., and Hendry, D. F. (1999b). On winning forecasting competitions in economics. *Spanish Economic Review*, **1**, 123–160.

Clements, M. P., and Hendry, D. F. (2001a). Explaining the results of the M3 forecasting competition. *International Journal of Forecasting*, **Forthcoming**.

Clements, M. P., and Hendry, D. F. (2001b). An historical perspective on forecast errors. *National Institute Economic Review*, **177**, 100–112.

Clements, M. P., and Hendry, D. F. (2001c). Structural change and economic forecasting. Unpublished typescript, Economics Department, University of Oxford.

Eitrheim, Ø., Husebø, T. A., and Nymoen, R. (1999). Equilibrium-correction versus differencing in macroeconometric forecasting. *Economic Modelling*, **16**, 515–544.

Ericsson, N. R., and Irons, J. S. (1995). The Lucas critique in practice: Theory without measurement. In Hoover, K. D. (ed.), *Macroeconometrics: Developments, Tensions and Prospects*. Dordrecht: Kluwer Academic Press.

Favero, C., and Hendry, D. F. (1992). Testing the Lucas critique: A review. *Econometric Reviews*, **11**, 265–306.

Feige, E. L., and Pearce, D. K. (1976). Economically rational expectations. *Journal of Political Economy*, **84**, 499–522.

Fildes, R. (1992). The evaluation of extrapolative forecasting methods. *International Journal of Forecasting*, **8**, 81–98.

Fildes, R., and Ord, K. (2001). Forecasting competitions – their role in improving forecasting practice and research. In Clements, M. P., and Hendry, D. F. (eds.), *A Companion to Economic Forecasting*. Oxford: Blackwells.

Fildes, R. A., and Makridakis, S. (1995). The impact of empirical accuracy studies on time series analysis and forecasting. *International Statistical Review*, **63**, 289–308.

Flemming, J. S. (1976). *Inflation*. Oxford: Oxford University Press.

Gordon, R. J. (1976). Can econometric policy evaluations be salvaged?. in Brunner, and Meltzer (1976) *op. cit*.

Hendry, D. F. (1988). The encompassing implications of feedback versus feedforward mechanisms in econometrics. *Oxford Economic Papers*, **40**, 132–149. Reprinted in Ericsson, N. R. and Irons, J. S. (eds.) *Testing Exogeneity*, Oxford: Oxford University Press, 1994.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.

Hendry, D. F. (1996). On the constancy of time-series econometric equations. *Economic and Social Review*, **27**, 401–422.

Hendry, D. F. (1997). The econometrics of macro-economic forecasting. *Economic Journal*, **107**, 1330–1357. Reprinted in T.C. Mills (ed.), Economic Forecasting. Edward Elgar, 1999.

Hendry, D. F. (2000a). A general forecast-error taxonomy. Mimeo, Nuffield College, Oxford.

Hendry, D. F. (2000b). On detectable and non-detectable structural change. *Structural Change and Economic Dynamics*, **11**, 45–65.

Hendry, D. F., and Clements, M. P. (2000). Economic forecasting in the face of structural breaks. In Holly, S., and Weale, M. (eds.), *Econometric Modelling: Techniques and Applications*, pp. 3–37. Cambridge: Cambridge University Press.

Hendry, D. F., and Doornik, J. A. (1997). The implications for econometric modelling of forecast failure. *Scottish Journal of Political Economy*, **44**, 437–461. Special Issue.

Hendry, D. F., and Mizon, G. E. (1999). The pervasiveness of Granger causality in econometrics. In Engle, R. F., and White, H. (eds.), *Cointegration, Causality and Forecasting*. Oxford: Oxford University Press.

Hendry, D. F., and Mizon, G. E. (2000a). On selecting policy analysis models by forecast accuracy. In Atkinson, A. B., Glennerster, H., and Stern, N. (eds.), *Putting Economics to Work: Volume in Honour of Michio Morishima*, pp. 71–113. London School of Economics: STICERD.

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

Johansen, S., and Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration – With application to the demand for money. *Oxford Bulletin of Economics and Statistics*, **52**, 169–210.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. in Brunner, and Meltzer (1976), pp. 19–46.

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

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

Marcellino, M., and Salmon, M. (2000). Robust decision theory and the Lucas critique. unpublished paper, City University Business School, London.

Pesaran, M. H. (ed.)(1987). *The Limits of Rational Expectations*. Oxford: Basil Blackwell.

Sims, C. A. (1982). Policy analysis with econometric models. *Brookings Papers on Economic Activity*,

1, 107–164.

Sims, C. A. (1986). Are forecasting models useful for policy analysis?. *Federal Reserve Bank of Minneapolis Quarterly Review*, **Winter**, 2–16.

Turner, D. S. (1990). The role of judgement in macroeconomic forecasting. *Journal of Forecasting*, **9**, 315–345.

Wallis, K. F. (1986). Forecasting with an econometric model: The ‘ragged edge’ problem. *Journal of Forecasting*, **5**, 1–14.