

Forecast Failure, Expectations Formation, and the Lucas Critique

David F. Hendry*
Nuffield College, Oxford.

March 13, 2000

Abstract

Since forecast failure is due to unanticipated large shifts in deterministic factors, ‘sensible’ agents should adopt ‘robust forecasting rules’. Unless the model coincides with the generating mechanism, one cannot even prove that causal variables will dominate non-causal in forecasting. In such a non-stationary world, ‘rational expectations’ do not have an epistemologically-sound basis: agents cannot know how all relevant information enters the joint data density at every point in time. Thus, although econometric models ‘break down’ intermittently when deterministic shifts occur, that is not due to the Lucas critique and need not impugn their value for policy analyses.

1 Introduction

There are important implications for modelling expectations formation in economics of 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. To summarize, in such a non-stationary world, the first implication is that one can disprove the basic theorem that forecasting based on causal variables will dominate that using non-causal, unless the model coincides with the generating mechanism – an impossible requirement in macroeconomics: see Hendry (1997). Since ‘rational expectations’ rely on causal claims – being the correct conditional expectations – they do not in fact have a sound theoretical basis. The second implication is that the methods which do best in forecasting competitions are those which are relatively immune to unanticipated structural breaks: see Clements and Hendry (1999b). Robust methods are generally different from ‘rational expectations’, which are not robust in the face of unanticipated deterministic shifts. 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 ‘robust forecasting rules’. Fourthly, should they do so, such robust forecasts have the property that the rules do not usually vary with policy changes (although the resulting forecasts do), rendering the Lucas (1976) critique otiose. Fifthly, these criticisms apply *a fortiori* to ‘model consistent’ expectations, which seem to embody the worst of all possible worlds: 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.

Finally, because ‘deterministic shifts’ are pernicious for forecasting based on ‘causal variables’, they are easy to detect. Unfortunately, changes to other parameters transpire to be difficult to detect in many

*Financial support from the UK Economic and Social Research Council under grant R000237500 is gratefully acknowledged.

circumstances: see Hendry (2000b). Thus, forecast failure derives from deterministic shifts and not from changed reaction parameters, quite unlike the implications of the Lucas (1976) critique. Yet the latter are precisely the parameters of relevance to economic policy. Thus, the final implication is that analyses based on impulse responses are not reliable, since they are sensitive to the signs of parameters whose changes are hard to detect. Conversely, although econometric models might forecast poorly when deterministic shifts occur, that need not impugn their value for policy analyses: see Hendry and Mizon (1999). Such models face the same problems in detecting shifts in mean-zero parameters as other approaches, but since policy experiments generally involve mean shifts, they will reveal previously-hidden mis-specifications. Surprisingly, therefore, forecast failure should be viewed as a learning device, not a problem. The conclusion 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). The remainder of the paper establishes these implications.

The structure of the paper is as follows. Section 2 discusses the sources of forecast failure, leading to a taxonomy of forecast errors in section 3. Section 4 reviews the detectability of deterministic shifts, and the difficulty of detecting other structural breaks, applied in section 5 to the problems confronting impulse-response analyses. Section 6 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. Section 7 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 8 notes 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 9 concludes.

2 Sources of forecast failure

Forecast failure is a significant deterioration in forecast performance relative to the anticipated outcome, usually based on the historical performance of a model. Consequently, forecast failure would 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: 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 is central to the implications for models of expectations formation. 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 (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 system,
- (4) mis-estimating its parameters,
- (5) using inaccurate observations,
- (6) generated by an integrated-cointegrated process,
- (7) subject to intermittent structural breaks.

Such a framework closely mimics the empirical setting, and 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. 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, but delivers the same implications as the simpler exemplar used here: see Hendry (2000a).

3 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 $\mathbf{E}_t[\mathbf{x}_t|\mathbf{X}_{t-1}^1, \mathbf{q}_t]$ (where that exists), conditional on information $\mathbf{X}_{t-1}^1 = (\mathbf{x}_1 \dots \mathbf{x}_{t-1})$ available at the time, when \mathbf{q}_t denotes the relevant deterministic factors (such as intercepts, trends, and indicators), and $\mathbf{Q}_t^1 = (\mathbf{q}_1 \dots \mathbf{q}_t)$. Because the underlying densities may be changing, 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 $h = 1, \dots, H$, from a forecast origin at T . Forecast accuracy is to be judged by a criterion function $\mathbf{C}_e(\mathbf{e}_{T+1|T} \dots \mathbf{e}_{T+H|T})$, which we take to depend only on the forecast errors $\mathbf{e}_{T+h|T} = \mathbf{x}_{T+h} - \widehat{\mathbf{x}}_{T+h|T}$, where ‘smaller’ values of $\mathbf{C}_e(\cdot)$ are preferable.

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

$$\mathbf{y}_{T+h} = \boldsymbol{\phi} + \mathbf{\Pi} \mathbf{y}_{T+h-1} + \boldsymbol{\epsilon}_{T+h}, \quad (1)$$

with $\boldsymbol{\epsilon}_t \sim \text{IN}_n[\mathbf{0}, \boldsymbol{\Omega}_\epsilon]$, using the forecasting model:

$$\widehat{\mathbf{y}}_{T+h|T+h-1} = \widehat{\boldsymbol{\phi}} + \widehat{\mathbf{\Pi}} \mathbf{y}_{T+h-1}. \quad (2)$$

In (1), when \mathbf{x}_t is $I(1)$, the \mathbf{y}_t comprise r elements $\boldsymbol{\beta}'\mathbf{x}_t$ (when there are $r < n$ cointegrating vectors), and $n - r$ elements $\Delta\mathbf{x}_t$, so are $I(0)$ 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 $\widehat{\epsilon}_{T+h|T} = \mathbf{y}_{T+h} - \widehat{\mathbf{y}}_{T+h|T}$, and $\mathbf{C}_{\widehat{\epsilon}}(\cdot)$ will comprise only the mean and variance of $\{\widehat{\epsilon}_{T+h|T}\}$. It is important to note 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:

$$\mathbb{E}[\mathbf{y}_t] = (\mathbf{I}_n - \mathbf{\Pi})^{-1} \boldsymbol{\phi} = \boldsymbol{\varphi} \quad (3)$$

and hence, in deviation form, the DGP is:

$$\mathbf{y}_t - \boldsymbol{\varphi} = \mathbf{\Pi} (\mathbf{y}_{t-1} - \boldsymbol{\varphi}) + \boldsymbol{\epsilon}_t.$$

The h -step ahead forecasts at time T for $h = 1, \dots, H$ from (2) can be written as:

$$\widehat{\mathbf{y}}_{T+h} - \widehat{\boldsymbol{\varphi}} = \widehat{\mathbf{\Pi}} (\widehat{\mathbf{y}}_{T+h-1} - \widehat{\boldsymbol{\varphi}}) = \widehat{\mathbf{\Pi}}^h (\widehat{\mathbf{y}}_T - \widehat{\boldsymbol{\varphi}}), \quad (4)$$

where $\widehat{\boldsymbol{\varphi}} = (\mathbf{I}_n - \widehat{\mathbf{\Pi}})^{-1} \widehat{\boldsymbol{\phi}}$, and ‘ \wedge ’s denote estimates on parameters, and forecasts on random variables. Although \mathbf{y}_T is unknown in practice, we assume $\mathbb{E}[\widehat{\mathbf{y}}_T] = \boldsymbol{\varphi}$ here, so that on average no systematic bias results from data inaccuracy – in practice, however, the forecast origin might be less well measured than earlier data (see e.g., Wallis, 1993) 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 $I(0)$. Then, from $T + 1$ onwards, the data are generated by:

$$\begin{aligned} \mathbf{y}_{T+h} &= \boldsymbol{\phi}^* + \mathbf{\Pi}^* \mathbf{y}_{T+h-1} + \boldsymbol{\epsilon}_{T+h} \\ &= \boldsymbol{\varphi}^* + \mathbf{\Pi}^* (\mathbf{y}_{T+h-1} - \boldsymbol{\varphi}^*) + \boldsymbol{\epsilon}_{T+h} \\ &= \boldsymbol{\varphi}^* + (\mathbf{\Pi}^*)^h (\mathbf{y}_T - \boldsymbol{\varphi}^*) + \sum_{i=0}^{h-1} \boldsymbol{\epsilon}_{T+h-i}, \end{aligned}$$

so both the slope and the intercept alter. Let $\widehat{\epsilon}_{T+h|T} = \mathbf{y}_{T+h} - \widehat{\mathbf{y}}_{T+h|T}$, then the forecast-error taxonomy in (5) decomposes the various sources into eight interpretable, and potentially freely-varying, components:¹

VAR forecast-error taxonomy

$$\begin{aligned} \widehat{\epsilon}_{T+h|T} \simeq & \left(\mathbf{I}_n - (\mathbf{\Pi}^*)^h \right) (\boldsymbol{\varphi}^* - \boldsymbol{\varphi}) && (ia) \text{ equilibrium-mean change} \\ & + \left((\mathbf{\Pi}^*)^h - \mathbf{\Pi}^h \right) (\mathbf{y}_T - \boldsymbol{\varphi}) && (ib) \text{ dynamic change} \\ & + \left(\mathbf{I}_n - \mathbf{\Pi}_p^h \right) (\boldsymbol{\varphi} - \boldsymbol{\varphi}_p) && (iia) \text{ equilibrium-mean mis-specification} \\ & + \left(\mathbf{\Pi}^h - \mathbf{\Pi}_p^h \right) (\mathbf{y}_T - \boldsymbol{\varphi}) && (iib) \text{ dynamic mis-specification} \\ & - \left(\mathbf{I}_n - \mathbf{\Pi}_p^h \right) (\widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}) && (iiia) \text{ equilibrium-mean estimation} \\ & - \mathbf{F}_h \left(\widehat{\mathbf{\Pi}} - \mathbf{\Pi} \right)^\nu && (iiib) \text{ dynamic estimation} \\ & - \left(\mathbf{\Pi}_p^h + \mathbf{C}_h \right) (\mathbf{y}_T - \widehat{\mathbf{y}}_T) && (iv) \text{ forecast-origin uncertainty} \\ & + \sum_{i=0}^{h-1} (\mathbf{\Pi}^*)^i \boldsymbol{\epsilon}_{T+h-i} && (v) \text{ error accumulation.} \end{aligned} \quad (5)$$

¹The 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.

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 (5), 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 $\widehat{\boldsymbol{\Pi}}$ for $\boldsymbol{\Pi}$ will not induce biased forecasts (see e.g., Clements and Hendry, 1998a), and the $\boldsymbol{\epsilon}_{T+h}$ have zero means by construction. Consequently, the primary sources of systematic forecast failure are (ia), (iia), (iiaa), and (iv). However, on *ex post* evaluation, (iv) will be removed, and in congruent models with freely-estimated intercepts, (iia) and (iiaa) will be zero on average. That leaves (ia) as the primary source. 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 (5) in section 7.

3.1 Mean effects

The key effect derives from changes to the ‘equilibrium mean’ (not necessarily the intercept in a model, as seen in (5)), namely the unconditional expectation $\mathbf{E}[\mathbf{y}_t]$ (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, $\mathbf{E}_{T+h}[\mathbf{y}_{T+h}] - \mathcal{E}_T[\mathbf{y}_{T+h}]$ 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. 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 6 below. Consequently, even relative success or failure is not a reliable basis for selecting between models – other than for forecasting purposes. Conversely, 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 (see the concept of extended constancy in Hendry, 1996).

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 $\mathcal{E}_T[\mathbf{y}_{T+h}]$, when $\mathbf{E}_{T+h}[\mathbf{y}_{T+h}]$ 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.

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

3.2 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 $O(1/T)$ for stationary components – and $O(1/T^2)$ for non-stationary – for samples of size T . However, mis-estimation of coefficients of deterministic terms could be deleterious to forecast accuracy when $E_{T+h}[\mathbf{y}_{T+h}] - \widehat{\mathcal{E}}_T[\mathbf{y}_{T+h}]$ is large by chance. 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 (5) 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.

With $I(1)$ data generated from a cointegrated VAR, the detectability of a change is not well reflected by the original VAR parameterization. Apparently-large shifts in both the VAR intercept, ϕ , and dynamic coefficient matrix, Π , need not be detectable, so long as φ remains constant; whereas seemingly-small changes can have a substantial, and easily detected, effect. The vector equilibrium-correction (VEqCM) parameterization clarifies this outcome: shifts in the equilibrium-mean, φ , are readily detectable, whereas mean-zero shifts are not. 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 perniciousness in forecasting; whereas the difficulty of detecting other changes has consequential benefits of robustness in forecasting but drawbacks of non-detection in modelling.

5 Impulse-response analyses

Such findings are potentially serious for ‘impulse-response’ analyses of economic policy. Since the changes in VAR intercepts and dynamic coefficient matrices may not be detected even when tested for, but the recorded estimates are a weighted average across the different regimes, the resulting impulse responses need not represent the policy outcomes that will in fact occur.

A Monte Carlo simulation illustrates the problem, using an unrestricted $I(0)$ VAR:

$$\begin{aligned} y_{1,t} &= \phi_1 + \pi_{11}y_{1,t-1} + \pi_{12}y_{2,t-1} + \epsilon_{1,t} \\ y_{2,t} &= \phi_2 + \pi_{21}y_{1,t-1} + \pi_{22}y_{2,t-1} + \epsilon_{2,t} \end{aligned} \tag{6}$$

where $\epsilon_{i,t} \sim \text{IN}[0, \sigma_{ii}]$, with $E[\epsilon_{1,t}\epsilon_{2,s}] = 0 \forall t, s$. We consider breaks in the π_{ij} with constant unconditional expectations of zero. The full-sample size is $T = 120$, with a single break at $t = 0.5T$: 1000 replications were used. The unrestricted VAR with intercept and one lag is estimated, and then tested for breaks. The statistics reported below for testing parameter constancy are recursively-computed, system variants of Chow (1960) tests, as described in Doornik and Hendry (1997, Ch. 14). In a univariate setting, the tests are calculated from the residual sums of squares as the sample size is increased, and so are related to the CUSUMSQ statistic in Brown, Durbin and Evans (1975). The vector F-tests use the F-approximation in Rao (1952), as described in Rao (1973, sect. 8c.5) or Anderson (1984, sect. 8.5.4). The critical values are those for a *known* break point, which deliver the highest-possible power for the test used. Test rejection frequencies at both 0.05 and 0.01 nominal sizes are reported graphically

(standard errors of about 0.007 and 0.003 respectively). The graphs serve to illustrate the outcomes visually, showing that rejection frequencies are everywhere low in most cases, but confirming that the highest power is immediately before the break.

5.1 Test size

As fig. 1 reveals, the null rejection frequencies in the $I(0)$ baseline data are reassuring: with 1000 replications, the approximate 95% confidence intervals are (0.036, 0.064) and (0.004, 0.016) for 5% and 1% nominal, and these are shown on the graphs as dotted and dashed lines respectively. The actual test sizes are close to their nominal levels.

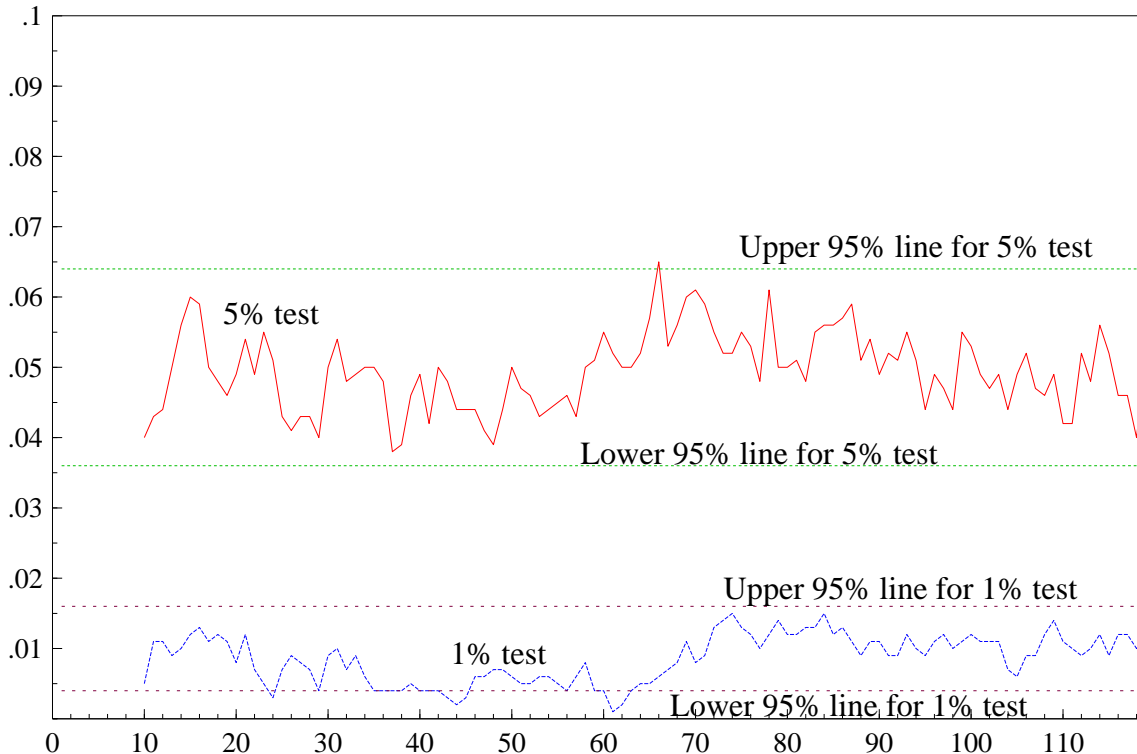


Figure 1 Constancy-test rejection frequencies for the $I(0)$ null.

5.2 $I(0)$ dynamic shift

The detectability of a shift in dynamics is low when the DGP is an $I(0)$ VAR with $\phi = \mathbf{0}$. We consider a large parameter shift, from:

$$\mathbf{\Pi} = \begin{pmatrix} 0.50 & -0.20 \\ -0.20 & -0.25 \end{pmatrix} \text{ to } \mathbf{\Pi}^* = \begin{pmatrix} 0.50 & 0.20 \\ 0.20 & 0.25 \end{pmatrix}. \quad (7)$$

Almost every coefficient changes sign, with only the first element left constant to highlight the changes in the other impulses. This break delivers the rejection-frequency graph in fig. 2. The highest power is less than 25%, even though the change constitutes a major structural break for this ‘model economy’. As shown in Hendry (2000b), shifts in ϕ that are small relative to the σ_{ii} are easily detected.

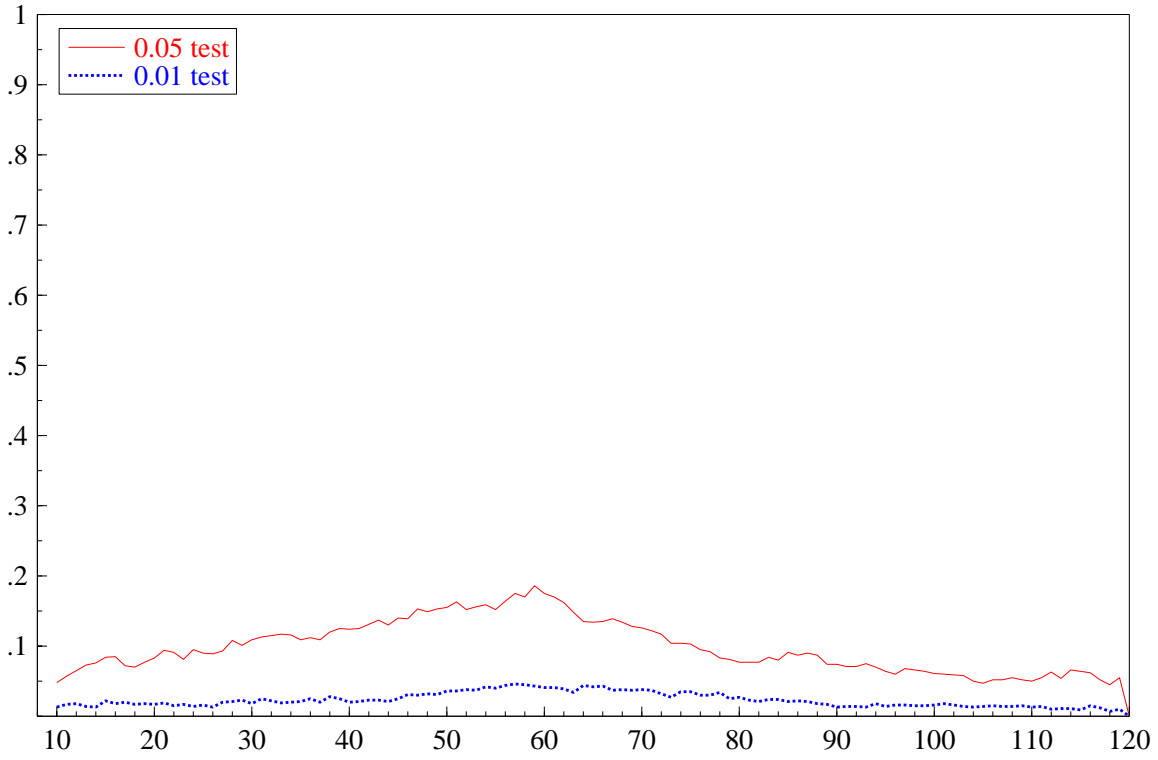


Figure 2 Constancy-test rejection frequencies for the $I(0)$ structural break.

5.3 Impulse responses

Finally, we record the impulse responses from the pre- and post- break models, and the model fitted across the regime shifts in fig. 3. The contrast is marked: despite the near undetectability of the break, the signs of most of the impulses have altered, and those obtained from the fitted model sometimes reflect one regime, and sometimes the other. Overall, mis-leading policy advice would follow.

6 Robustness to breaks

Returning to the $n \times 1$ vector of $I(1)$ time-series variables \mathbf{x}_t , and maintaining the concrete example of a VEqCM, write the first-order dynamic linear system as:

$$\mathbf{x}_t = \mathbf{\Upsilon}\mathbf{x}_{t-1} + \boldsymbol{\tau} + \mathbf{v}_t \quad (8)$$

where $\mathbf{v}_t \sim \text{IN}_n[\mathbf{0}, \boldsymbol{\Omega}_v]$. In (8), 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 $\mathbf{x}_t \sim I(1)$ and the cointegrating rank is r , (8) can be reparameterized as:

$$\Delta\mathbf{x}_t = \boldsymbol{\Psi}\mathbf{x}_{t-1} + \boldsymbol{\tau} + \mathbf{v}_t \quad (9)$$

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

$$\Delta\mathbf{x}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \boldsymbol{\tau} + \mathbf{v}_t. \quad (10)$$

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 (8) is not $I(2)$ if $\boldsymbol{\alpha}'_\perp\boldsymbol{\beta}_\perp$ has rank $(n - r)$, which we assume, as well as sufficient restrictions to ensure

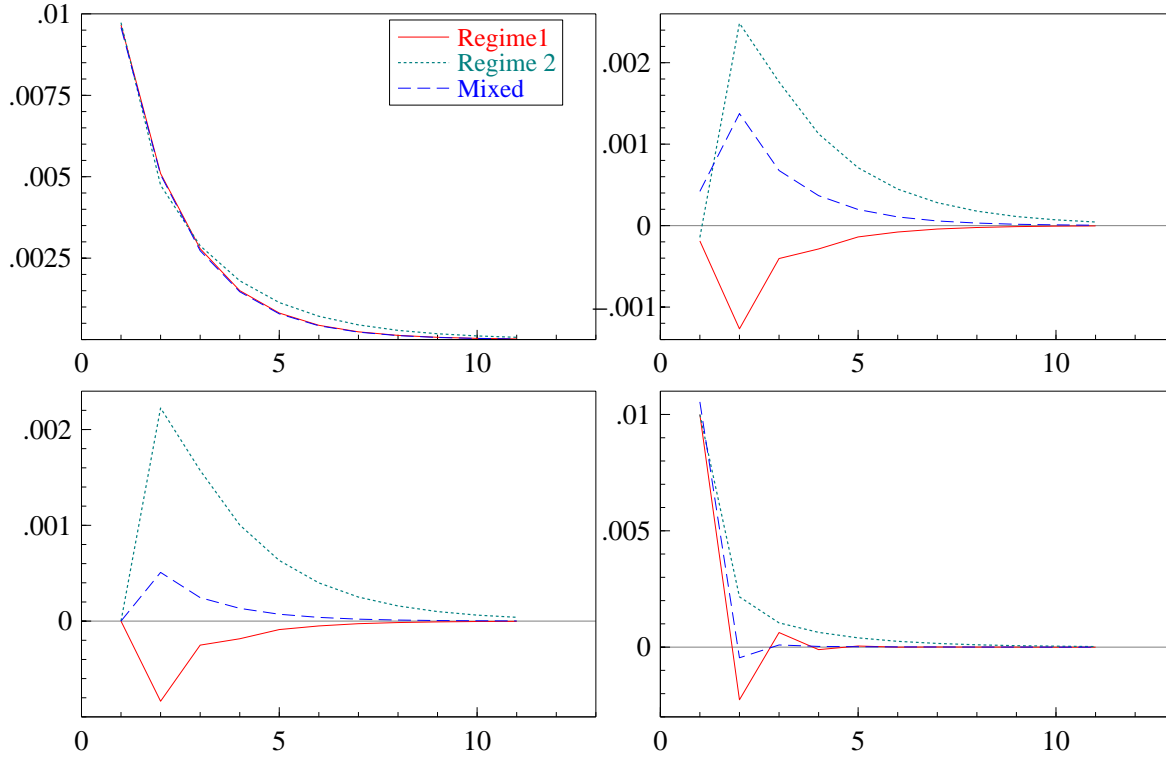


Figure 3 Impulse response comparisons in an I(0) VAR.

uniqueness in α_{\perp} , β_{\perp} , α , and β . Let the expectation of $\Delta \mathbf{x}_t$ be γ ($n \times 1$), which defines the growth in the system, and let $E[\beta' \mathbf{x}_t] = \mu$, which is $r \times 1$, then from (10), $\tau = \gamma - \alpha \mu$ (see Johansen and Juselius, 1990), and the DGP becomes:

$$\Delta \mathbf{x}_t = \gamma + \alpha (\beta' \mathbf{x}_{t-1} - \mu) + \mathbf{v}_t. \quad (11)$$

In relation to (5), the first r \mathbf{y}_t are the cointegrating combinations $\beta' \mathbf{x}_t$ and the remaining $(n - r)$ are $\alpha'_{\perp} \Delta \mathbf{x}_t$. The parameters γ and μ correspond directly to φ : thus shifts in these induce forecast failure.

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

$$\Delta \mathbf{x}_t = \gamma + \xi_t, \quad (12)$$

which is generally mis-specified by omitting any cointegrating vectors (unless $\alpha = \mathbf{0}$); and a DV in the differences of the variables (DDV), defined by:

$$\Delta^2 \mathbf{x}_t = \zeta_t \text{ or } \Delta \mathbf{x}_t = \Delta \mathbf{x}_{t-1} + \zeta_t. \quad (13)$$

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, (13) will rarely provide a congruent data characterization due to 'over-differencing'.

To illustrate in the simplest setting, consider 1-step forecasting where the in-sample parameters are constant and of negligible variance, estimated from accurate data for the correctly-specified first-order VEqCM (11) with no additional dynamics, immediately following a deterministic shift at time $T - 1$, so:

$$\Delta \mathbf{x}_{T-1} = \gamma + \alpha (\beta' \mathbf{x}_{T-2} - \mu) + \mathbf{v}_{T-1}$$

$$\begin{aligned}\Delta \mathbf{x}_T &= \gamma^* + \alpha (\beta' \mathbf{x}_{T-1} - \boldsymbol{\mu}^*) + \mathbf{v}_T \\ \Delta \mathbf{x}_{T+1} &= \gamma^* + \alpha (\beta' \mathbf{x}_T - \boldsymbol{\mu}^*) + \mathbf{v}_{T+1},\end{aligned}\tag{14}$$

whereas:

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

Then the taxonomy simplifies to

$$\mathbf{e}_{T+1|T} = \Delta \mathbf{x}_{T+1} - \Delta \widehat{\mathbf{x}}_{T+1|T} = (\gamma^* - \gamma) - \alpha (\boldsymbol{\mu}^* - \boldsymbol{\mu}) + \mathbf{v}_{T+1},\tag{15}$$

with a forecast-error bias of:

$$E_{T+1} [\mathbf{e}_{T+1|T}] = (\gamma^* - \gamma) - \alpha (\boldsymbol{\mu}^* - \boldsymbol{\mu}).$$

By way of contrast, consider using $\Delta \widetilde{\mathbf{x}}_{T+1|T} = \Delta \mathbf{x}_T$ from (13). Here, $\Delta \mathbf{x}_{t-1}$ does not enter the DGP (11), so is not a causally-relevant variable. The corresponding taxonomy for (13) from (14) is:

$$\begin{aligned}\Delta \mathbf{x}_{T+1} - \Delta \widetilde{\mathbf{x}}_{T+1|T} &= \gamma^* + \alpha (\beta' \mathbf{x}_T - \boldsymbol{\mu}^*) + \mathbf{v}_{T+1} - \Delta \mathbf{x}_T \\ &= \Delta \mathbf{v}_{T+1} + \alpha \beta' \Delta \mathbf{x}_T,\end{aligned}\tag{16}$$

with a forecast-error bias of:

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

which is zero unless the structural change also induced a new mean in the cointegration relationships. Until the VEqCM switches to having growth of γ^* and equilibrium mean of $\boldsymbol{\mu}^*$, (13) will outperform, despite its non-causal basis.

Clements and Hendry (1999b) demonstrate both analytically and empirically that in the face of deterministic shifts immediately prior to a forecast horizon, (12) and (13) will outperform (11), with the latter being best over very short horizons. This behaviour is precisely what was observed by Eitheim, Husebø and Nymoen (1997) in their study of the forecasting performance of the Norges Bank model. Over the longest (12 quarter) evaluation horizon, the Bank's model performed well (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 creates more forecasts after breaks, the DDV did best more often than any other method.

Since forecast failure regularly occurs, we infer that deterministic shifts must do so as well. Thus, models related to (13), namely ones that are highly adaptive, will perform successfully in forecasting competitions, as seems to be the case empirically: see e.g., Makridakis (1982), Fildes (1992), and Allen and Fildes (2000). 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).

7 Implications for 'rational expectations'

In an economy made non-stationary by unanticipated deterministic shifts, '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, when many of the events, and their consequences, cannot be unanticipated. 'Sensible' agents will learn that they cannot form expectations any better than the best forecasting models, and so adopt 'robust forecasting rules' – which have the property that they

do not usually vary with policy changes, although their forecasts do. ‘Rational agents’ would be wise not to use model-based expectations, when models – including their own – are manifestly mis-specified.

These claims are clear from (5): the model error $\widehat{\epsilon}_{T+h|T}$ equals the ‘rational expectations’ error $\epsilon_{T+h|T}$ (i.e., term (v) in (5)) if and only if every other term is zero.³ When $\widehat{y}_{T+h|T} = E_{T+h}[y_{T+h} | \mathbf{Y}_T^1, \{\mathbf{Q}^*\}_{T+h}^1]$, where the actual values of the deterministic factors over the forecast period (thereby incorporating any deterministic shifts), are denoted by $\{\mathbf{Q}^*\}_{T+h}^{T+1}$ and $[\mathbf{Q}_T^1, \{\mathbf{Q}^*\}_{T+h}^{T+1}] = \{\mathbf{Q}^*\}_{T+h}^1$, then indeed (5) collapses to leave only $\epsilon_{T+h|T}$. However, the required implicit knowledge of recent and *future* deterministic shifts is untenable. Moreover, once $\{\mathbf{Q}^*\}_{T+h}^1$ is replaced by \mathbf{Q}_{T+h}^1 and E_{T+h} by E_T (or even \mathcal{E}_T), the previous section showed that the resulting forecasting device can be dominated by methods that involve no causally-relevant variables. $\sum_{i=0}^{h-1} (\mathbf{\Pi}^*)^i \epsilon_{T+h-i}$

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

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.

7.1 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].

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 Δx_t entered representing $\Delta \widetilde{x}_{t|t-1}$ would not necessarily change even when forecast failure occurred (‘gear changes’ of the kind discussed by Flemming, 1976, shifting variables from $l(1)$ to $l(2)$ say, could induce changed forecasting rules). In fact, this is precisely the formulation discussed by Favero and Hendry (1992), 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, and 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

³Except cancellation by fluke.

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.

The final implicit assumption 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 – as shown above, estimated impulse responses could have the wrong sign. In a progressive research context (i.e., from the perspective of learning), this 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.

8 Selecting models by forecast accuracy

A statistical forecasting system is one having no economic-theory basis, in contrast to econometric models for which economic theory is the hallmark.⁴ 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 explained by Hendry and Mizon (1999).

The rationale for this analysis follows from the taxonomy of forecast errors in section 2 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, 1999). 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. Conversely, post-forecasting policy changes will induce breaks in models that did not embody the relevant policy links, whereas 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: this suggests that they should be combined, and Hendry and Mizon (1999) propose, and empirically illustrate, one approach to doing so.

8.1 Congruent modelling

As a usable knowledge base, theory-related, congruent, encompassing econometric models remain undominated by matching the data in all measurable respects (see, e.g., Hendry, 1995). For empirical understanding, such models seem likely to remain an integral component of any progressive research

⁴This section draws heavily on Hendry and Mizon (1999) and is only included for completeness.

strategy. Nevertheless, even the ‘best-available model’ will be caught out when forecasting by a sudden, unanticipated outbreak of (say) a major war or other crisis for which no effect was included in the forecast. However, if empirical models which are congruent within sample remain subject to a non-negligible probability of failing out of sample, then a critic might doubt their worth. Our defence of the program of attempting to discover such models rests on the fact that empirical research is part of a progressive strategy, in which knowledge gradually accumulates. This includes knowledge about general causes of structural changes, such that later models incorporate measures accounting for previous events, and hence are more robust (e.g., to wars, changes in credit rationing, financial innovations, etc.). For example, the dummies for purchase-tax changes in Davidson, Hendry, Srba and Yeo (1978) that at the time ‘mopped up’ forecast failure, later successfully predicted the effects of introducing VAT, as well as the consequences of its doubling in 1979; and the First World-War shift in money demand in Ericsson, Hendry and Prestwich (1998) matched that needed for the Second World War.

9 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. 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 unanticipated. 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*, but 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. Their problems in detecting shifts in mean-zero parameters are as hard as for other approaches, but since policy experiments generally involve mean shifts, they 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.

References

- Allen, P. G., and Fildes, R. A. (2000). Econometric forecasting strategies and techniques. Forthcoming, J. S. Armstrong, (ed.) Principles of Forecasting, Kluwer Academic Press, USA.

- Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd edn. New York: John Wiley & Sons.
- Brown, R. L., Durbin, J., and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time (with discussion). *Journal of the Royal Statistical Society B*, **37**, 149–192.
- Calzolari, G. (1981). A note on the variance of ex post forecasts in econometric models. *Econometrica*, **49**, 1593–1596.
- Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, **28**, 591–605.
- 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: The Marshall Lectures on Economic Forecasting*. 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: The Zeuthen Lectures on Economic Forecasting*. Cambridge, Mass.: MIT Press.
- Clements, M. P., and Hendry, D. F. (1999b). On winning forecasting competitions in economics. *Spanish Economic Review*, **1**, 123–160.
- Davidson, J. E. H., Hendry, D. F., Srba, F., and Yeo, J. S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal*, **88**, 661–692. Reprinted in Hendry, D. F. (1993), *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.
- Doornik, J. A., and Hendry, D. F. (1997). *Modelling Dynamic Systems using PcFiml 9 for Windows*. London: Timberlake Consultants Press.
- Eitrheim, Ø., Husebø, T. A., and Nymoen, R. (1997). Error correction versus differencing in macro econometric forecasting. mimeo, Department of Economics, University of Oslo, Norway.
- Ericsson, N. R., Hendry, D. F., and Prestwich, K. M. (1998). The demand for broad money in the United Kingdom, 1878–1993. *Scandinavian Journal of Economics*, **100**, 289–324.
- 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.
- Fildes, R. (1992). The evaluation of extrapolative forecasting methods. *International Journal of Forecasting*, **8**, 81–98.
- Flemming, J. S. (1976). *Inflation*. Oxford: Oxford University Press.
- 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*, **Anniversary**, 00–00. Forthcoming.
- Hendry, D. F., and Clements, M. P. (1999). Economic forecasting in the face of structural breaks. In Holly, S., and Weale, M. (eds.), *Econometric Modelling: Techniques and Applications*. Cambridge: Cambridge University Press. Forthcoming.
- Hendry, D. F., and Mizon, G. E. (1999). On selecting policy analysis models by forecast accuracy. Festschrift in Honour of Michio Morishima, STICERD, London School of Economics.
- 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, K., and Meltzer, A. (eds.), *The Phillips Curve and Labor Markets*, Vol. 1 of *Carnegie-Rochester Conferences on Public Policy*, pp. 19–46. Amsterdam: North-Holland Publishing Company.
- Makridakis, S. e. a. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, **1**, 111–153.
- Rao, C. R. (1952). *Advanced Statistical Methods in Biometric Research*. New York: John Wiley.
- Rao, C. R. (1973). *Linear Statistical Inference and its Applications*, 2nd edn. New York: John Wiley & Sons.
- Turner, D. S. (1990). The role of judgement in macroeconomic forecasting. *Journal of Forecasting*, **9**, 315–345.
- Wallis, K. F. (1993). Comparing macroeconometric models: A review article. *Economica*, **60**, 225–237.