

# The Implications for Econometric Modelling of Forecast Failure

David F. Hendry and Jurgen A. Doornik  
Nuffield College,  
Oxford \*

January 31, 2002

## Abstract

To reconcile forecast failure with building congruent empirical models, we analyze the sources of mis-prediction. This reveals that ex ante forecast failure is purely a function of forecast-period events, not determinable from in-sample information. The primary causes are unmodelled shifts in deterministic factors, rather than model mis-specification, collinearity, or a lack of parsimony. We examine the effects of deterministic breaks on equilibrium-correction mechanisms, and consider the role of causal variables. Throughout, Monte Carlo simulation and empirical models illustrate the analysis, and support a progressive research strategy based on learning from past failures.

## 1 Introduction

It is a pleasure to contribute to the centenary volume of the *Scottish Journal of Political Economy*, and we wish it a further century of continued success. We record with interest the coincidence of timing in that the *Scandinavian Journal of Economics* is also celebrating its 100<sup>th</sup> birthday this year: Northern Europe in the late 1890s clearly witnessed a large percentage increase in Journal production!

Over much of that century, economic forecasting has suffered from bouts of serious predictive failure, including the ABC curves based on Persons (1924) apparently missing the onset of the Great Depression in 1929; the post-war mis-prediction of consumers' expenditure (see e.g. Duesenberry, 1949); the Korean War jump in inflation; the general demise of Keynesian-style macro-econometric systems during the oil-crises of the 1970s (see e.g., Cooper and Nelson, 1975); and the underpredictions of consumers' expenditure in the UK in the mid 1980s followed by overpredictions in the early 1990s (see Cook, 1995, for a review of UK Treasury forecasts of consumers' expenditure). Indeed, despite advances in economic theory, in econometric techniques and methodology, in computing, and especially in accumulated empirical knowledge, forecast failures and parameter non-constancies still abound. The empirical demise has been matched by professional scepticism, especially the early claims in (say) Morgenstern (1928) that the notion of economic forecasting is flawed in principle (but see Marget, 1929, reprinted with discussion in Hendry and Morgan, 1995).

This paper presents an analysis which seeks to reconcile the strategy of building congruent empirical models with the repeated rebuffs that forecasters have faced historically. Recent research on economic forecasting has highlighted some salient issues that help account for the mis-match, suggesting alternative approaches to forecasting, and explaining why models can fit better than they forecast, even in the absence of overfitting by 'data mining'. We distinguish ex-ante forecast failure (incorrect statements

---

\*Financial support from the UK Economic and Social Research Council under grant R000234954 is gratefully acknowledged. We thank Mike Clements, Rob Engle, Neil Ericsson, Grayham Mizon and Bent Nielsen for helpful discussions about the material herein.

about as yet unobserved events) from ex-post predictive failure (failure on a parameter constancy test against the observed outcomes). The former could be due to many causes, including data errors or false assumptions about non-modelled variables which are corrected later, so the model remains constant when updated. Ex-ante forecast failure transpires to be purely a function of forecast-period events, a primary cause of which is shown to be unmodelled shifts in equilibrium means (see §2). The latter occurs when a model is non-constant on the whole available information set, and is a well-established notion. Forecast failure may be predictable ex ante: for example, correct anticipation of the effects of the 1984 legal change permitting interest payments on retail sight deposits would entail mis-forecasting by a transactions-demand model which excluded that variable. However, because of ambiguities in the concept of a constant model, predictive failure is not in general determinable from the in-sample information used in modelling (see Hendry, 1996b, and §6 below).

The use of equilibrium-correction mechanisms based on multivariate cointegration is also examined. The analysis reveals the potential problems inherent in modelling for forecasting, based on the result in Hendry and Mizon (1996) that ‘either the dynamics must ensure correction, or the target implicit in the econometric model must move when the regime alters’. A fundamental difficulty is that causal variables need not dominate non-causal factors for forecasting when models are mis-specified in a world subject to breaks in deterministic factors (constants, trends etc.): see Hendry (1996a). This phenomenon allows for the possibility of designing forecasting devices that are robust to important classes of structural breaks, driving a wedge between forecasting and modelling, and in particular precluding claims that the ‘best forecasting’ model must be selected for policy. The paper draws together these apparently disparate developments to check the validity of modelling congruent relations despite repeated forecast failures.

The structure of the paper is as follows. Section 2 discusses the sources of predictive failure in congruent models of cointegrated processes, and, using analysis and Monte Carlo simulation, highlights the central role therein of deterministic shifts. The roles of model mis-specification, collinearity, and parsimony in forecast failure are considered in §3, and shown not to account for major systematic mis-forecasting per se; however, when interacting with breaks in regressors, all three can induce serious forecast errors. Section 4 discusses the role of causal variables. Then §5 considers the impacts of mean shifts on equilibrium-correction mechanisms, and the obverse issue of why non-congruent (e.g., differenced data) models need not fail when forecasting, drawing on results about robustifying forecasts against breaks. Finally, §6 considers the possibility of obtaining constancy despite forecast failure, and investigates whether ex ante non-constancy is a fatal weakness. Section 7 concludes.

## 2 Predictive failure in congruent models

In this section, the data generation process (DGP) is a cointegrated, vector autoregression (VAR) which coincides with the model in-sample. We introduce our general notation, and consider the non-integrated (denoted  $I(0)$ ) representation. The only possible causes of forecast failure in such a setting are post-estimation sample breaks. By analysis, and a Monte Carlo illustration in §2.2, intercept shifts are shown to be the dominant cause of predictive failure, both ex ante and ex post. As discussed in Hendry (1996b), UK M1 is a potential example following the introduction of interest-bearing retail sight deposits: these sharply lowered the opportunity costs of holding M1, shifting the long-run equilibrium mean, which — when not modelled appropriately — induced substantial forecast errors.

## 2.1 The consequences of mean shifts

We first show that intercept shifts (either from direct changes in deterministic factors, or induced by changes in dynamic adjustments) are a dominant cause of predictive failure in linear dynamic systems. Other breaks can cause forecast failures, but their ease of detection depends on the magnitudes of the long-run means (of the stationary combinations) relative to their error standard deviations. Intuitively, when the long-run mean is non-zero, breaks shift the location of the data, inducing a short-run ‘trend’ to the new equilibrium mean which is more easily detected than a variance change around the origin.

Consider a first-order VAR in the  $n$  variables  $\mathbf{x}_t$ :

$$\mathbf{x}_t = \boldsymbol{\kappa} + \boldsymbol{\Upsilon}\mathbf{x}_{t-1} + \boldsymbol{\nu}_t, \quad (1)$$

where  $\boldsymbol{\nu}_t \sim \text{IN}_n[\mathbf{0}, \boldsymbol{\Omega}_\nu]$  (an independent, normal error with expectation  $\mathbf{E}[\boldsymbol{\nu}_t] = \mathbf{0}$  and variance matrix  $\mathbf{V}[\boldsymbol{\nu}_t] = \boldsymbol{\Omega}_\nu$ ). The data are integrated of order unity ( $I(1)$ ) as:

$$\boldsymbol{\Upsilon} = \mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}', \quad (2)$$

where  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  are  $n \times r$  matrices of rank  $r$ . No roots of  $|\mathbf{I}_n - \boldsymbol{\Upsilon}L| = 0$  lie inside the unit circle (where  $L^s x_t = x_{t-s}$ ), with  $\text{rank}(\boldsymbol{\alpha}'_\perp \boldsymbol{\Upsilon} \boldsymbol{\beta}_\perp) = (n - r)$ , where  $\boldsymbol{\alpha}_\perp$  and  $\boldsymbol{\beta}_\perp$  are  $n \times (n - r)$  matrices of rank  $(n - r)$  and  $\boldsymbol{\alpha}'\boldsymbol{\alpha}_\perp = \boldsymbol{\beta}'\boldsymbol{\beta}_\perp = \mathbf{0}$ . In (1):

$$\boldsymbol{\kappa} = \boldsymbol{\gamma} - \boldsymbol{\alpha}\boldsymbol{\mu}, \quad (3)$$

where  $\boldsymbol{\mu}$  is  $r \times 1$  and  $\boldsymbol{\beta}'\boldsymbol{\gamma} = \mathbf{0}$ . Substituting (2) and (3) into (1) yields the vector equilibrium-correction model (VEqCM):

$$\Delta\mathbf{x}_t - \boldsymbol{\gamma} = \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\nu}_t \quad (4)$$

where  $\Delta\mathbf{x}_t$  and  $\boldsymbol{\beta}'\mathbf{x}_t$  are  $I(0)$ , with  $\mathbf{E}[\Delta\mathbf{x}_t] = \boldsymbol{\gamma}$  and equilibrium  $\mathbf{E}[\boldsymbol{\beta}'\mathbf{x}_t] = \boldsymbol{\mu}$ .

To highlight the impacts of parameter changes on (4), we draw on the forecast-error taxonomy proposed by Clements and Hendry (1994). Write the system in  $I(0)$  space in terms of the  $n$  variables  $\mathbf{y}_t$ , the first  $r$  of which are  $\boldsymbol{\beta}'\mathbf{x}_t$  and the remaining  $n - r$  are the relevant elements of  $\Delta\mathbf{x}_t$  (e.g.,  $\boldsymbol{\alpha}'_\perp \Delta\mathbf{x}_t$ ). Then:

$$\mathbf{y}_t = \boldsymbol{\phi} + \boldsymbol{\Pi}\mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t \quad \text{with} \quad \boldsymbol{\epsilon}_t \sim \text{IN}_n[\mathbf{0}, \boldsymbol{\Omega}_\epsilon], \quad (5)$$

so the unconditional mean of  $\mathbf{y}_t$  over  $t = 1, \dots, T$  is:

$$\mathbf{E}[\mathbf{y}_t] = (\mathbf{I}_n - \boldsymbol{\Pi})^{-1} \boldsymbol{\phi} = \boldsymbol{\varphi} \quad (6)$$

and so, from (5):

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

Assume that, at the end of time period  $T$ ,  $(\boldsymbol{\phi} : \boldsymbol{\Pi})$  changes to  $(\boldsymbol{\phi}^* : \boldsymbol{\Pi}^*)$ , so from  $T + 1$  onwards, the data are generated by:

$$\mathbf{y}_{T+1} = \boldsymbol{\phi}^* + \boldsymbol{\Pi}^*\mathbf{y}_T + \boldsymbol{\epsilon}_{T+1}.$$

Letting  $\boldsymbol{\phi}^* = (\mathbf{I}_n - \boldsymbol{\Pi}^*) \boldsymbol{\varphi}^*$  (we assume  $\boldsymbol{\Pi}^*$  still has all its eigenvalues less than unity in absolute value, and the number of cointegrating vectors remains the same):

$$\mathbf{y}_{T+1} - \boldsymbol{\varphi}^* = \boldsymbol{\Pi}^*(\mathbf{y}_T - \boldsymbol{\varphi}^*) + \boldsymbol{\epsilon}_{T+1}. \quad (7)$$

Let  $\hat{\cdot}$  denote estimates, then the residual at  $T + 1$ ,  $\hat{\boldsymbol{\epsilon}}_{T+1} = \mathbf{y}_{T+1} - \hat{\mathbf{y}}_{T+1}$  is:

$$\hat{\boldsymbol{\epsilon}}_{T+1} = \boldsymbol{\phi}^* + \boldsymbol{\Pi}^*\mathbf{y}_T + \boldsymbol{\epsilon}_{T+1} - \hat{\boldsymbol{\phi}} - \hat{\boldsymbol{\Pi}}\mathbf{y}_T. \quad (8)$$

Treating finite-sample biases as negligible (see Hendry, 1996a), but allowing for (possibly) inconsistent parameter estimates,  $E[\hat{\Pi}] = \Pi_p$  with  $E[\hat{\phi}] = \phi_p = (\mathbf{I}_n - \Pi_p)\varphi$  (since almost all estimation methods match data means in-sample), then conditional on  $\mathbf{y}_T$ :

$$\begin{aligned} E[\hat{\epsilon}_{T+1} | \mathbf{y}_T] &\simeq \phi^* + \Pi^* \mathbf{y}_T + E[\epsilon_{T+1}] - E[\hat{\phi}] - E[\hat{\Pi}] \mathbf{y}_T \\ &= \phi^* - \phi_p + (\Pi^* - \Pi_p) \mathbf{y}_T. \end{aligned} \quad (9)$$

From (5), taking the unconditional mean of each term in (9):

$$\begin{aligned} E[\hat{\epsilon}_{T+1}] &= (\mathbf{I}_n - \Pi^*) \varphi^* - (\mathbf{I}_n - \Pi_p) \varphi + (\Pi^* - \Pi_p) \varphi \\ &\simeq (\mathbf{I}_n - \Pi^*) (\varphi^* - \varphi). \end{aligned} \quad (10)$$

There are four implications from (10). First, the outcome is zero in mean-zero processes. Secondly, the forecast bias is also zero when the long-run mean is constant: if  $\varphi^* = \varphi$ , changes in the dynamics ( $\Pi^*$ ) vanish on average. Indeed, rather startlingly from (8), breaks in dynamics and parameter-estimation uncertainty would have *no effect* if  $\mathbf{y}_T = \mathbf{0}$ . Equally surprisingly,  $E[\hat{\epsilon}_{T+1}]$  is zero when shifts in  $\phi^*$  offset those in  $\Pi^*$  to leave  $\varphi$  unaffected so both dynamics and intercepts shift. This result is the basis for our claim that changes in deterministic factors matter most for predictive failure. Finally from (10), a given differential  $(\varphi^* - \varphi)$ , deriving from changes in  $\Pi$  only or  $\phi$  only, will induce a larger (smaller) forecast bias as  $(\mathbf{I}_n - \Pi^*)$  becomes ‘larger’ (‘nearer zero’). In all cases, the detectability of failure is dependent on the magnitudes of the shifts relative to the error standard deviations. We now illustrate the implications of these analytical forecast results by simulating some tests of parameter constancy.

## 2.2 An $l(0)$ Monte Carlo illustration

The first Monte Carlo simulation considers the following five contrasting experiments, implemented in Ox by PcNaive for Windows (see Doornik, 1996, and Doornik and Hendry, 1997).

- (A) A constant-parameter equation;
- (B) a break in the dynamics of a zero-mean  $l(0)$  process;
- (C) a break in the dynamics of a non-zero mean  $l(0)$  process;
- (D) a break in the intercept so the long-run mean is equal to that induced in (C);
- (E) the break in (C) offset by the break in (D) to leave the long-run mean unaltered.

The baseline DGP is the following  $l(0)$  VAR for a full-sample size of  $T = 200$ :

$$\begin{aligned} y_{1,t} &= \alpha_1 + 0.3y_{2,t} + \beta_1 y_{1,t-1} - 0.15y_{2,t-1} + \epsilon_{1,t} \\ y_{2,t} &= \alpha_2 - 0.5y_{1,t-1} + 0.75y_{2,t-1} + \epsilon_{2,t}. \end{aligned} \quad (11)$$

The  $\{\epsilon_{i,t}\}$  are mutually and serially independent, normal, mean zero, with variances of 0.1, and 0.5 respectively. All breaks occurred at  $T = 100$ , followed by a return to the original parameter value at  $T = 150$ : the regime over  $T = 100$  to  $T = 150$  is denoted 2, the remaining regime is 1. The model is the first equation in (11), estimated consistently by OLS.<sup>1</sup> We used  $M = 1000$  replications, with common random numbers across all experiments.

<sup>1</sup>Depending on the test statistic selected for analysis, non-centrality parameters could be calculated in this scalar  $l(0)$  process to reveal the factors determining test power (see e.g., Hendry, 1995a).

The experimental design parameter values are:

Case	A1	A2	B1	B2	C1	C2	D1	D2	E1	E2
$\beta_1$	0.85	0.85	0.85	0.70	0.85	0.70	0.85	0.85	0.85	0.70
$\alpha_1$	1.5	1.5	0.0	0.0	1.5	1.5	1.5	0.375	0.375	1.5
$\alpha_2$	5.0	5.0	0.0	0.0	5.0	5.0	5.0	5.0	5.0	5.0

Thus, the intercept change in the first equation is about four error standard deviations. Since the process in (11) can be written as:

$$\begin{pmatrix} 1 - \beta_1 L & -0.3 + 0.15L \\ 0.5L & 1 - 0.75L \end{pmatrix} \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}, \quad (12)$$

the long-run solution is:

$$\begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} = \frac{1}{(1.3 - \beta_1)} \begin{pmatrix} \alpha_1 + 0.6\alpha_2 \\ 4(1 - \beta_1)\alpha_2 - 2\alpha_1 \end{pmatrix}.$$

This generates the following outcomes for the equilibria corresponding to  $\varphi^*$ :

Case	A1	A2	B1	B2	C1	C2	D1	D2	E1	E2
$\varphi_1$	10	10	0	0	10	7.5	10	7.5	7.5	7.5
$\varphi_2$	0	0	0	0	0	5	0	5	5	5

Figure 1 shows the rejection frequencies from  $T = 50$  to 200 of the break-point Chow (1960) tests for parameter constancy.<sup>2</sup> Case (A) shows that the actual size of the test is close to the nominal 5%. As predicted, the power in the zero-mean process (B) is nearly equal to the size, so the change in the dynamics is essentially undetectable. However, it is easily detected in case (C). Next, (D) does indeed generate nearly equal ‘powers’ to (C), with the latter somewhat higher since the dynamics were reduced. Further, in (C) and (D), on reversion to the state prior to the first break, the tests become biased: this matches our experience with macroeconomic data.

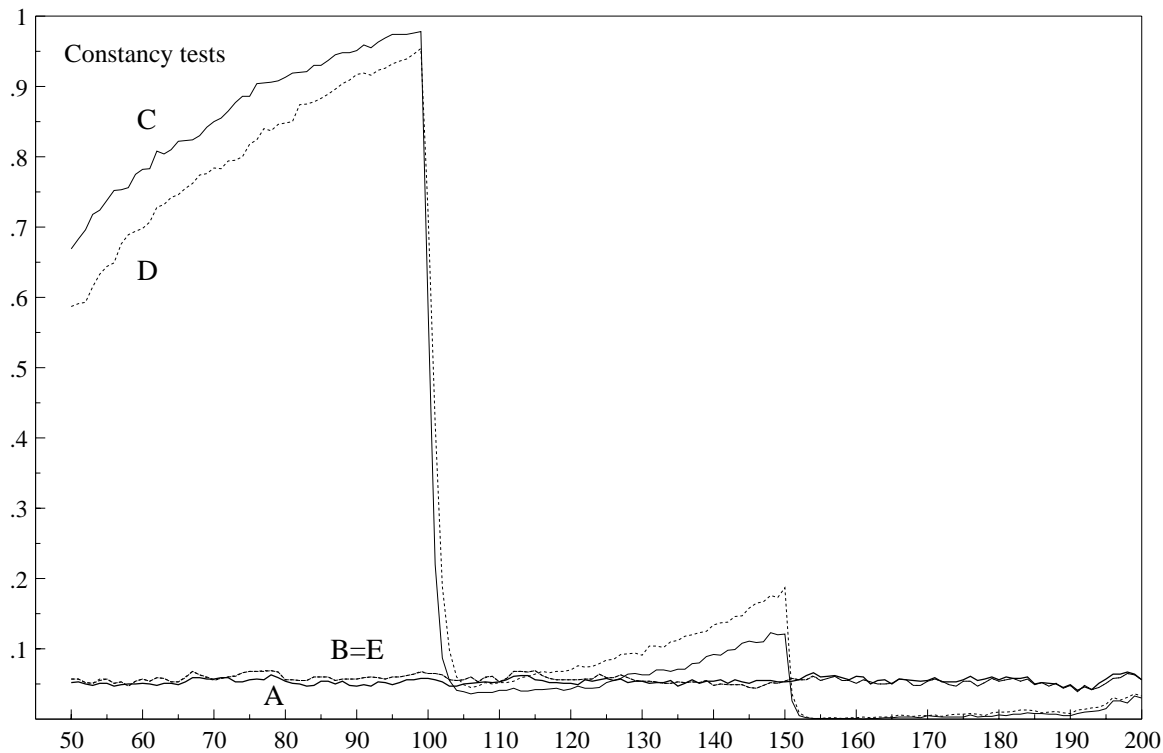
The key implications of the above analysis are that case (B) would not be detectable on such tests, and that (E) would be similar to (B). To understand why the outcome in (E) is in fact identical to (B), consider the reduced forms of the first equations of their DGPs, expressed in deviations about means, denoted  $\tilde{y}$ , where in (B),  $\tilde{y}_{1,t} = y_{1,t}$  with  $\tilde{y}_{2,t} = y_{2,t}$ , whereas in (E),  $\tilde{y}_{1,t} = y_{1,t} - 7.5$  and  $\tilde{y}_{2,t} = y_{2,t} - 5$ :

$$\begin{aligned} \text{(B1)} &: \tilde{y}_{1,t} = 0.70\tilde{y}_{1,t-1} + 0.075\tilde{y}_{2,t-1} + (\epsilon_{1,t} + 0.3\epsilon_{2,t}) \\ \text{(B2)} &: \tilde{y}_{1,t} = 0.55\tilde{y}_{1,t-1} + 0.075\tilde{y}_{2,t-1} + (\epsilon_{1,t} + 0.3\epsilon_{2,t}), \end{aligned} \quad (13)$$

$$\begin{aligned} \text{(E1)} &: \tilde{y}_{1,t} = 0.70\tilde{y}_{1,t-1} + 0.075\tilde{y}_{2,t-1} + (\epsilon_{1,t} + 0.3\epsilon_{2,t}) \\ \text{(E2)} &: \tilde{y}_{1,t} = 0.55\tilde{y}_{1,t-1} + 0.075\tilde{y}_{2,t-1} + (\epsilon_{1,t} + 0.3\epsilon_{2,t}). \end{aligned} \quad (14)$$

Thus, in mean-deviation form, their DGPs are identical both pre and post break, even at the break point. This outcome confirms the analytic calculations for the single forecast at  $T + 1$ . The main message is that direct or induced deterministic shifts are more important for predictive failure than changes in

<sup>2</sup>The constancy (or break-point) Chow test performs an F-test on the current residual sum of squares (RSS) versus the final RSS, so uses  $RSS_t$  versus  $RSS_{201}$  in fig. 1, and  $RSS_t$  versus  $RSS_{105}$  in fig. 2 for  $t = 50, \dots$ . The forecast Chow test uses  $RSS_t$  versus  $RSS_{49}$ .



**Figure 1** Rejection frequencies of the parameter constancy test.

parameters such as reaction speeds, feedbacks or latencies. Below, we note the impact of that conclusion on the formulation of forecasting models.

Since break-point tests are not optimal for long post-break samples, and their power depends on the date against which tests are conducted, we ran the same experiments with only 4 post-break observations to mimic a forecasting, as against a modelling, context: this delivers fig. 2, where the upper graph is the forecast test. The rejection frequencies are higher than for the full-sample, but the pattern of results is unchanged: shifts in zero-mean processes remain a minor problem compared to non-zero means.

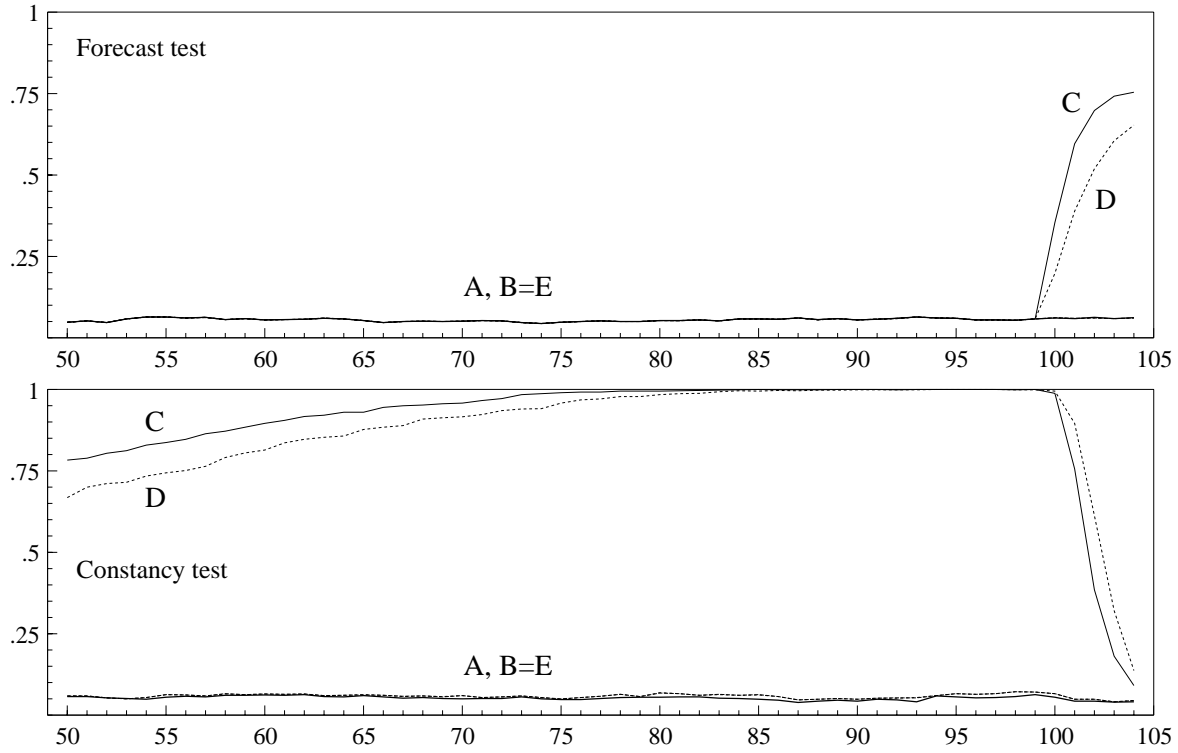
The next section considers some alternative potential causes of serious forecast errors, and explains how they can be differentiated post hoc from the mean-shift problem.

### 3 Alternative explanations

We now consider three other potential explanations for forecast failure, namely model mis-specification, collinearity, and a lack of parsimony. By themselves, none of these can explain that phenomenon, but interacting with breaks elsewhere in the economic system, can induce actual or apparent failure.

#### 3.1 Model mis-specification

In a stationary world, least-squares estimated models are consistent for their associated conditional expectations (when second moments exist), so forecasts on average attain their expected accuracy unconditionally (see, e.g., Hendry, 1979). Thus, model mis-specification *per se* cannot account for forecast failure, except to the extent that inconsistent standard errors are used to judge forecast accuracy: for, example, untreated residual autocorrelation can downward bias estimated standard errors and induce excess rejections on parameter constancy tests (see e.g., Corsi, Pollock and Prakken, 1982). Although congruent models remain false in general, by having innovation residuals they are not prone to this prob-



**Figure 2** Rejections of the forecast and constancy tests for four forecasts.

lem. We believe such a result extends to congruent models of systems that are stationary after suitable cointegration and differencing, so we look elsewhere for the causes of forecast failure.

Reconsider the  $l(0)$  VAR in (5), where before the break:

$$\begin{aligned} E[y_t] &= \varphi \\ E[(y_t - \varphi)(y_t - \varphi)'] &= M = \Omega_\epsilon + \Pi M \Pi' \\ E[(y_t - \varphi)(y_{t-1} - \varphi)'] &= \Pi M. \end{aligned}$$

In more detail, we let the DGP comprise of two blocks:

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} + \begin{pmatrix} \Pi_{1,1} & \Pi_{1,2} \\ \mathbf{0} & \Pi_{2,2} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}. \quad (15)$$

Only the first block is modelled, with the model of  $y_{1,t}$  mis-specified by omitting  $y_{2,t-1}$ . Thus, from (15):

$$\begin{aligned} y_{1,t} &= \phi_1 + \Pi_{1,1}y_{1,t-1} + \Pi_{1,2}y_{2,t-1} + \epsilon_{1,t} \\ &= (\phi_1 + \Pi_{1,2}\rho) + (\Pi_{1,1} + \Pi_{1,2}\Psi_{1,1})y_{1,t-1} + (\Pi_{1,2}u_{2,t-1} + \epsilon_{1,t}) \\ &= \delta_1 + \Gamma_{1,1}y_{1,t-1} + v_{1,t}, \end{aligned} \quad (16)$$

using:

$$y_{2,t} = \rho + \Psi_{1,1}y_{1,t} + u_{2,t}, \quad (17)$$

with  $E[y_{1,t}u_{2,t}'] = \mathbf{0}$ , so that:

$$\rho = E[y_{2,t} - \Psi_{1,1}y_{1,t}] = \varphi_2 - \Psi_{1,1}\varphi_1 = (-\Psi_{1,1} : \mathbf{I}_{n_2})(\mathbf{I}_n - \Pi)^{-1}\phi,$$

where:

$$\Psi_{1,1} = \mathbf{M}_{21}\mathbf{M}_{11}^{-1}. \quad (18)$$

Despite the mis-specification, (17) is well defined, and its error variance in any forecast period will on average match that insample: however, as in Corsi *et al.* (1982), the residual autocorrelation will lead to some excess rejection. In practice, an investigator is likely to add extra lags of  $y_1$  to remove this.

Let the  $n_2 \times 1$  intercept  $\phi_2$  of the second block change at  $T_1$  to  $\phi_2^*$ , so after  $T_1$ :

$$\mathbb{E}[y_t \mid T > T_1] = (\mathbf{I}_n - \mathbf{\Pi})^{-1} \phi^* = \varphi^*,$$

the second moments around  $\varphi^*$  remaining constant. The analogous equation to (17) becomes:

$$y_{2,t} = \rho^* + \Psi_{1,1}y_{1,t} + u_{2,t}$$

with:

$$\rho^* = (-\Psi_{1,1} : \mathbf{I}_{n_2})(\mathbf{I}_n - \mathbf{\Pi})^{-1} \phi^*.$$

Thus, after  $T_1$ , as in (16):

$$y_{1,t} = \delta_1^* + \Gamma_{1,1}y_{1,t-1} + v_{1,t}. \quad (19)$$

Now rejection should match that in §2.2; and when the sample spans the break, severe residual autocorrelation will ensue, analogous to the creation of apparent unit roots (see e.g., Hendry and Neale, 1991).

The corresponding Monte Carlo again corroborates the analysis. We used the bivariate DGP:

$$\begin{aligned} y_{1,t} &= 3 + 0.85y_{1,t-1} - 0.50y_{2,t-1} + \epsilon_{1,t} \\ y_{2,t} &= \alpha_2 + 0.80y_{2,t-1} + \epsilon_{2,t}. \end{aligned} \quad (20)$$

There were five experiments, all with  $\epsilon_{i,t} \sim \text{IN}[0, \sigma_{ii}]$ , using  $\sigma_{ii}$  of 0.25, 0.36. In the first three,  $\alpha_2$  was changed from 5 to 4 at  $T = 100$ , reverting at  $T = 150$ ; in the last two,  $\alpha_2$  remained constant. In all but the first experiment, a scalar autoregression in  $y_{1t}$  (denoted AR( $\cdot$ )) was estimated.

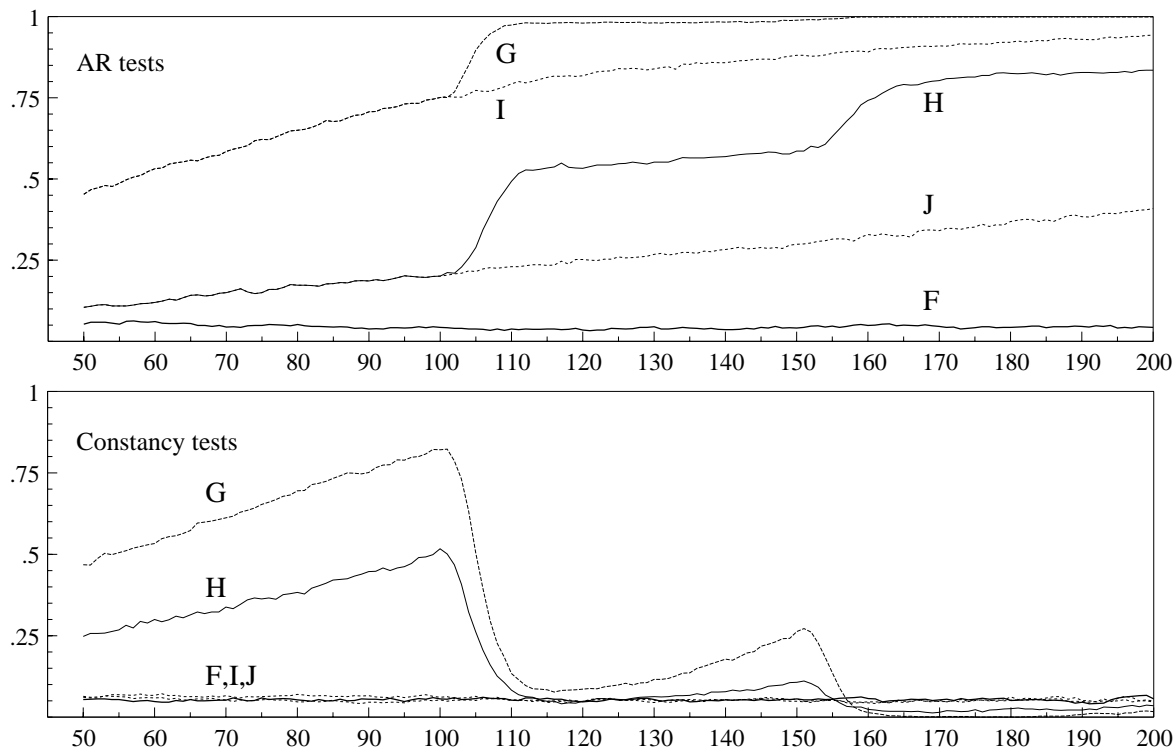
- (F) The first equation of (20),  $\alpha_2$  changed;
- (G) AR(1) in  $y_{1,t}$ ,  $\alpha_2$  changed;
- (H) AR(2) in  $y_{1,t}$ ,  $\alpha_2$  changed;
- (I) AR(1) in  $y_{1,t}$ ,  $\alpha_2$  constant;
- (J) AR(2) in  $y_{1,t}$ ,  $\alpha_2$  constant.

Figure 3 shows the resulting Chow constancy test and autocorrelation test rejection frequencies<sup>3</sup>. The autocorrelation is marked for 1 lag, but greatly attenuated for 2, when there is no break. In both these cases, the Chow test rejects somewhat more than its nominal size of 5%. However, in the cases with a break, the constancy rejections are high for both lag specifications. Thus, an entailed mean shift is the culprit once more. Adding further lags of  $y_1$  eliminates the remaining residual autocorrelation (little more than 5% rejection with 5 lags), thereby reducing the power of this constancy test to about 30%.

---

<sup>3</sup>The autocorrelation test is the F-version of the Lagrange multiplier test of Breusch (1978) and Godfrey (1978), testing for first order autocorrelation, following Kiviet (1986).





**Figure 3** Autocorrelation and constancy test rejection frequencies for an omitted variable.

### 3.2 Collinearity

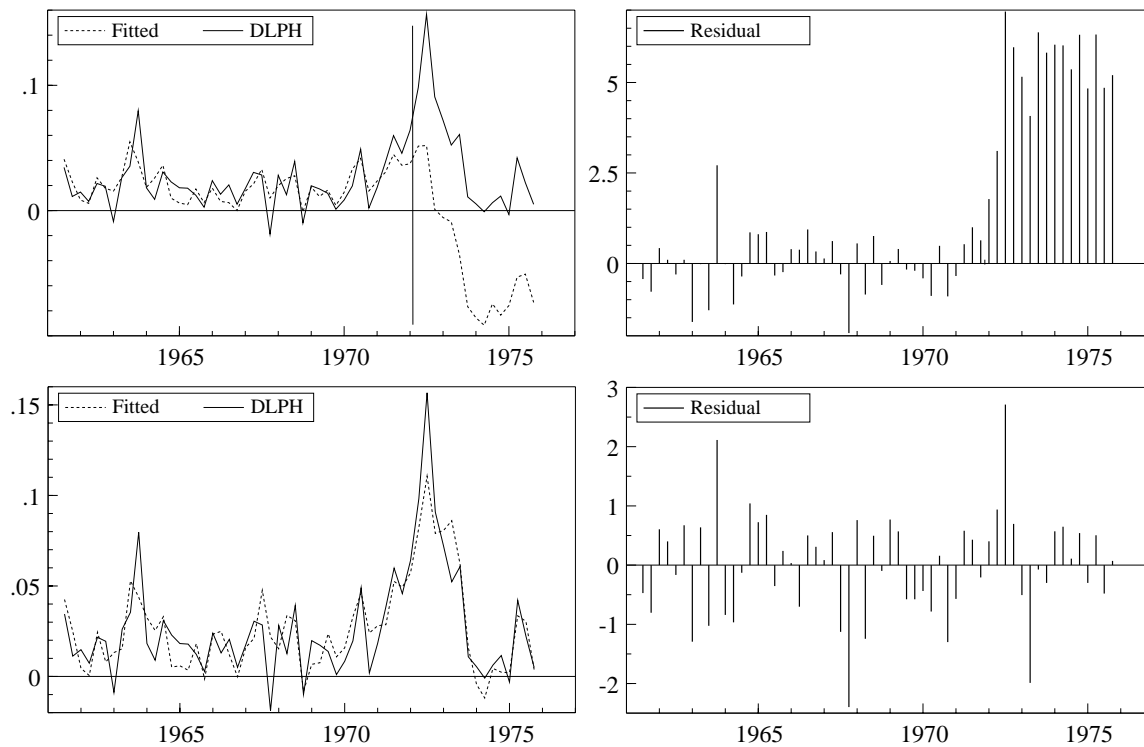
Next, it may be thought that severe ‘collinearity’ might help account for forecast failure, since very imprecise parameter estimates could be far from the correct values, and hence produce misleading forecasts. However, in a linear model, collinearity cannot do so by itself for reasons given in Clements and Hendry (1995) and Hendry (1996a), namely that forecasts are invariant under non-singular, scale-preserving, linear transformations, and hence are identical to those delivered for orthogonal variables.

Nevertheless, poor ex-ante forecasts can result from a constant model when there is a break in the process generating the regressors. Denote by  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$  the eigenvalues of the second-moment matrix  $\mathbf{Z}'\mathbf{Z}$  of the  $n$  regressors  $\mathbf{z}_t$  in a linear model, where the ratio  $\lambda_n/\lambda_1$  is large. Let the impact of the forecast-period observations change the  $\lambda_i$  to  $\lambda_i^*$ . Then, from the usual formula for the 1-step ahead forecast error,  $\hat{\nu}_{T+1}$  has an expected squared value of:

$$E[\hat{\nu}_{T+1}^2] \simeq \sigma_v^2 \left( 1 + T^{-1} \sum_{i=1}^n \frac{\lambda_i^*}{\lambda_i} \right),$$

which can be large if  $\lambda_1$  changes markedly (e.g., from 0.0001 to 0.1, say, so a factor of 1000 is involved). Consequently, variance effects interacting with breaks in ‘exogenous’ processes can produce poor forecasts. In practice, this situation would be signalled by a large increase in forecast confidence intervals when these were calculated; as this is not always the case, we proceed as if none were.

Forecasting UK house prices over 1973–75 provides a possible empirical example, with ‘forecast failure’ being induced by the many changes in financial markets following the Competition and Credit Control regulations. Indeed, the models in Hendry (1984) produce poor ex-ante forecasts for the resultant price leap, as the top row of fig. 4 shows for a model like his equation (17) (replacing  $(\Delta p_{h,t-1})^3$  by  $\Delta p_{h,t-1}$ ). The forecasts suggest a large fall in house prices (in the midst of one of the most rapid booms ever!), and the forecast-period residuals are large and of the same sign.



**Figure 4** UK house price inflation: fitted values, forecasts and errors over 1972(1)–1975(4).

Similar outcomes of ex-ante forecast failure in conditional models have occurred many times in the history of econometrics: parochial examples include consumers' expenditure in the mid-1970s and early 1990s (see Davidson, Hendry, Srba and Yeo, 1978, Hendry, 1994, and Clements and Hendry, 1997a, for a follow up), and UK M1 (see Hendry and Mizon, 1993, and Hendry, 1996b). The first of these seems due to omitting the loss on liquid assets induced by the rapid increase in inflation, and a model including inflation became constant; the next (consumers' expenditure in the 1990s) remains a puzzle to some extent (see Muellbauer, 1994), but appears to be associated with the consequences of credit deregulation and negative equity in housing; and the last (UK M1) reflects the introduction of interest-bearing checking accounts which greatly reduced the opportunity cost of holding money, and was rectified by re-measuring that notion. Thus, all three examples seem to be breaks within the pre-existing relation relative to the relevant information set. How can one differentiate those outcomes (which deliver similar, if less dramatic, graphs for forecast residuals to fig. 4) from the present case where a break in the regressor set induces apparent failure in what is in fact a constant relation? Fortunately, the discrimination is direct from the postulated source: the collinear situation is detectable either from forecast confidence intervals widening markedly (see fig. 10d below), or by the ex-post fit being similar to the ex ante. For the house-price equation, both are the case: e.g., it does not fit significantly worse to the ex-post data ( $F(16, 28) = 1.75$ ,  $p = 0.094$ ), despite the large change in the data. Updating the consumers' expenditure, or M1 equations, without the required modifications, produces significantly worse fits. Here, the contrast in Fig. 4 between the ex-ante forecasts and forecast errors (top row) and the ex-post fit (second row), computed with the model specification unchanged, confirms no break in the house-price equation under analysis: similar contrasts hold for other specifications. An alternative explanation is provided by Richard and Zhang (1996), who use the same underlying theory to show that considerable heteroscedasticity will result from Hendry's non-linear adjustment equation. Measuring the misprediction relative to an assumed constant error variance will then lead to a mistaken inference, in so far as using the correct heteroscedastic variance would entail an insignificant constancy-test outcome.

### 3.3 Lack of parsimony

This notion is taken to mean both incorrectly including (proxy) variables, and correctly including those that have small partial effects (conditional on the remaining specification), the elimination of which would improve forecast performance on some measure. Since the main consequence of the second form of lack of parsimony is imprecise parameter estimates, it too could at best explain ‘excess forecast variance’, so that somewhat inaccurate forecasts are produced. However, systematic forecast failure (the main problem discussed above) could occur in the first case following a change in data properties somewhere in the associated economic system. Thus, falsely including regressors that alter substantially in a forecast period could induce forecast failure (see e.g., Clements and Hendry, 1995).

To mimic this case in our Monte Carlo, we let  $y_{2,t}$  depend on  $y_{1,t}$  directly (i.e., contemporaneously) with cross-correlated errors, then incorrectly include  $y_{2,t}$  as a regressor in the model of  $y_{1,t}$  when the equation determining  $y_{2,t}$  undergoes a structural break in its equilibrium mean. This example therefore also embodies invalid weak exogeneity, and is an illustration of testing for it by a change in a marginal process (see Engle and Hendry, 1993, and Favero and Hendry, 1992). The upper row of Fig. 5 shows the rejection frequencies for both the break-point and forecast tests for  $T = 104$  when:

- (a) there is no break, using the correctly specified  $y_{1,t}$  model;
- (b) there is no break, with  $y_{2,t}$  falsely included in the  $y_{1,t}$  model;
- (c) the intercept shifts in the  $y_{2,t}$  equation, using the correctly specified  $y_{1,t}$  model;
- (d) the intercept shifts in the  $y_{2,t}$  equation, with  $y_{2,t}$  falsely included in the  $y_{1,t}$  model.

Thus, the DGP has the form:

$$\begin{aligned} y_{1,t} &= 1 + 0.85y_{1,t-1} + \epsilon_{1,t} \\ y_{2,t} &= \alpha_2 + 1.95y_{1,t} + 0.80y_{2,t-1} + \epsilon_{2,t}, \end{aligned} \quad (21)$$

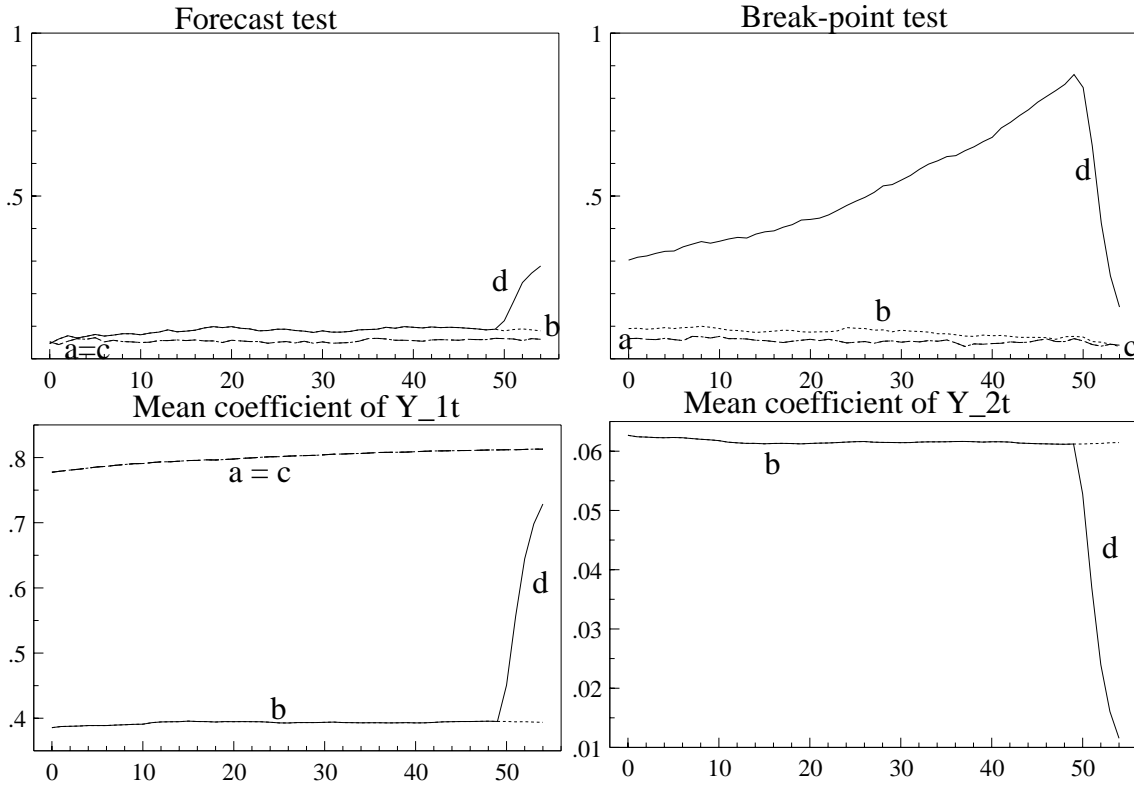
where:

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \sim \text{IN}_2 \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} (0.25)^2 & (0.28)^2 \\ (0.28)^2 & (0.36)^2 \end{pmatrix} \right].$$

We set high inter-equation correlations and a large change from 7 to 0 for the shift in  $\alpha_2$  to emphasise that important non-constancies can derive from falsely including variables that change.

Figures 5a,b show that there is slight over-rejection in case (b) due to the ‘simultaneity’, although both (a) and (c) deliver the correct size. The rejection frequency in (d) exceeds 75% just prior to the break using the break-point test, but is much lower on the forecast test. If that outcome empirically led to deleting  $y_{2,t}$ , then the full-sample fit of the reduced model would again be constant. Indeed, the main diagnosis of this case is the reduced significance of wrongly-included variables after re-estimation over the enlarged sample, and the increased constancy that should result when they are deleted from the model. Figures 5c,d show the recursively-computed mean values of the coefficients of  $y_{1,t}$  and  $y_{2,t}$  in case (d): four post-break observations are rather few to be definitive, but the coefficient of  $y_{2,t}$  dives almost immediately to zero, and that of  $y_{1,t}$  rises to the DGP value: their respective standard errors lay between  $(0.02) - (0.005)$  and  $(0.15) - (0.05)$ . There was no evidence of coefficient non-constancy in the other cases, as their graphs demonstrate.

There would be a failure of variance dominance here, in that the whole-sample constant model would fit less well than the non-constant over the period before the break, requiring care in evaluating between non-congruent models. Unfortunately, such an analysis would also apply in an  $I(1)$  process when the forecast failure is induced by a shift in the equilibrium mean, since dropping the ECM could produce a more constant model in differences. We investigate that issue in §5 after discussing the potential role of causal information.



**Figure 5** Constancy test rejection frequencies and recursive coefficients.

## 4 Causal information

Once models are mis-specified for the DGP, and the system is subject to structural breaks, it cannot be proved that causally-relevant variables must produce better forecasts than non-causal. Thus, forecasts based on the latter can dominate those based on causal influences, as we now show by an example. The DGP is the  $I(1)$  system:

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} \Pi_{1,2} y_{2,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \quad (22)$$

which holds till time  $\tau$ , then changes to:

$$\begin{pmatrix} y_{1,\tau+i} \\ y_{2,\tau+i} \end{pmatrix} = \begin{pmatrix} \Pi_{1,2}^* y_{2,\tau+i-1} \\ y_{2,\tau+i-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,\tau+i} \\ \epsilon_{2,\tau+i} \end{pmatrix}. \quad (23)$$

Only the first block in (22) is modelled, with forecasts generated by the correct insample system:

$$\hat{y}_{1,\tau+i} = \Pi_{1,2} y_{2,\tau+i-1}, \quad (24)$$

so that after the break, the forecast error is:

$$E[y_{1,\tau+i} - \hat{y}_{1,\tau+i} | y_{2,\tau+i-1}] = (\Pi_{1,2}^* - \Pi_{1,2}) y_{2,\tau+i-1}.$$

Since  $y_{2,t}$  is a vector random walk, it will almost always be non-zero, so the forecast error could be large as well as persistent.

By way of contrast, consider the following non-causal forecasting rule:

$$\tilde{y}_{1,\tau+i} = y_{1,\tau+i-1}.$$

This is a purely extrapolative mechanism, but ‘error corrects’ to the previous level. Thus, using the fact that the unconditional growth,  $E[\Delta \mathbf{y}_{2,\tau+i-1}] = \mathbf{0}$ :

$$\begin{aligned} E[\mathbf{y}_{1,\tau+i} - \tilde{\mathbf{y}}_{1,\tau+i} \mid \mathbf{y}_{2,\tau+i-1}] &= E[\Delta \mathbf{y}_{1,\tau+i} \mid \mathbf{y}_{2,\tau+i-1}] \\ &= \begin{cases} (\mathbf{\Pi}_{1,2}^* - \mathbf{\Pi}_{1,2}) \mathbf{y}_{2,\tau-1} & \text{for } i = 1 \\ \mathbf{0} & \text{for } i > 1. \end{cases} \end{aligned}$$

By construction, the lagged  $\mathbf{y}_1$  is non-causal, yet forecasts based on it are less biased after  $\tau + 1$ . For possible changes in parameter values, such shifts could dominate any variance losses. This example leads into the next section’s discussion of VEqCMs and differenced models.

## 5 Equilibrium correction and error correction

The preceding analysis has important implications for the use in forecasting of equilibrium-correction mechanisms embodying cointegration relations. The formulation in (4) clarifies that once the ECMs are expressed as deviations from their means, the remaining intercepts are the growth rates. Since growth rates are small numbers (0.001 – 0.02 for quarterly real variables) – and are usually small even relative to residual standard deviations – and cannot change by large amounts, the main problems from breaks concern shifts in the equilibrium means. Both forms of shift can induce large shifts in levels, which will appear in practice as sustained unidirectional changes in some endogenous variables. Because estimated ECMs drive relations back to the previous equilibrium means, they will take on large values — of precisely the wrong sign. This follows because the changed levels are viewed by the estimated model as disequilibria; the higher those levels, the more it seeks to ‘correct’ them back to the old equilibrium, and hence the worse the forecast. Indeed, graphs similar to fig. 4 may be observed, despite in-sample congruency, well-determined estimates, and theoretically-supported cointegration (see, for example, figure 2 in Hendry, 1996b). Such an outcome is far from ‘error correction’, and prompted the renaming to equilibrium correction.

From (4), the expected deviation from the outcome of a forecast for  $\Delta \mathbf{x}_t$  based on an existing equilibrium mean  $\boldsymbol{\mu}_0$  after a shift to  $\boldsymbol{\mu}_1$  is:

$$E[\Delta \mathbf{x}_t - \gamma - \alpha(\boldsymbol{\beta}' \mathbf{x}_{t-1} - \boldsymbol{\mu}_0)] = \alpha(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1). \quad (25)$$

Thus, the growth in  $\mathbf{x}_t$  is mispredicted by  $\alpha(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)$  every period, which could be large relative to  $\gamma$ . Consequently, ECMs will be reliable in forecasting only if they contain all the variables needed to track changed states of nature. Nevertheless, in policy analyses, they will frequently play a central role, so a method of retaining their economic importance and statistical efficacy without losing forecast robustness is highly desirable. Hendry and Mizon (1996) consider some approaches to obtaining error correction for non-stationary processes, but conclude that common factors of unity are required, entailing differencing, and hence the elimination of the ECM. Thus, forecast robustness ideas, such as intercept corrections (see Clements and Hendry, 1996), and forecast-period corrections along the lines of exponentially-weighted moving averages, merit consideration.

### 5.1 An I(1) Monte Carlo illustration

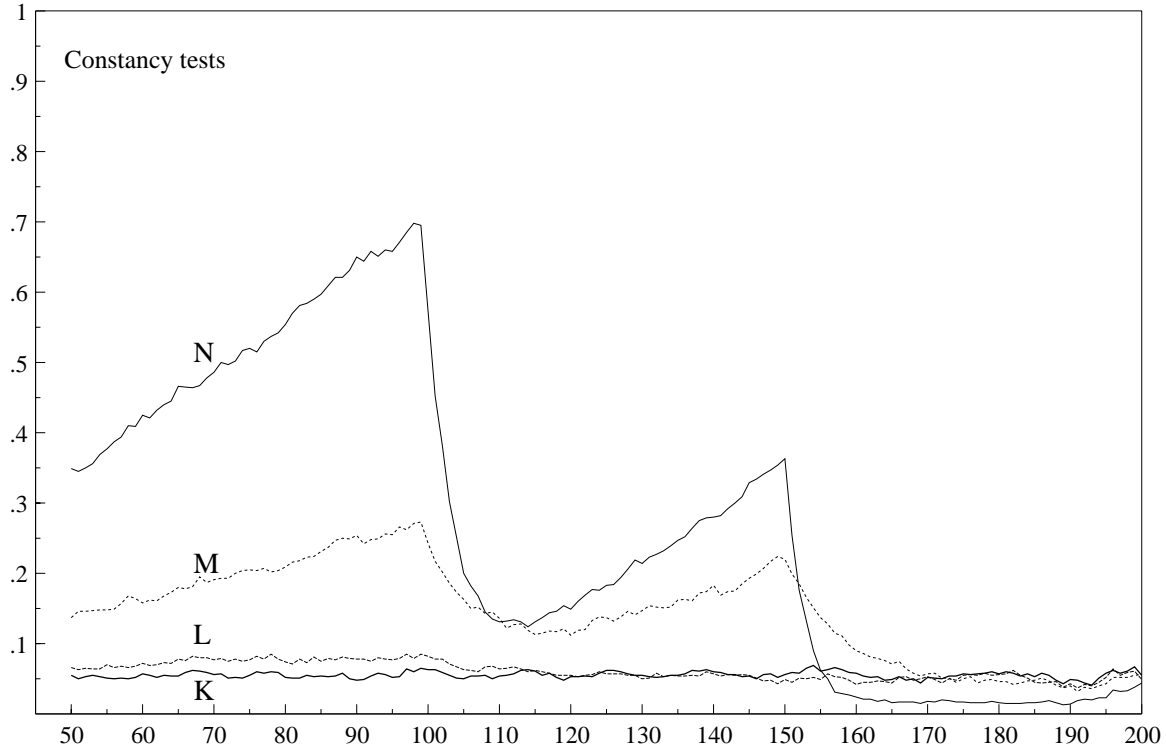
We illustrate this analysis using an I(1) cointegrated system, estimating the model in levels: §5.2 reports estimates in differences. The data are generated by the bivariate VAR in (26), and we consider 4 experiments:

- (K) constant DGP;
- (L) a break in the coefficient of the ECM;
- (M) a break in the long-run mean;
- (N) a break in the growth rate.

The I(1) DGP is:

$$\begin{aligned}\Delta y_{1,t} &= \gamma_1 + 0.3\Delta y_{2,t} + \pi_1 (y_{1,t-1} - y_{2,t-1} - \mu_1) + \epsilon_{1,t} \\ \Delta y_{2,t} &= 0.025 + \epsilon_{2,t}.\end{aligned}\tag{26}$$

The  $\epsilon_{i,t} \sim \text{IN}[0, \sigma_{ii}]$ , with  $\sqrt{\sigma_{ii}}$  of 0.025, 0.036. As before, breaks occur at  $T = 100$ , reverting to the original value at  $T = 150$  with the full-sample size of  $T = 200$ . When the dynamics are constant,  $\pi_1 = -0.15$ , but for its break (case (L)),  $\pi_1$  changes from  $-0.15$  to  $-0.3$ . We set  $\mu_1 = 1$ , changed to 1.33 for the break in (M); and  $\gamma_1 = 0.025$ , changed to 0.05 in (N). Thus, the drift and intercept changes are one and two error standard deviations respectively (see Hylleberg and Mizon, 1989). The formulation is invariant to what induces changes in  $\gamma_1 - \pi_1\mu_1$ , but large growth-rate changes are unlikely in OECD countries, whereas a shift in (say) the ratio of money to income from 25% to 10% (i.e., 0.15) is within the observed range empirically.



**Figure 6** Constancy-test rejection frequencies in a cointegrated process.

In these experiments, the first equation is estimated unrestrictedly, in levels:

$$y_{1,t} = \beta_0 + \beta_1 y_{2,t} + \beta_2 y_{1,t-1} + \beta_3 y_{2,t-1} + \epsilon_{1,t}\tag{27}$$

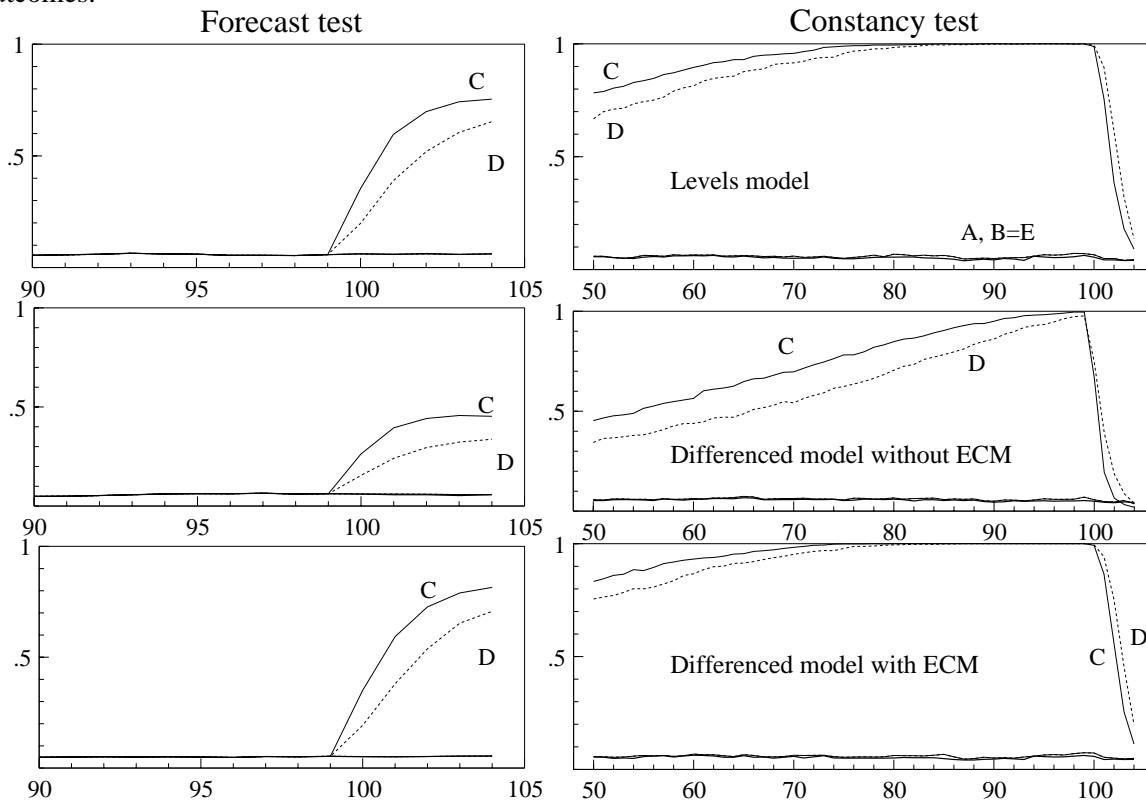
Experiment (K) checks the size of the constancy test in an I(1), cointegrated, setting, (L) demonstrates that a change in the strength of reaction to a zero-mean disequilibrium is not detectable, and mimics (B) and (E) above, whereas (M) mimics the break effects found in case (D) above, and (N) examines a doubling of the growth rate. Figure 6 confirms the anticipated outcomes: the break-point test rejection

frequencies under the null are close to their nominal size of 5%; the break in the equilibrium mean is easily detected; but that in the dynamics alone is not detectable. (N) is a large change in real growth, but is equivalent to half the change in (M), and at about 25% rejection, is far less perceptible on these tests.

## 5.2 Non-congruent devices need not fail

The preceding example suggests the need to avoid systematic mis-forecasting by ‘robustifying’ to structural breaks. To the extent that such can be achieved, the resulting model need not fail, but need not be congruent either, as in the following differencing example.

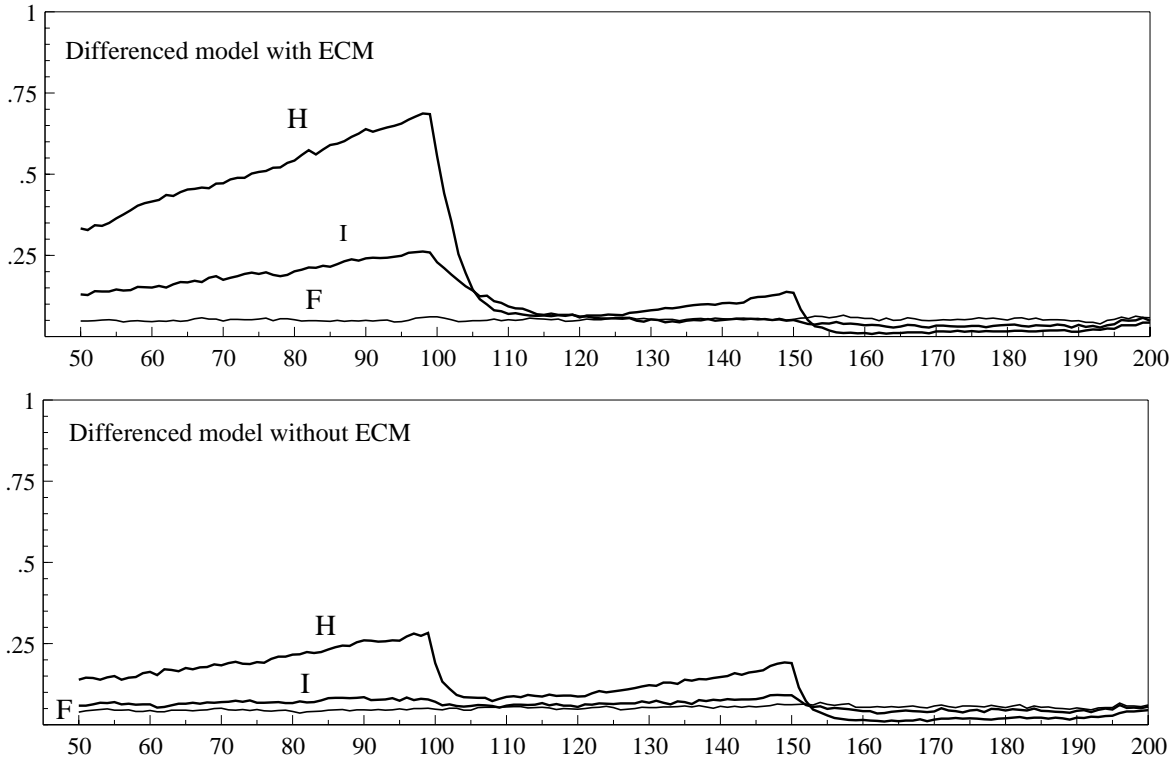
Since step shifts in intercepts are reduced to blips by differencing, we might anticipate less failure for first-differenced models. In the Monte Carlo experiments in §2.2 and §5.1, as well as the unrestricted levels model (27), we therefore also estimated models in differences (i.e., omitting the ECM), and as ECMs with the cointegration restriction imposed (the first equation in (26)). Figures 7 and 8 show the outcomes.



**Figure 7** Constancy-test rejection frequencies for the  $I(0)$  process.

The first column in fig. 7 shows the forecast-test outcomes, and the second column the constancy test for models in levels (i.e., fig. 2), differences, and levels reparameterized as an ECM with the long-run coefficient imposed. In this  $I(0)$  situation, the breaks are still easily detected in the differenced model. However, that is not so clearly the case for (26) when the system is  $I(1)$  as the upper graph in fig. 8 shows. Since there is no natural scale for the intercept, a step break in the intercept can ‘swamp’ other terms even after becoming a blip on differencing, leading to substantial — and easily detected — non-constancy. That is unlike the outcome for a cointegrated DGP, where intercepts are growth rates (as in (4) above): only small changes in  $\gamma$  are likely, and any changes in  $\mu$  are irrelevant to models in differences. The empirical finding that differencing is effective in reducing predictive failure becomes indirect evidence that there are unit roots in economic processes.

For cointegrated processes, Clements and Hendry (1996) show that a VEqCM and DVAR for (1)



**Figure 8** Constancy-test rejection frequencies for the  $I(1)$  process.

have identical forecast-error biases when a forecast is made before a break occurs for a horizon that includes the break. This is so despite the former including, and the latter excluding, all the cointegration information, although their forecast-error variances will differ. The biases for the VEqCM do not depend on whether the forecast starts pre or post the break: thus, there is no error correction after the break. However, the DVAR has different biases pre and post for breaks in  $\mu$ , and these are usually smaller than the corresponding biases from the VEqCM. Extending the VEqCM in (4) for an extra lagged difference:

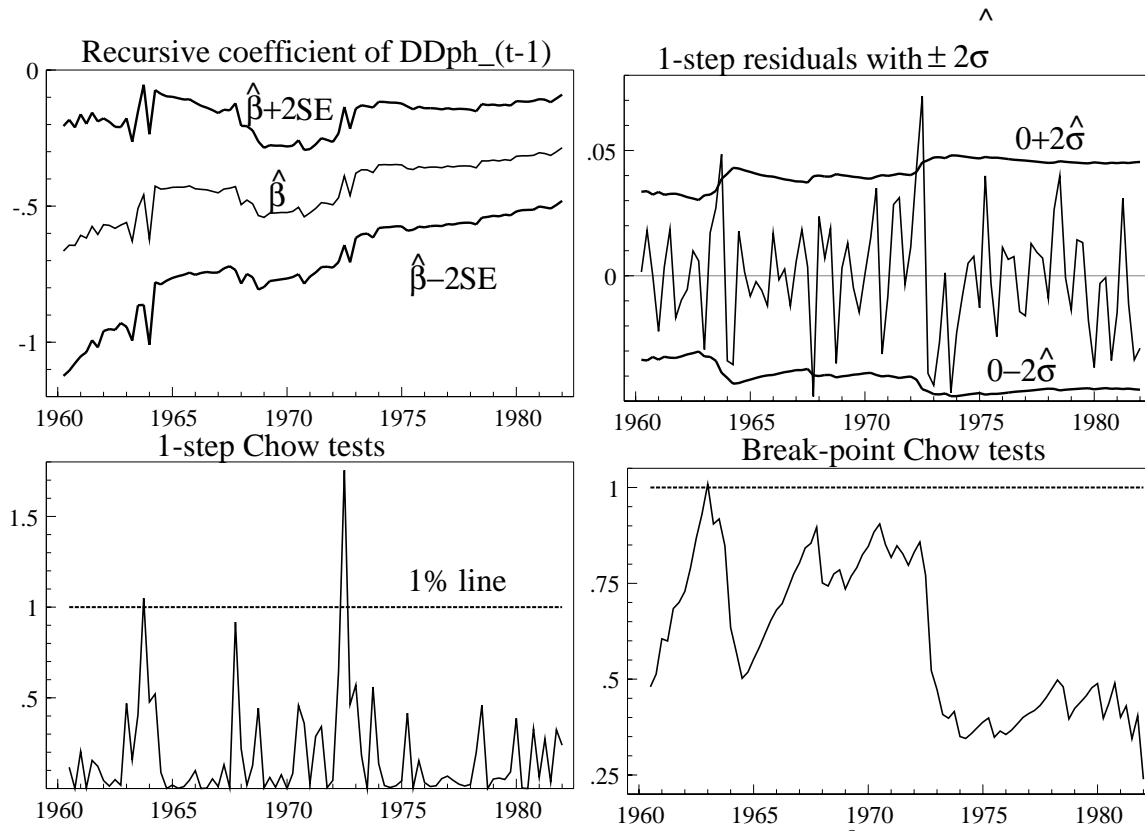
$$\begin{aligned}\Delta \mathbf{x}_t &= \gamma_0 + \Gamma \Delta \mathbf{x}_{t-1} + \alpha (\beta' \mathbf{x}_{t-1} - \mu) + \nu_t \\ &\simeq \gamma_0 + \Gamma_0 \Delta \mathbf{x}_{t-1} + \mathbf{u}_t,\end{aligned}\tag{28}$$

where the second line shows the model in differences, with  $\Gamma_0$  defined by the condition  $E[\Delta \mathbf{x}_{t-1} \mathbf{u}_t'] = \mathbf{0}$ . So long as the ECM has expectation zero, forecasting models like (28), or its counterpart with a vector unit root imposed,  $\Delta^2 \tilde{\mathbf{x}}_{T+1} = \Delta^2 \mathbf{x}_T$ , become robust to many of the shifts, but may ‘over-insure’ by not predicting any developments of interest, merely ‘tracking’ by never being badly wrong (see Clements and Hendry, 1997b).

Although such devices can improve forecast accuracy, especially on bias measures, they entail nothing about the usefulness for other purposes of the forecasting model. Even if the resulting (intercept-corrected or differenced) forecast is more accurate, this does not entail choosing the ‘robustified’ model for policy, or later modelling exercises.

As an empirical illustration, we reconsider the house-price series, since that experienced many major perturbations over the period 1955–82. Figure 9 shows the recursively-estimated coefficient for a first-order autoregression in the second difference of the log of UK house prices,  $\Delta^2 p_{h,t}$  together with the 1-step ahead forecast errors ( $\pm 2SE$ ), and 1-step and break-point Chow tests computed recursively. Despite all the known institutional changes, and the lack of any economic explanatory variables, no





**Figure 9** UK house price autoregression for  $\Delta^2 p_{h,t}$ .

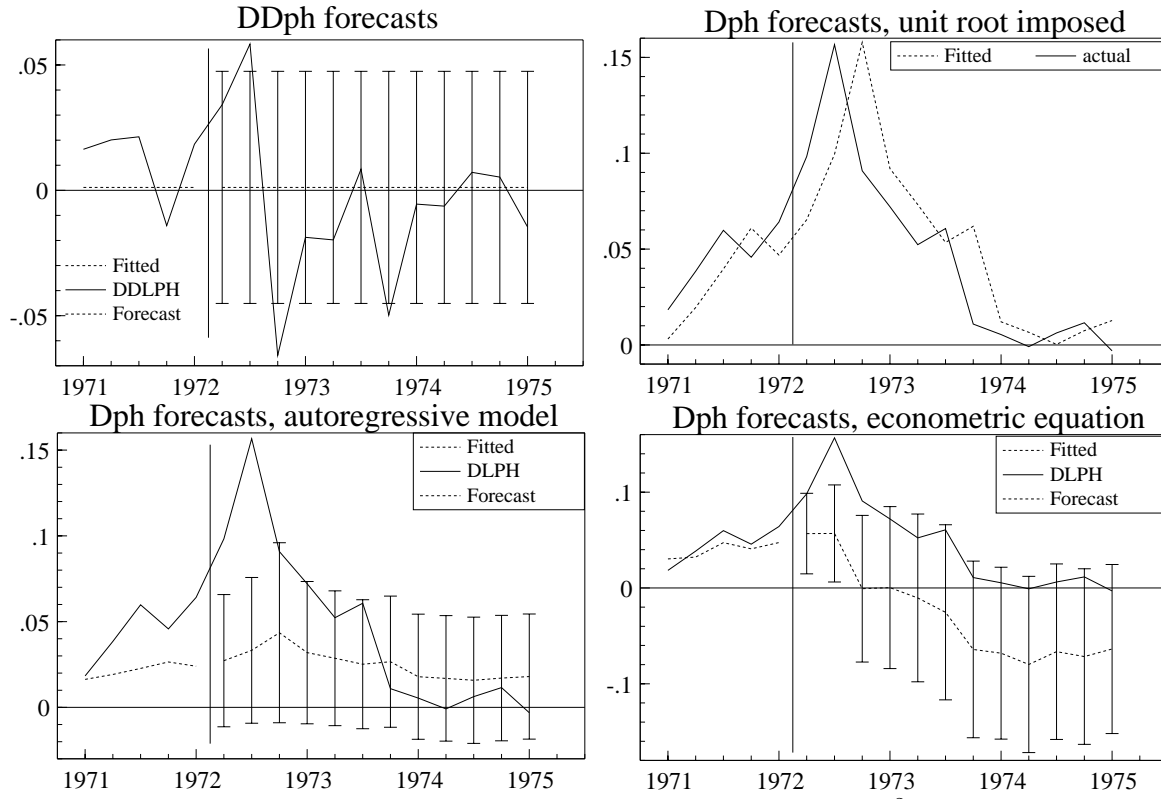
break-point test is significant at a (one-off) 1% level, and only one outlier is seen (1972(3)).<sup>4</sup> A more detailed comparison of the forecasts and outcomes over 1972(1)–1975(1) is shown in the four panels of fig. 10a–d: for  $\Delta^2 p_{h,t}$  and  $\Delta p_{h,t}$  from the second-differenced model (forecast standard errors are not available for the latter), and for  $\Delta p_{h,t}$  from the autoregressive and econometric equations. Two significant forecast errors are apparent for  $\Delta^2 p_{h,t}$ , leading to systematic over-prediction in terms of  $\Delta p_{h,t}$ . However, these seem respectable in comparison to the second row, where large forecast errors result or where the forecast confidence bars mushroom from the third forecast on (in the last case), after the jump in house prices, consistent with the analysis in §3.2.

## 6 Extended model constancy

Given the occurrence of non-constancy in practice, we now consider possible responses. First, if the failure is due to model mis-specification revealed by a change in the omitted variables' behaviour as in §3.1, a respecification is essential and could recreate a full-sample constant representation. This situation corresponds to non-constancy in the pre-existing model, but is distinct progress, rather than failure, as it represents a genuine increase in understanding. Since several breaks could be attributable to a single omitted influence, small extensions could in principle account for many forecast failures. Of course, discovering the culprit may require considerable ingenuity as well as improved theoretical analyses and possibly new measurements.

A second possibility is that the extended model does not entail non-constant parameters in the initial model. This may seem paradoxical, but can be explained as follows: the definitions here are from

<sup>4</sup>All is not lost however: notwithstanding its apparently impressive constancy, this autoregressive equation in  $\Delta^2 p_{h,t}$  fails a directed constancy test for the period of the house-price boom (over 1973–75,  $F(16, 40) = 2.21$ ,  $p = 0.022^*$ ), even though its residual standard deviation is 50% larger than that of the 'structural model' used in §3 ( $\hat{\sigma} = 2.8\%$  as against  $\hat{\sigma} = 1.5\%$ ).



**Figure 10** UK house price forecasts for  $\Delta p_{h,t}$  and  $\Delta^2 p_{h,t}$ .

Hendry (1995a), and the following analysis builds on Hendry (1996b). A parameter  $\theta \in \Theta \subseteq \mathbb{R}^k$  must be constant across realizations of the stochastic process it indexes, but need not be constant over time. It is constant over a time interval  $\mathcal{T} = \{\dots, -1, 0, 1, 2, \dots\}$  if  $\theta$  takes the same value for all  $t \in \mathcal{T}$ . A model is constant if all its parameters are. Even so, it is well known that constant models can have time-varying coefficients, which are, in effect, latent variables that depend on more basic parameters, when these in turn are constant: the models in Harvey and Shephard (1992) are an example.

All 1 – 1 transforms of parameters are valid, so zero can become a population value of a parameter  $\psi$  in  $\psi = \mathbf{h}(\theta) \in \Phi \subseteq \mathbb{R}^k$ . Let the original parameters for the density  $D_{y|z}(\mathbf{y}_t | \mathbf{z}_t; \theta)$  of  $\mathbf{y}_t$  over  $T_1$  be  $\theta$  given a conditioning vector  $\mathbf{z}_t$ , where sequential factrization of the  $T_1$ -sample joint density yields:

$$\prod_{t=1}^{T_1} D_{y|z}(\mathbf{y}_t | \mathbf{z}_t; \theta). \quad (29)$$

Consider a setting where to adequately characterize the whole  $T$ -sample requires an extended model with parameters  $\rho \subseteq \mathbb{R}^K$  (for  $K > k$ ) given  $K$  variables  $\mathbf{x}_t$ :

$$\prod_{t=1}^T D_{y|z}(\mathbf{y}_t | \mathbf{x}_t; \rho). \quad (30)$$

In (30), we assume that  $\rho$  is constant. Prima facie, the model in (29) seems non-constant, but if there exists a 1 – 1 mapping from  $\rho$  to  $\delta$  (say), such that:

$$\delta = \mathbf{g}(\rho) = (\theta, \mathbf{0}), \quad (31)$$

then the model is constant despite the apparent expansion. For example, when  $\mathbf{x}_t$  coincides with  $\mathbf{z}_t$  over the  $T_1$ -sample, but is an alternative measure for the same construct over the forecast period, then (29)

is:

$$\prod_{t=1}^{T_1} D_{y|z}(\mathbf{y}_t | \mathbf{z}_t; \boldsymbol{\theta}) = \prod_{t=1}^{T_1} D_{y|z}(\mathbf{y}_t | \mathbf{x}_t; \boldsymbol{\theta}), \quad (32)$$

and appropriate transforms of (30) can re-express it as:

$$\prod_{t=1}^T D_{y|z}(\mathbf{y}_t | \mathbf{x}_t; \boldsymbol{\delta}) = \prod_{t=1}^T D_{y|z}(\mathbf{y}_t | \mathbf{x}_t; \boldsymbol{\theta}). \quad (33)$$

An empirical example of this situation is when an outside interest rate  $r_c$  is used as a measure of the opportunity cost of holding money over a period where the own rate  $r_o$  is zero; the introduction of a non-zero own rate can induce forecast failure, perhaps countered by adding  $r_o$  to the model, thereby demonstrating apparent non-constancy. However, successfully replacing the previous measure by the differential  $r_c - r_o$  as the correct measure of opportunity cost, with the coefficient on the added variable becoming zero, then reduces the model back to the original dimension with unchanged parameters.

Constancy has not lost its operational content because of this possibility: (29) is not constant in the space of  $(\mathbf{y}_t : \mathbf{z}_t)$  for all  $T$ . Moreover, there need be no  $\mathbf{g}(\cdot)$  which reproduces the original parameters augmented by zeroes. However, the result does imply that there are no possible in-sample tests (up to  $T_1$ ) for later predictive failure, since the observation of failure depends on how the model is extended, not on its within-sample properties.

### 6.1 Is ex-ante non-constancy a fatal flaw?

A critic might doubt the role for empirical models that are congruent within sample when there is a non-negligible probability that they will fail out of sample. We must emphasise 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, dummies for purchase tax changes in Davidson *et al.* (1978) later successfully accommodated both the introduction of VAT and its near doubling in 1979.

Conversely, that some breaks can be attributed to a scenario of a constant, congruent model confronted with post-sample change does not entail all are: for example, models may already have broken down in-sample, and later analysis of their predictive failure just reveals that they are not useful even as data descriptions.

A hypothetical analogy of why forecast failure need not be fatal is Halley's prediction of the return of the comet now named for him. Several planets were then unknown, and had one slightly perturbed that comet's trajectory such that it then collided in the Oort cloud with another object, it would never have returned – and become a classic case of forecast failure. However, a Mystic Meg might have claimed in Halley's time that his comet would never reappear, and then have been correct. Nevertheless, Newtonian gravitation theory is not refuted by such predictive failure, nor are Meg's claims to mystic sources validated. Progressive understanding would have uncovered the explanation in due course, and been justly regarded as success.

## 7 Conclusion

The main conclusion is that there is no essential conflict between the existence of structural breaks and congruent modelling: the former highlight weaknesses in models, which the latter uses to help us

to learn from our mistakes and eventually thereby improve empirical knowledge. The best available models would gradually incorporate all the main shift factors stimulated through time, and thereafter be more robust. Nevertheless, one could still be caught out when forecasting: e.g., by the sudden outbreak of a war for which no dummy was included. Here our analogy is of a spacecraft forecast to reach the moon at a specific time, prior to being hit by a meteor and knocked off course: no implications arise for the physical principles on which trajectories are calculated, nor even for the forecasting rules in use at NASA, unless their methods failed to adapt to the altered state.

Next, the primary factor precipitating forecast failure is a shift in the intercept: all the Monte Carlo simulations illustrate various facets of this analysis for both  $I(0)$  and  $I(1)$  processes. Conversely, model mis-specification, collinearity, and lack of parsimony *per se* are not the primary cause of forecast failure, albeit that they may exacerbate that problem. An implication is that modelling methods seem unlikely to account for forecast failure either, a conclusion reinforced by the results in Hoover and Perez (1996).

By post-hoc modelling of unpredicted changes, equations may be ‘overfitted’ relative to their likely forecast performance. If future unanticipated breaks occur, the average forecast mean-square error will exceed that found on fitting the past. This raises the issue as to whether we should measure innovation variances with or without structural-break dummies. The latter may be a better guide to likely forecast errors, but the former may be more useful in a policy context by enabling less imprecise parameter estimates. In terms of modelling strategy, we also favour the former (i.e., including dummies for externally-identified breaks – not just observed outliers), perhaps computing forecast confidence intervals from the model with no dummies.

A more fundamental difficulty is that causal variables need not dominate non-causal ones for forecasting when models are mis-specified in a world subject to breaks in deterministic factors (constants, trends etc.): see Hendry (1996a). At first sight, the basis of congruent modelling seems under threat since such a phenomenon allows for the possibility of designing non-congruent forecasting devices that are robust to important classes of structural breaks. Certainly, that drives a wedge between forecasting and modelling, and in particular precludes any claim that the ‘best forecasting’ model should be selected for policy. However, as argued above, explaining the forecast failure allows progress in modelling. Moreover, the concept of co-breaking (see Hendry, 1995b) suggests the possibility of eliminating structural breaks in deterministic factors in systems of equations by taking linear combinations of variables, analogous to cointegration removing unit roots. To be sustainable, co-breaking combinations essentially require causal relations.

Finally, there are several empirical examples of ‘extended constancy’, whereby, despite initial forecast failure, a model may be found to provide a constant *ex post* explanation with unchanged parameters: see Hendry (1996b) and Ericsson, Hendry and Prestwich (1997). In both of cases, inappropriate measurement of some of the explanatory variables over the forecast horizon accounted for the initial mis-forecasting, which vanished when more plausible measures were adopted. En route, an expanded (and therefore non-constant) model which included additional variables was thought necessary to achieve a constant fit, but a re-parameterization existed that reduced it to the original model, confirming constancy.

## References

- Breusch, T. S. (1978). Testing for autocorrelation in dynamic linear models. *Australian Economic Papers*, **17**, 334–355.
- 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. (1994). Towards a theory of economic forecasting. In Hargreaves, C. (ed.), *Non-stationary Time-series Analysis and Cointegration*, pp. 9–52. Oxford: Oxford University Press.
- Clements, M. P., and Hendry, D. F. (1995). Macro-economic forecasting and modelling. *Economic Journal*, **105**, 1001–1013.
- Clements, M. P., and Hendry, D. F. (1996). Intercept corrections and structural breaks. *Journal of Applied Econometrics*, **11**, 475–494.
- Clements, M. P., and Hendry, D. F. (1997a). Forecasting economic processes. mimeo, Institute of Economics and Statistics, University of Oxford. Forthcoming, *International Journal of Forecasting*.
- Clements, M. P., and Hendry, D. F. (1997b). Macroeconomic forecasting and structural change. mimeo, Institute of Economics and Statistics, University of Oxford.
- Cook, S. (1995). Treasury economic forecasting. mimeo, Institute of Economics and Statistics, University of Oxford.
- Cooper, J. P., and Nelson, C. R. (1975). The ex ante prediction performance of the St. Louis and FRB-MIT-PENN econometric models and some results on composite predictors. *Journal of Money, Credit, and Banking*, **7**, 1–32.
- Corsi, P., Pollock, R. E., and Prakken, J. C. (1982). The chow test in the presence of serially correlated errors. In Chow, G. C., and Corsi, P. (eds.), *Evaluating the Reliability of Macro-Economic Models*. New York: John Wiley.
- 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. (1996). *Object-Oriented Matrix Programming using Ox*. London: International Thomson Business Press and Oxford: <http://www.nuff.ox.ac.uk/Users/Doornik/>.
- Doornik, J. A., and Hendry, D. F. (1997). Monte Carlo simulation using PcNaive for Windows. Unpublished typescript, Nuffield College, University of Oxford.
- Duesenberry, J. S. (1949). *Income, Saving and the Theory of Consumer Behaviour*. Cambridge: Harvard University Press.
- Engle, R. F., and Hendry, D. F. (1993). Testing super exogeneity and invariance in regression models. *Journal of Econometrics*, **56**, 119–139.
- Ericsson, N., Hendry, D., and Prestwich, K. (1997). The demand for broad money in the United Kingdom, 1878–1993. mimeo, Nuffield College, Oxford. forthcoming, *Scandinavian Journal of Economics*.
- Favero, C., and Hendry, D. F. (1992). Testing the Lucas critique: A review. *Econometric Reviews*, **11**, 265–306.
- Godfrey, L. G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica*, **46**, 1303–1313.
- Harvey, A. C., and Shephard, N. (1992). Structural time series models. In Maddala, G. S., Rao, C. R., and Vinod, H. D. (eds.), *Handbook of Statistics*, Vol. 11. Amsterdam: North-Holland.
- Hendry, D. F. (1979). Predictive failure and econometric modelling in macro-economics: The transactions demand for money. In Ormerod, P. (ed.), *Economic Modelling*, pp. 217–242. London: Heinemann. Reprinted in Hendry, D. F. (1993), *Econometrics: Alchemy or Science?* Oxford:

Blackwell Publishers.

- Hendry, D. F. (1984). Econometric modelling of house prices in the United Kingdom. In Hendry, D. F., and Wallis, K. F. (eds.), *Econometrics and Quantitative Economics*, pp. 135–172. Oxford: Basil Blackwell.
- Hendry, D. F. (1994). HUS revisited. *Oxford Review of Economic Policy*, **10**, 86–106.
- Hendry, D. F. (1995a). *Dynamic Econometrics*. Oxford: Oxford University Press.
- Hendry, D. F. (1995b). A theory of co-breaking. Mimeo, Nuffield College, University of Oxford.
- Hendry, D. F. (1996a). The econometrics of macro-economic forecasting. *Economic Journal*. Forthcoming.
- Hendry, D. F. (1996b). On the constancy of time-series econometric equations. *Economic and Social Review*, **27**, 401–422.
- Hendry, D. F., and Mizon, G. E. (1993). Evaluating dynamic econometric models by encompassing the VAR. In Phillips, P. C. B. (ed.), *Models, Methods and Applications of Econometrics*, pp. 272–300. Oxford: Basil Blackwell.
- Hendry, D. F., and Mizon, G. E. (1996). The influence of A. W. H. Phillips on econometrics. In Leeson, R. (ed.), *A. W. H. Phillips: Collected Works in Contemporary Perspective*. Cambridge: Cambridge University Press. Forthcoming.
- Hendry, D. F., and Morgan, M. S. (1995). *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press.
- Hendry, D. F., and Neale, A. J. (1991). A Monte Carlo study of the effects of structural breaks on tests for unit roots. In Hackl, P., and Westlund, A. H. (eds.), *Economic Structural Change, Analysis and Forecasting*, pp. 95–119. Berlin: Springer-Verlag.
- Hoover, K. D., and Perez, S. J. (1996). Data mining reconsidered: Encompassing and the general-to-specific approach to specification search. Mimeo, Economics department, University of California, Davis.
- Hylleberg, S., and Mizon, G. E. (1989). Cointegration and error correction mechanisms. *Economic Journal*, **99**, 113–125. Supplement.
- Kiviet, J. F. (1986). On the rigor of some mis-specification tests for modelling dynamic relationships. *Review of Economic Studies*, **53**, 241–261.
- Marget, A. W. (1929). Morgenstern on the methodology of economic forecasting. *Journal of Political Economy*, **37**, 312–339.
- Morgenstern, O. (1928). *Wirtschaftsprognose: eine Untersuchung ihrer Voraussetzungen und Möglichkeiten*. Vienna: Julius Springer.
- Muellbauer, J. N. J. (1994). The assessment: Consumer expenditure. *Oxford Review of Economic Policy*, **10**, 1–41.
- Persons, W. M. (1924). *The Problem of Business Forecasting*. No. 6 in Pollak Foundation for Economic Research Publications. London: Pitman.
- Richard, J.-F., and Zhang, W. (1996). Econometric modelling of UK house prices using accelerated importance sampling. *Oxford Bulletin of Economics and Statistics*, **58**, 601–613.