

On Winning Forecasting Competitions in Economics

Michael P. Clements*
Economics Department,
Warwick University.

and

David F. Hendry†
Nuffield College,
Oxford.

March 30, 2008

Abstract

To explain which methods might win forecasting competitions on economic time series, we consider forecasting in an evolving economy subject to structural breaks, using mis-specified, data-based models. ‘Causal’ models need not win when facing deterministic shifts, a primary factor underlying systematic forecast failure. We derive conditional forecast biases and unconditional (asymptotic) variances to show that when the forecast evaluation sample includes sub-periods following breaks, non-causal models will outperform at short horizons. This suggests using techniques which avoid systematic forecasting errors, including improved intercept corrections. An application to a small monetary model of the UK illustrates the theory.

1 Introduction

Modern economies are often subject to major institutional, political, financial, and technological changes which manifest themselves as structural breaks in econometric models relative to the underlying data generation process. Consequently, over any long forecast-evaluation horizon in economics, a break after forecasts are announced is a likely event. We know of no methods that fare well in such a setting, and beyond pre-cognition of the sign, size, timing, and location of the break, none seem possible. Conversely, any forecaster operating for a prolonged period will experience episodes when forecasts are made after breaks have occurred, and some methods are more robust than others in that situation. The usual forecast-evaluation scenario reserves H observations at the end of an estimation sample, and often also divides H into several shorter-horizon tests, such as k periods of length H/k . If any breaks occur prior to the start of the final evaluation period, then they must have occurred before some of the forecasts are calculated. Robust methods should perform well over those periods, and so the rankings of methods not only depend on their intrinsic properties, but also on the presence or absence of breaks over the evaluation periods used, and on the horizons forecasted (e.g., 1-step or 4-steps ahead).

In an earlier paper (Clements and Hendry, 1995), assuming constant parameters, we obtained formulae for unconditional forecast-error biases and asymptotic variances for a vector equilibrium-correction model (VEqCM) and a vector autoregression in differences (DV) considering levels, differences and cointegrating combinations. We also showed how these differed depending on the transformation inspected. We have also derived unconditional forecast-error biases for systems with breaks (see Hendry and Clements, 1998), and note these below. Here, we investigate conditional biases, and unconditional

*All correspondence should be addressed to the first author.

†Financial support from the U.K. Economic and Social Research Council under grant L116251015 is gratefully acknowledged by both authors. We are indebted to participants in seminars at the Bank of Norway, Bocconi University, Ekonometrídagarna, European University Institute, Stockholm School of Economics, and an anonymous referee for helpful comments on an earlier version.

variances for 1- and 2-step ahead forecasts for a VEqCM, DV, and also for a second-differenced system (DDV), both before and after deterministic breaks, so that a better appraisal of likely mean squared forecast errors (MSFEs) is possible. These three models are differentially susceptible to deterministic shifts, and hence make a useful comparative set. The formulae presented below are only for the levels of the variables, but specialize to our previous results when no breaks occur.

The structure of the paper is as follows. Section 2 describes the data generation process (DGP) used as the basis for the analytic calculations. It also sets out the forecasting models, since we do not assume that the models and DGP coincide. Then §3 reconsiders deterministic shifts in forecast failure, following the evidence in Hendry and Doornik (1997) of their role in dynamic, cointegrated, systems. Next, §4 derives the conditional biases of the three forecasting models, and §5 the (asymptotic) variances, in each case both before and after structural breaks in the deterministic terms. Then §5.7 discusses the biases that result in the variance formulae programmed under the assumption of white-noise errors. Section 6 provides an extensive empirical illustration of the theory using a four-equation monetary model based on UK M1, and §7 undertakes forecast comparisons. Then §8 considers intercept corrections and §9 discusses empirical forecast accuracy comparisons. Section 10 concludes.

2 The data generation process and forecasting models

Let \mathbf{x}_t be an $n \times 1$ vector of time-series variables, satisfying the first-order dynamic linear system for $t = 1, 2, \dots, T$:

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

where $\mathbf{v}_t \sim \text{IN}_n[\mathbf{0}, \Omega]$. In (1), the initial value \mathbf{x}_0 is fixed, Υ is an $n \times n$ matrix of coefficients, and $\boldsymbol{\tau}$ an $n \times 1$ vector of intercepts. When $\mathbf{x}_t \sim \text{I}(1)$ and the cointegrating rank is r , (1) can be reparameterized as:

$$\Delta \mathbf{x}_t = \Pi \mathbf{x}_{t-1} + \boldsymbol{\tau} + \mathbf{v}_t \quad (2)$$

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

$$\Delta \mathbf{x}_t = \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{x}_{t-1} + \boldsymbol{\tau} + \mathbf{v}_t. \quad (3)$$

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

Let the expectation of $\Delta \mathbf{x}_t$ be $\boldsymbol{\gamma}$ ($n \times 1$), which defines the growth in the system, and let $E[\boldsymbol{\beta}' \mathbf{x}_t] = \boldsymbol{\mu}$, which is $r \times 1$, then from (3):

$$E[\boldsymbol{\beta}' \mathbf{x}_t] = (\mathbf{I}_r + \boldsymbol{\beta}' \boldsymbol{\alpha}) E[\boldsymbol{\beta}' \mathbf{x}_{t-1}] + \boldsymbol{\beta}' \boldsymbol{\tau} = \boldsymbol{\mu},$$

so that $\boldsymbol{\mu} = -(\boldsymbol{\beta}' \boldsymbol{\alpha})^{-1} \boldsymbol{\beta}' \boldsymbol{\tau}$, and hence:

$$\begin{aligned} E[\Delta \mathbf{x}_t] &= E[\boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{x}_{t-1} + \boldsymbol{\tau} + \mathbf{v}_t] = \boldsymbol{\alpha} E[\boldsymbol{\beta}' \mathbf{x}_{t-1}] + \boldsymbol{\tau} \\ &= [\mathbf{I}_n - \boldsymbol{\alpha} (\boldsymbol{\beta}' \boldsymbol{\alpha})^{-1} \boldsymbol{\beta}'] \boldsymbol{\tau} = \mathbf{K} \boldsymbol{\tau} = \boldsymbol{\gamma}. \end{aligned} \quad (4)$$

Thus, $\boldsymbol{\tau} = \boldsymbol{\gamma} - \boldsymbol{\alpha} \boldsymbol{\mu}$ (see Johansen and Juselius, 1990). The matrix \mathbf{K} is non-symmetric but idempotent with $\boldsymbol{\beta}' \mathbf{K} = \mathbf{0}'$ and $\mathbf{K} \boldsymbol{\alpha} = \mathbf{0}$, so $\Upsilon \mathbf{K} = \mathbf{K}$ and $\boldsymbol{\beta}' \boldsymbol{\gamma} = \mathbf{0}$. Consequently, the DGP becomes:

$$\Delta \mathbf{x}_t = \boldsymbol{\gamma} + \boldsymbol{\alpha} (\boldsymbol{\beta}' \mathbf{x}_{t-1} - \boldsymbol{\mu}) + \mathbf{v}_t. \quad (5)$$

The first model is the VEqCM which matches the DGP (5), but will not coincide with the DGP in the forecast period when there is structural change. The second model is a VAR in the differences of the variables (DV):

$$\Delta \mathbf{x}_t = \gamma + \boldsymbol{\xi}_t, \quad (6)$$

so is correctly specified only when $\alpha = \mathbf{0}$ in (5), in which case $\boldsymbol{\xi}_t = \mathbf{v}_t$. Otherwise, it is mis-specified by omitting the cointegrating vectors, and hence omits causal information. The third ‘model’ we examine is a DV in the differences of the variables (DDV), defined by:

$$\Delta^2 \mathbf{x}_t = \boldsymbol{\zeta}_t \text{ or } \mathbf{x}_t = \mathbf{x}_{t-1} + \Delta \mathbf{x}_{t-1} + \boldsymbol{\zeta}_t. \quad (7)$$

These models are convenient for the analytic calculations, but are generalized in an obvious manner to allow for longer lag structures in the empirical work below.

3 Determinants of forecast failure

The analysis in Hendry and Doornik (1997) suggested that the primary determinant of a model experiencing forecast failure – defined as a significant deterioration in its forecast performance relative to earlier fit – was a deterministic shift, namely a change in the unconditional mean of the stationary component. Such a change could result from an intercept shift, a trend change (perhaps due to incorrectly omitting a trend term), changes in other parameters of included variables that shifted the equilibrium mean relative to its sample value, or changes in the means of incorrectly excluded variables that changed. In each instance, the key factor was the induced mean shift, as illustrated below for changes that do not involve deterministic components (as in zero-mean processes). This section seeks to clarify the basis for that view.

Clearly, any sufficiently large and prolonged change in the data relative to a model will induce forecast failure. Most tests of constancy are based on comparisons between in-sample and post-sample residual sums of squares, and any event that greatly increases the latter must be detectable. However, the force of the focus on deterministic shifts is that the mean value of the variable experiencing the parameter change is a major determinant of the detectability of the break for parameter shifts of a (percentage) size that seem likely in economics. This is illustrated in fig.1 which records Monte Carlo rejection frequencies of sequences of break-point Chow (1960) tests for a shift in a parameter ϕ_1 at time $T_1 = 55$ when the sample size is $T = 60$ in the model:

$$y_t = \phi_0 + \phi_1 z_t + \epsilon_t \text{ where } \epsilon_t \sim \text{IN} [0, \sigma_\epsilon^2], \quad (8)$$

when the explanatory variable $z_t \sim \text{IN}[\eta, \sigma_z^2]$ with $\phi_0 = 0$, $\sigma_z^2 = 1$ and $\sigma_\epsilon^2 = 1$. The parameter η takes the values 0, 1, 2, and 4, and we consider breaks in ϕ_1 to $\phi_1 + \delta$ where $\delta = \rho\phi_1$ when $\rho = \pm 0.25$, ± 0.50 and ± 1.00 from a baseline of $\phi_1 = 1$. When $\eta \neq 0$, changes in ϕ_1 induce shifts in $\text{E}[y_t]$ from $\phi_1\eta$ to $(\phi_1 + \delta)\eta$.

For a known breakpoint at time $T_1 = 55$, based on the appendix derivations, the expectation of the Chow test is (setting $k = T^{-1}T_2$ where $T_2 = T - T_1$):

$$\text{E} \left[F_{T_1-1}^{T_2+1} \right] = \text{E} \left[\frac{RSS_T - RSS_{T_1}}{RSS_{T_1}} \times \frac{T_1 - 1}{T_2 + 1} \right] \simeq \frac{[\sigma_\epsilon^2 + \delta^2 (\sigma_z^2 + \eta^2) (1 - k)] T_2}{(T_2 + 1) \sigma_\epsilon^2}.$$

Since $\text{E}[F_{T_1-1}^{T_2+1}] = (T_2 + 1)/(T_2 - 1)$, the non-centrality is approximately $\rho^2 (1 + \eta^2) (1 - k)$, and hence depends on the squared mean of z as well as on the squared size of the break, both relative to the

equation standard deviation. Although data means are often arbitrary for economic time series, such as indices, or nominal magnitudes, the detectability of breaks depends on the scaling, as fig.1 shows (200 replications were used).¹

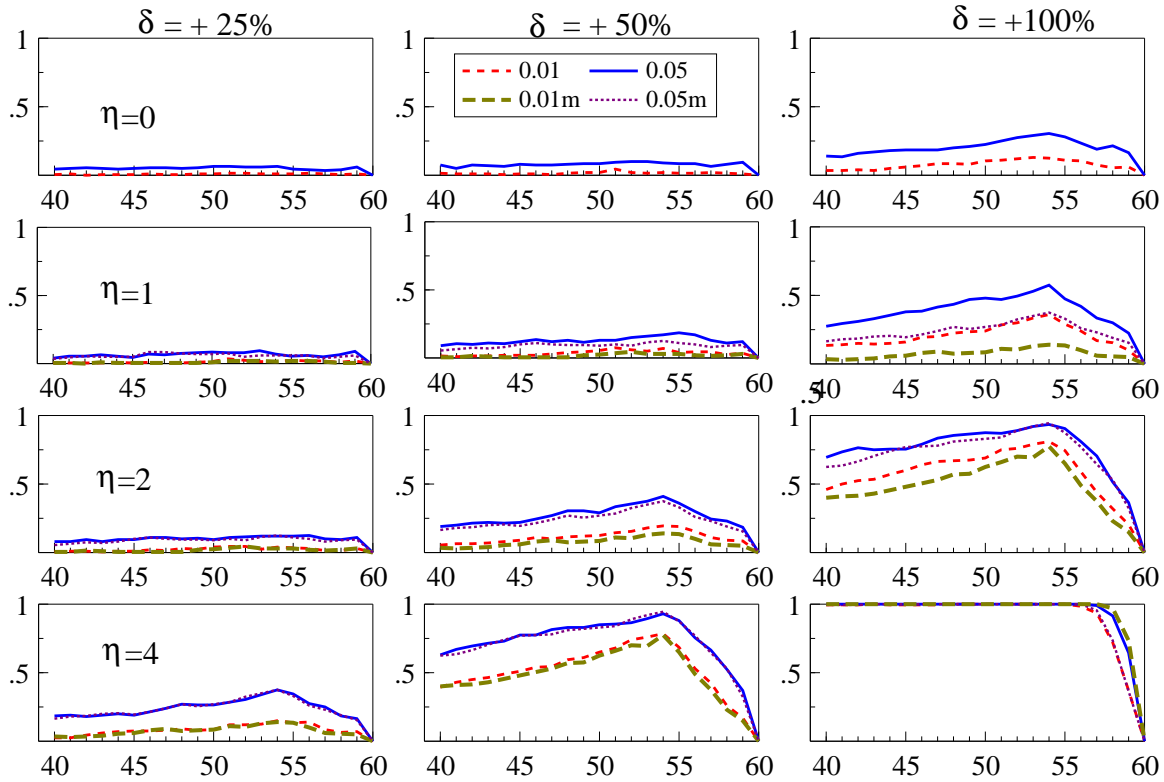


Figure 1 Recursively-computed Chow test rejection frequencies.

Only the outcomes for upward shifts are shown, as the results were closely similar for the downward shifts: both 5% and 1% decisions are recorded. The figures also show the rejection frequencies for shifts in $E[y_t]$ equal to $\delta\eta$, but with ϕ_1 held constant (denoted 0.05m and 0.01m respectively). These are achieved by changing ϕ_0 . The marked impact of increasing η on rejection frequencies for a given shift in ϕ_1 is clear from reading down the columns. Sufficiently large changes in zero-mean processes can be detected (see the first row), but the magnitude of the required change is large relative to that likely to be encountered in practice. There is a close correspondence between rejections for shifts in ϕ_1 and those in $E[y_t]$ alone, so that changes in parameters (ϕ_1) and changes in deterministic components (ϕ_0) that have a similar effect on the mean have similar rejection rates. We expect changes in economic parameters to be small relative to base-parameter values, making the first two columns of the figure most relevant. Hendry and Doornik (1997) work through a similar analysis but with dynamic models, and find that it is shifts in long-run means that are detectable and responsible for forecast failure. Deterministic shifts would then appear to be a primary cause of forecast failure, and are the focus of the analysis below.

4 Conditional biases

When a break occurs, its timing matters greatly relative to the forecast origin, and we consider forecasts made both pre and post breaks to mimic the real-time nature of economic forecasting where eventually, a

¹In figures, 2×2 panels of graphs are notionally labelled a, b, c, d , such that $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ denotes the location; 3×3 panels as $a, b, c, d, e, f, g, h, i$; and so on.

forecast will be made after a break. Throughout, we assume the parameter values are known; estimation would add sampling uncertainty, and model selection could introduce additional biases. We also ignore model revision in the light of forecast failure: to the extent that such occurs correctly, it will bring the model back to the DGP and hence its forecasts to those made when no change occurs. There could be changes in any or all of the parameters of (5), namely α , β , γ , or μ , but here β will be held constant. We consider the effects of all the other parameter changes on the VEqCM, DV and DDV, both 1-step and 2-periods ahead.

The 1-step forecasts of the levels at forecast origin time $T + i$ ($i = 0, 1$) from the VEqCM (5) are:

$$\Delta \hat{\mathbf{x}}_{T+i+1|T+i} = \gamma + \alpha (\beta' \mathbf{x}_{T+i} - \mu) \quad (9)$$

whereas from the DV (6):

$$\Delta \tilde{\mathbf{x}}_{T+i+1|T+i} = \gamma, \quad (10)$$

and the DDV (7):

$$\Delta \bar{\mathbf{x}}_{T+i+1|T+i} = \Delta \mathbf{x}_{T+i}. \quad (11)$$

The 1-step forecast errors of the three models in terms of the levels are denoted by $\hat{\mathbf{e}}_{T+i+1|T+i} = \mathbf{x}_{T+i+1} - \hat{\mathbf{x}}_{T+i+1|T+i}$, $\tilde{\mathbf{e}}_{T+i+1|T+i} = \mathbf{x}_{T+i+1} - \tilde{\mathbf{x}}_{T+i+1|T+i}$, and $\bar{\mathbf{e}}_{T+i+1|T+i} = \mathbf{x}_{T+i+1} - \bar{\mathbf{x}}_{T+i+1|T+i}$. The corresponding 2-step forecasts are:

$$\hat{\mathbf{x}}_{T+i+2|T+i} = \Upsilon \hat{\mathbf{x}}_{T+i+1|T+i} + \gamma - \alpha \mu = \Upsilon^2 \mathbf{x}_{T+i} + (\mathbf{I}_n + \Upsilon) (\gamma - \alpha \mu); \quad (12)$$

$$\tilde{\mathbf{x}}_{T+i+2|T+i} = \mathbf{x}_{T+i} + 2\gamma; \quad (13)$$

and:

$$\bar{\mathbf{x}}_{T+i+2|T+i} = \bar{\mathbf{x}}_{T+i+1|T+i} + \Delta \bar{\mathbf{x}}_{T+i+1|T+i} = \mathbf{x}_{T+i} + 2\Delta \mathbf{x}_{T+i}, \quad (14)$$

from (11). Similar expressions hold for their forecast errors.

After time T , the data are in fact generated by:

$$\Delta \mathbf{x}_{T+1} = \gamma^* + \alpha^* (\beta' \mathbf{x}_T - \mu^*) + \nu_{T+1} \quad (15)$$

so we consider forecasting from T to $T + 1$; T to $T + 2$; $T + 1$ to $T + 2$; and $T + 1$ to $T + 3$: the last two are forecasts prepared after the break, but still with no knowledge of the break having occurred. Changes in the feedback parameters are retained to clarify their role in affecting forecast variances. The detailed calculations are shown in Appendix B (§12); here we focus on the resulting formulae. We distinguish unconditional and conditional forecast biases, and additional variance effects; as the former are analyzed in Clements and Hendry (1996) and Hendry and Clements (1998), here we derive the *conditional* biases, and the asymptotic variance formulae.

4.1 Forecasting from T to $T + 1$

Let $\beta' \mathbf{x}_T - \mu = \kappa_T$, then:

$$\mathbf{E} [\hat{\mathbf{e}}_{T+1|T} | I_T] = (\gamma^* - \gamma) - \alpha^* (\mu^* - \mu) + (\alpha^* - \alpha) \kappa_T, \quad (16)$$

$$\mathbf{E} [\tilde{\mathbf{e}}_{T+1|T} | I_T] = (\gamma^* - \gamma) - \alpha^* (\mu^* - \mu) + \alpha^* \kappa_T, \quad (17)$$

and:

$$\mathbf{E} [\bar{\mathbf{e}}_{T+1|T} | I_T] = (\gamma^* - \Delta \mathbf{x}_T) - \alpha^* (\mu^* - \mu) + \alpha^* \kappa_T. \quad (18)$$

Since $\mathbf{E}[\kappa_T] = \mathbf{0}$ and $\mathbf{E}[\Delta \mathbf{x}_T] = \gamma$, the three forecasting procedures deliver identical unconditional biases for forecasting from T to $T + 1$ when an unknown break occurs at T after the forecast is made. The conditional biases only differ to the extent that $\beta' \mathbf{x}_T \neq \mu$ and $\Delta \mathbf{x}_T \neq \gamma$, noting that $\Delta \mathbf{x}_T = \gamma + \alpha \kappa_{T-1} + \nu_T$. When no break occurs, (16) is zero, whereas both (17) and (18) remain non-zero.

4.2 Forecasting from $T + 1$ to $T + 2$

Having reached time $T + 1$, with data at $T + 1$ assumed available, the outcomes are remarkably different: the three predictors now deliver distinct outcomes. As the unconditional expectations of the levels are altering along the adjustment path, and must be time dated due to the non-stationarity induced by the breaks, all terms have to be tracked. First, and assuming that no additional changes occur at time $T + 1$, the DGP produces:

$$\Delta \mathbf{x}_{T+2} = \gamma^* + \alpha^* (\beta' \mathbf{x}_{T+1} - \mu^*) + \nu_{T+2} \quad (19)$$

and as:

$$\mathbf{x}_{T+1} = \mathbf{x}_T + \gamma^* + \alpha^* (\beta' \mathbf{x}_T - \mu^*) + \nu_{T+1},$$

then:

$$\beta' \mathbf{x}_{T+1} - \mu^* = \Lambda^* (\beta' \mathbf{x}_T - \mu^*) + \beta' \nu_{T+1}, \quad (20)$$

where $\Lambda^* = (\mathbf{I}_r + \beta' \alpha^*)$ has all its eigenvalues inside the unit circle. Further, let $\Upsilon^* = (\mathbf{I}_n + \alpha^* \beta')$, and note that $\alpha^* \Lambda^* = \Upsilon^* \alpha^*$. To avoid dependence of the new growth rate on the changed parameters, we maintain $\beta' \gamma^* = 0$ and set $\beta' \mathbf{x}_{T+1} - \mu^* = \kappa_{T+1}^*$ so from (20):

$$\kappa_{T+1}^* = \Lambda^* (\kappa_T - [\mu^* - \mu]) + \beta' \nu_{T+1}.$$

Finally:

$$\mathbb{E} [\Delta \mathbf{x}_{T+h}] = \gamma^* - \alpha^* (\Lambda^*)^{h-1} (\mu^* - \mu) \xrightarrow{h \rightarrow \infty} \gamma^*$$

The conditional biases of the three 1-step forecast errors are therefore:

$$\mathbb{E} [\hat{\mathbf{e}}_{T+2|T+1} | I_{T+1}] = (\gamma^* - \gamma) - \alpha (\mu^* - \mu) + (\alpha^* - \alpha) \kappa_{T+1}^* \quad (21)$$

$$\mathbb{E} [\tilde{\mathbf{e}}_{T+2|T+1} | I_{T+1}] = (\gamma^* - \gamma) + \alpha^* \kappa_{T+1}^* \quad (22)$$

and:

$$\mathbb{E} [\bar{\mathbf{e}}_{T+2|T+1} | I_{T+1}] = (\gamma^* - \Delta \mathbf{x}_{T+1}) + \alpha^* \kappa_{T+1}^*, \quad (23)$$

where now $\mathbb{E} [\Delta \mathbf{x}_{T+1}] = \gamma^* - \alpha^* (\mu^* - \mu)$. Consequently, the first and second differ depending on the relative size of $(\alpha^* - \alpha)$ compared to α^* , but the third shows considerable ‘error-correction’ behaviour at this post-break 1-step ahead horizon, completely removing the effect of $\mathbb{E} [\bar{\mathbf{e}}_{T+1|T}]$.

In particular, when $\alpha^* = \alpha$, so only deterministic terms change, then:

$$\begin{aligned} \mathbb{E} [\hat{\mathbf{e}}_{T+2|T+1} | I_{T+1}] &= (\gamma^* - \gamma) - \alpha (\mu^* - \mu) \\ \mathbb{E} [\tilde{\mathbf{e}}_{T+2|T+1} | I_{T+1}] &= (\gamma^* - \gamma) + \alpha \kappa_{T+1}^* \\ \mathbb{E} [\bar{\mathbf{e}}_{T+2|T+1} | I_{T+1}] &= (\gamma^* - \Delta \mathbf{x}_{T+1}) + \alpha \kappa_{T+1}^*. \end{aligned} \quad (24)$$

These three expressions highlight the effects of first and second differencing relative to the VEqCM, since:

$$\mathbb{E} [\kappa_{T+1}^*] = -\Lambda^* (\mu^* - \mu).$$

Consequently, κ_{T+1}^* on average is ‘smaller than’ $(\mu^* - \mu)$ and so induces less bias. Further, as:

$$\mathbb{E} [\Delta \mathbf{x}_{T+1}] = \gamma^* - \alpha^* (\mu^* - \mu),$$

in this case, then second differencing removes the impact of the changing growth rate, and as $\beta' \alpha$ corresponds to the roots inside the unit circle, dampens the effect of shifts in the equilibrium mean. The DV has not avoided the impact of changes in growth, and only partly attenuated the effect of shifts in μ : this is because when a VEqCM is the DGP, the DV arises by dropping equilibrium-correction terms, not by first differencing.

4.3 Forecasting from T to $T + 2$

This comprises 2-steps ahead forecasts made prior to the break using (12), (13), and (14), to illustrate how the results change for multi-step forecasts. Now the corresponding 2-step ahead conditional forecast-error biases are:

$$E [\hat{e}_{T+2|T} | I_T] = E [\hat{e}_{T+2|T}] + (\alpha^* [I_r + \Lambda^*] - \alpha [I_r + \Lambda]) \kappa_T \quad (25)$$

$$E [\tilde{e}_{T+2|T} | I_T] = E [\tilde{e}_{T+2|T}] + \alpha^* (I_r + \Lambda^*) \kappa_T \quad (26)$$

$$E [\bar{e}_{T+2|T} | I_T] = E [\bar{e}_{T+2|T}] + \alpha^* (I_r + \Lambda^*) \kappa_T - 2\alpha \kappa_{T-1} \quad (27)$$

where, for $\check{e} = \hat{e}$, \tilde{e} , and \bar{e} , the unconditional biases are:

$$E [\check{e}_{T+2|T}] = 2(\gamma^* - \gamma) - \alpha^* (I_r + \Lambda^*) (\mu^* - \mu) \quad (28)$$

Thus, all three unconditional biases suffer from the same impact after a change in the equilibrium mean or the growth rate. If α is constant (so $\Lambda^* = \Lambda$ also), (25) simplifies to $E [\hat{e}_{T+2|T}]$ which appears to favour VEqCM over DV and DDV when the economy is in considerable disequilibrium. Otherwise, the conditional biases deviate from the unconditional by amounts that depend on past disequilibria.

4.4 Forecasting from $T + 1$ to $T + 3$

First, the forecasting rules are just time-index shifted versions of (12)–(14). Again, the expectations are changing as adjustments to the new parameters continue, and the feature of practical importance lies in the changed expectations of the variables, $\beta'x_{T+1}$ and Δx_{T+1} . Thus:

$$\begin{aligned} E [\hat{e}_{T+3|T+1} | I_{T+1}] &= 2(\gamma^* - \gamma) - \alpha (I_r + \Lambda) (\mu^* - \mu) + [\alpha^* (I_r + \Lambda^*) - \alpha (I_r + \Lambda)] \kappa_{T+1}^* \\ &= E [\hat{e}_{T+3|T+1}] + [\alpha^* (I_r + \Lambda^*) - \alpha (I_r + \Lambda)] (\kappa_{T+1}^* - E [\kappa_{T+1}^*]) \end{aligned}$$

where:

$$E [\hat{e}_{T+3|T+1}] = 2(\gamma^* - \gamma) - [\alpha^* (I_r + \Lambda^*) \Lambda^* - \alpha (I_r + \Lambda) \beta' \alpha^*] (\mu^* - \mu).$$

In the special case that α is constant:

$$E [\hat{e}_{T+3|T+1} | I_{T+1}] = E [\hat{e}_{T+3|T+1}] = 2(\gamma^* - \gamma) - \alpha (I_r + \Lambda) (\mu^* - \mu). \quad (29)$$

Next, for the DV:

$$\begin{aligned} E [\tilde{e}_{T+3|T+1} | I_{T+1}] &= 2(\gamma^* - \gamma) + \alpha^* (I_r + \Lambda^*) \kappa_{T+1}^* \\ &= E [\tilde{e}_{T+3|T+1}] + \alpha^* (I_r + \Lambda^*) (\kappa_{T+1}^* - E [\kappa_{T+1}^*]) \end{aligned}$$

where:

$$E [\tilde{e}_{T+3|T+1}] = 2(\gamma^* - \gamma) - \alpha^* \Lambda^* (I_r + \Lambda^*) (\mu^* - \mu),$$

so, if α is constant:

$$E [\tilde{e}_{T+3|T+1} | I_{T+1}] = 2(\gamma^* - \gamma) + \alpha (I_r + \Lambda) \kappa_{T+1}^*. \quad (30)$$

This differs from the VEqCM bias, but is smaller on average, as:

$$E [\kappa_{T+1}^*] = -\Lambda^* (\mu^* - \mu).$$

Finally, the DDV conditional forecast-error bias is:

$$\begin{aligned} E [\bar{\mathbf{e}}_{T+3|T+1} | I_{T+1}] &= 2(\gamma^* - \Delta \mathbf{x}_{T+1}) + \alpha^* (\mathbf{I}_r + \Lambda^*) \kappa_{T+1}^* \\ &= 2\alpha^* (\mu^* - \mu) + 2(E[\Delta \mathbf{x}_{T+1}] - \Delta \mathbf{x}_{T+1}) + \alpha^* (\mathbf{I}_r + \Lambda^*) \kappa_{T+1}^* \\ &= E[\bar{\mathbf{e}}_{T+3|T+1}] + 2(E[\Delta \mathbf{x}_{T+1}] - \Delta \mathbf{x}_{T+1}) + \alpha^* (\mathbf{I}_r + \Lambda^*) (\kappa_{T+1}^* - E[\kappa_{T+1}^*]), \end{aligned}$$

where:

$$E[\bar{\mathbf{e}}_{T+3|T+1}] = -\alpha^* (2\mathbf{I}_r + \Lambda^*) \beta' \alpha^* (\mu^* - \mu),$$

noting κ_{T+1}^* is ‘smaller’ than $(\mu^* - \mu)$ on average. Consequently, if α is constant:

$$E[\bar{\mathbf{e}}_{T+3|T+1}] = -\alpha (2\mathbf{I}_r + \Lambda) \beta' \alpha (\mu^* - \mu). \quad (31)$$

Hence the DDV avoids breaks in γ , and differs in its reaction to changes in μ .

5 Variance effects after structural breaks

Forecast-error biases must be evaluated against the additional forecast-error variances that the non-congruent models introduce, and this section analyzes these. The formulae take account of the information that is averaged over when unconditional biases are computed: such terms appear here as variance effects (e.g., the difference between $\beta' \mathbf{x}_T$ and μ). The unconditional forecast-error variance formulae derived in this section foreshadow the findings of the empirical forecast-accuracy comparison in §9. The variances and biases recorded there are unconditional in that they are formed by averaging over sixteen 1-step or thirteen 4-step ahead forecasts. Formulae for the s -step ahead asymptotic variances of VEqCM and DV in levels, differences and $l(0)$ transformations, when there are no breaks, are provided in Clements and Hendry (1995); here we focus on the results for levels.

First:

$$(\beta' \mathbf{x}_T - \mu) = \Lambda (\beta' \mathbf{x}_{T-1} - \mu) + \beta' \nu_T$$

and hence:

$$\mathbf{V}[\beta' \mathbf{x}_T] - \Lambda \mathbf{V}[\beta' \mathbf{x}_T] \Lambda' = \beta' \Omega \beta$$

or:

$$[(\mathbf{V}[\beta' \mathbf{x}_T])^v]^{sm} = \mathbf{V}[\beta' \mathbf{x}_T] = [(\mathbf{I}_r - \Lambda \otimes \Lambda)^{-1} (\beta' \Omega \beta)^v]^{sm} = \mathbf{V},$$

where $[\mathbf{A}(\mathbf{B})^v]^{sm}$ denotes the symmetric matrix formed from a column created by vectoring, denoted $(\cdot)^v$. Also:

$$\mathbf{V}[\Delta \mathbf{x}_T] = \alpha \mathbf{V} \alpha' + \Omega.$$

Analogous expressions hold for the variance matrices after the break. We can now deduce the corresponding outcomes for each of the four different combinations of forecast origins and horizons.

5.1 Variances from T to $T + 1$

When the parameters are known, a setting that favours the VEqCM, the 1-step forecast errors of the three models can be written as:

$$\hat{\mathbf{e}}_{T+1|T} - E[\hat{\mathbf{e}}_{T+1|T}] = (\alpha^* - \alpha) (\beta' \mathbf{x}_T - \mu) + \nu_{T+1}, \quad (32)$$

$$\tilde{\mathbf{e}}_{T+1|T} - E[\tilde{\mathbf{e}}_{T+1|T}] = \alpha^* (\beta' \mathbf{x}_T - \mu) + \nu_{T+1}, \quad (33)$$

and :

$$\bar{\mathbf{e}}_{T+1|T} - \mathbf{E} [\bar{\mathbf{e}}_{T+1|T}] = \boldsymbol{\alpha}^* (\boldsymbol{\beta}' \mathbf{x}_T - \boldsymbol{\mu}) + \boldsymbol{\nu}_{T+1} - (\Delta \mathbf{x}_T - \boldsymbol{\gamma}). \quad (34)$$

Conditionally, these have the same variance components, i.e.:

$$\begin{aligned} \mathbf{V} [\hat{\mathbf{e}}_{T+1|T} | \mathcal{I}_T] &\equiv \mathbf{E} \left[(\hat{\mathbf{e}}_{T+1|T} - \mathbf{E} [\hat{\mathbf{e}}_{T+1|T} | \mathcal{I}_T]) (\hat{\mathbf{e}}_{T+1|T} - \mathbf{E} [\hat{\mathbf{e}}_{T+1|T} | \mathcal{I}_T])' | \mathcal{I}_T \right] \\ &= \mathbf{V} [\tilde{\mathbf{e}}_{T+1|T} | \mathcal{I}_T] = \mathbf{V} [\bar{\mathbf{e}}_{T+1|T} | \mathcal{I}_T] \end{aligned}$$

Given earlier analyses, and to help interpret the empirical forecast-error variances, we record the unconditional variances:

$$\mathbf{V} [\hat{\mathbf{e}}_{T+1|T}] = \boldsymbol{\Omega} + (\boldsymbol{\alpha}^* - \boldsymbol{\alpha}) \mathbf{V} (\boldsymbol{\alpha}^* - \boldsymbol{\alpha})', \quad (35)$$

$$\mathbf{V} [\tilde{\mathbf{e}}_{T+1|T}] = \boldsymbol{\Omega} + \boldsymbol{\alpha}^* \mathbf{V} \boldsymbol{\alpha}^{*'}, \quad (36)$$

and:

$$\mathbf{V} [\bar{\mathbf{e}}_{T+1|T}] = [\boldsymbol{\alpha}^* \boldsymbol{\Lambda} - \boldsymbol{\alpha}] \mathbf{V} [\boldsymbol{\alpha}^* \boldsymbol{\Lambda} - \boldsymbol{\alpha}]' + \boldsymbol{\Omega} + (\mathbf{I}_n - \boldsymbol{\alpha}^* \boldsymbol{\beta}') \boldsymbol{\Omega} (\mathbf{I}_n - \boldsymbol{\alpha}^* \boldsymbol{\beta}')'. \quad (37)$$

In the special case that $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}$:

$$\mathbf{V} [\bar{\mathbf{e}}_{T+1|T}] = \boldsymbol{\alpha} (\boldsymbol{\beta}' \boldsymbol{\alpha}) \mathbf{V} (\boldsymbol{\alpha}' \boldsymbol{\beta}) \boldsymbol{\alpha}' + \boldsymbol{\Omega} + (\mathbf{I}_n - \boldsymbol{\alpha} \boldsymbol{\beta}') \boldsymbol{\Omega} (\mathbf{I}_n - \boldsymbol{\alpha} \boldsymbol{\beta}')'.$$

There are clear, and well known, variance costs to differencing or double differencing which need to be offset against the benefits of robustness to deterministic shifts. Even if $\boldsymbol{\alpha}$ does not change, (37) will generally not be less than (36), and will be more than double (35) in general. Thus, when there are no structural breaks, the three methods have the opposite variance ranking to their robustness, namely VEqCM is best, DV next and DDV third, as anticipated.

5.2 Variances from $T + 1$ to $T + 2$

At time $T + 1$, the DGP produces (19) and since:

$$\boldsymbol{\beta}' \mathbf{x}_{T+1} - \boldsymbol{\mu}^* + \boldsymbol{\Lambda}^* (\boldsymbol{\mu}^* - \boldsymbol{\mu}) = \boldsymbol{\Lambda}^* (\boldsymbol{\beta}' \mathbf{x}_T - \boldsymbol{\mu}) + \boldsymbol{\beta}' \boldsymbol{\nu}_{T+1}, \quad (38)$$

then:

$$\mathbf{V} [\boldsymbol{\beta}' \mathbf{x}_{T+1}] = \boldsymbol{\Lambda}^* \mathbf{V} \boldsymbol{\Lambda}^{*'} + \boldsymbol{\beta}' \boldsymbol{\Omega} \boldsymbol{\beta} = \mathbf{V}^*$$

and:

$$\mathbf{V} [\Delta \mathbf{x}_{T+1}] = \boldsymbol{\alpha}^* \mathbf{V} \boldsymbol{\alpha}^{*'} + \boldsymbol{\Omega},$$

with:

$$\mathbf{V} [\Delta \mathbf{x}_{T+2}] = \boldsymbol{\alpha}^* \mathbf{V} [\boldsymbol{\beta}' \mathbf{x}_{T+1}] \boldsymbol{\alpha}^{*'} + \boldsymbol{\Omega} = \boldsymbol{\alpha}^* \mathbf{V}^* \boldsymbol{\alpha}^{*'} + \boldsymbol{\Omega}.$$

Consequently:

$$\hat{\mathbf{e}}_{T+2|T+1} - \mathbf{E} [\hat{\mathbf{e}}_{T+2|T+1}] = (\boldsymbol{\alpha}^* - \boldsymbol{\alpha}) (\boldsymbol{\beta}' \mathbf{x}_{T+1} - \mathbf{E} [\boldsymbol{\beta}' \mathbf{x}_{T+1}]) + \boldsymbol{\nu}_{T+2},$$

so:

$$\mathbf{V} [\hat{\mathbf{e}}_{T+2|T+1}] = (\boldsymbol{\alpha}^* - \boldsymbol{\alpha}) \mathbf{V}^* (\boldsymbol{\alpha}^* - \boldsymbol{\alpha})' + \boldsymbol{\Omega}. \quad (39)$$

This is larger than (35) if there is any change in the feedback coefficient matrix, $\boldsymbol{\alpha}$, but otherwise is the same, confirming that changes in dynamics mainly induce variance effects (see table 2 in Hendry and Clements, 1998).

Next:

$$\mathbf{V} [\tilde{\mathbf{e}}_{T+2|T+1}] = \boldsymbol{\alpha}^* \mathbf{V}^* \boldsymbol{\alpha}^{*'} + \boldsymbol{\Omega}, \quad (40)$$

which coincides with (36), and the Clements and Hendry (1995) result when $\boldsymbol{\alpha}$ remains constant. Finally:

$$\mathbf{V} [\bar{\mathbf{e}}_{T+2|T+1}] = \boldsymbol{\alpha}^* (\boldsymbol{\beta}' \boldsymbol{\alpha}^*) \mathbf{V} (\boldsymbol{\alpha}^{*'} \boldsymbol{\beta}) \boldsymbol{\alpha}^{*'} + \boldsymbol{\Omega} + \mathbf{B}^* \boldsymbol{\Omega} \mathbf{B}^{*'} \quad (41)$$

where $\mathbf{B}^* = \mathbf{I}_n - \boldsymbol{\alpha}^* \boldsymbol{\beta}'$, which is generally less than the double of (40) that might have been anticipated from the extra differencing. Even when no parameters change, (41) remains:

$$\mathbf{V} [\bar{\mathbf{e}}_{T+2|T+1}] = \boldsymbol{\alpha} (\boldsymbol{\beta}' \boldsymbol{\alpha}) \mathbf{V} (\boldsymbol{\alpha}' \boldsymbol{\beta}) \boldsymbol{\alpha}' + \boldsymbol{\Omega} + \mathbf{B} \boldsymbol{\Omega} \mathbf{B}' \quad (42)$$

so no useful simplifications result.

5.3 Variances from T to $T + 2$

This comprises 2-steps ahead forecasts made prior to the break, to illustrate how the results change for multi-step forecasts. From before:

$$\hat{\mathbf{e}}_{T+2|T} - \mathbf{E} [\hat{\mathbf{e}}_{T+2|T}] = (\boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) - \boldsymbol{\alpha} (\mathbf{I}_r + \boldsymbol{\Lambda})) (\boldsymbol{\beta}' \mathbf{x}_T - \boldsymbol{\mu}) + \boldsymbol{\Upsilon}^* \boldsymbol{\nu}_{T+1} + \boldsymbol{\nu}_{T+2} \quad (43)$$

so, letting $(\boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) - \boldsymbol{\alpha} (\mathbf{I}_r + \boldsymbol{\Lambda})) = \mathbf{D}^*$:

$$\mathbf{V} [\hat{\mathbf{e}}_{T+2|T}] = \mathbf{D}^* \mathbf{V} \mathbf{D}^{*'} + \boldsymbol{\Upsilon}^* \boldsymbol{\Omega} \boldsymbol{\Upsilon}^{*'} + \boldsymbol{\Omega}. \quad (44)$$

Next:

$$\tilde{\mathbf{e}}_{T+2|T} - \mathbf{E} [\tilde{\mathbf{e}}_{T+2|T}] = \boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) (\boldsymbol{\beta}' \mathbf{x}_T - \boldsymbol{\mu}) + \boldsymbol{\Upsilon}^* \boldsymbol{\nu}_{T+1} + \boldsymbol{\nu}_{T+2}$$

so:

$$\mathbf{V} [\tilde{\mathbf{e}}_{T+2|T}] = \boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) \mathbf{V} (\mathbf{I}_r + \boldsymbol{\Lambda}^{*'}) \boldsymbol{\alpha}^{*'} + \boldsymbol{\Upsilon}^* \boldsymbol{\Omega} \boldsymbol{\Upsilon}^{*'} + \boldsymbol{\Omega}. \quad (45)$$

Finally as:

$$\boldsymbol{\beta}' \mathbf{x}_T - \boldsymbol{\mu} = \boldsymbol{\Lambda} (\boldsymbol{\beta}' \mathbf{x}_{T-1} - \boldsymbol{\mu}) + \boldsymbol{\beta}' \boldsymbol{\nu}_T,$$

and:

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

then:

$$\begin{aligned} \bar{\mathbf{e}}_{T+2|T} - \mathbf{E} [\bar{\mathbf{e}}_{T+2|T}] &= \boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) (\boldsymbol{\beta}' \mathbf{x}_T - \boldsymbol{\mu}) + \boldsymbol{\Upsilon}^* \boldsymbol{\nu}_{T+1} + \boldsymbol{\nu}_{T+2} - 2 (\Delta \mathbf{x}_T - \boldsymbol{\gamma}) \\ &= [\boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) \boldsymbol{\Lambda} - 2\boldsymbol{\alpha}] (\boldsymbol{\beta}' \mathbf{x}_{T-1} - \boldsymbol{\mu}) \\ &\quad + \boldsymbol{\Upsilon}^* \boldsymbol{\nu}_{T+1} + \boldsymbol{\nu}_{T+2} + [\boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) \boldsymbol{\beta}' - 2\mathbf{I}_n] \boldsymbol{\nu}_T. \end{aligned}$$

Letting $(\boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) \boldsymbol{\Lambda} - 2\boldsymbol{\alpha}) = \mathbf{F}^*$ and $(\boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) \boldsymbol{\beta}' - 2\mathbf{I}_n) = \mathbf{G}^*$:

$$\mathbf{V} [\bar{\mathbf{e}}_{T+2|T}] = \mathbf{F}^* \mathbf{V} \mathbf{F}^{*'} + \boldsymbol{\Upsilon}^* \boldsymbol{\Omega} \boldsymbol{\Upsilon}^{*'} + \boldsymbol{\Omega} + \mathbf{G}^* \boldsymbol{\Omega} \mathbf{G}^{*'} \quad (46)$$

Even when $\boldsymbol{\alpha}$ is constant, so:

$$\mathbf{F}^* = \boldsymbol{\alpha} [(\mathbf{I}_r + \boldsymbol{\Lambda}) \boldsymbol{\Lambda} - 2\mathbf{I}_r] \quad \mathbf{G}^* = [\boldsymbol{\alpha} (\mathbf{I}_r + \boldsymbol{\Lambda}) \boldsymbol{\beta}' - 2\mathbf{I}_n],$$

no simplification is obtained. However, if $\boldsymbol{\alpha} = \mathbf{0}$, we have:

$$\mathbf{V} [\bar{\mathbf{e}}_{T+2|T}] = 6\boldsymbol{\Omega}, \quad (47)$$

revealing that a large increase in error variance results as the horizon grows. Over longer horizons, the rate of increase suggested by comparing (37) and (47) (approximately threefold) would restrict DDV to being useful at short horizons.

5.4 Variances from $T + 1$ to $T + 3$

Now:

$$\hat{\mathbf{x}}_{T+3} = \Upsilon^2 \mathbf{x}_{T+1} + (\mathbf{I}_n + \Upsilon) (\gamma - \alpha \mu) \quad (48)$$

$$\tilde{\mathbf{x}}_{T+3} = \mathbf{x}_{T+1} + 2\gamma, \quad (49)$$

and:

$$\bar{\mathbf{x}}_{T+3} = \mathbf{x}_{T+1} + 2\Delta \mathbf{x}_{T+1}. \quad (50)$$

Then:

$$\hat{\mathbf{e}}_{T+3|T+1} - \mathbb{E} [\hat{\mathbf{e}}_{T+3|T+1}] = [\alpha^* (\mathbf{I}_r + \Lambda^*) - \alpha (\mathbf{I}_r + \Lambda)] (\kappa_{T+1}^* - \mathbb{E} [\kappa_{T+1}^*]) + \Upsilon^* \nu_{T+2} + \nu_{T+3},$$

(compare (43)) so that:

$$\mathbb{V} [\hat{\mathbf{e}}_{T+3|T+1}] = \mathbf{D}^* \mathbf{V}^* \mathbf{D}^{*'} + \Upsilon^* \Omega \Upsilon^{*'} + \Omega.$$

Apart from the changed variance of κ_{T+1}^* over κ_T , the formula has the same structure as (44) above, so new features appear for $T + 1$ to $T + 3$ (forecasting after the break) than we found for T to $T + 1$.

Next:

$$\tilde{\mathbf{e}}_{T+3|T+1} - \mathbb{E} [\tilde{\mathbf{e}}_{T+3|T+1}] = \alpha^* (\mathbf{I}_r + \Lambda^*) (\kappa_{T+1}^* - \mathbb{E} [\kappa_{T+1}^*]) + \Upsilon^* \nu_{T+2} + \nu_{T+3},$$

so that:

$$\mathbb{V} [\tilde{\mathbf{e}}_{T+3|T+1}] = \alpha^* (\mathbf{I}_r + \Lambda^*) \mathbf{V}^* (\mathbf{I}_r + \Lambda^{*'}) \alpha^{*'} + \Upsilon^* \Omega \Upsilon^{*'} + \Omega,$$

matching (45).

Finally:

$$\begin{aligned} \bar{\mathbf{e}}_{T+3|T+1} - \mathbb{E} [\bar{\mathbf{e}}_{T+3|T+1}] &= \alpha^* (\mathbf{I}_r + \Lambda^*) (\kappa_{T+1}^* - \mathbb{E} [\kappa_{T+1}^*]) + \Upsilon^* \nu_{T+2} + \nu_{T+3} \\ &\quad - 2(\Delta \mathbf{x}_{T+1} - \mathbb{E} [\Delta \mathbf{x}_{T+1}]) \\ &= \alpha^* ((\mathbf{I}_r + \Lambda^*) \Lambda^* - 2\mathbf{I}_r) \kappa_T \\ &\quad + \Upsilon^* \nu_{T+2} + \nu_{T+3} + (\alpha^* (\mathbf{I}_r + \Lambda^*) \beta' - 2\mathbf{I}_n) \nu_{T+1} \\ &= \mathbf{F}^{**} \kappa_T + \Upsilon^* \nu_{T+2} + \nu_{T+3} + \mathbf{G}^* \nu_{T+1} \end{aligned}$$

where $\alpha^* ((\mathbf{I}_r + \Lambda^*) \Lambda^* - 2\mathbf{I}_r) = \mathbf{F}^{**}$ and \mathbf{G}^* is as in (46), so that:

$$\mathbb{V} [\bar{\mathbf{e}}_{T+3|T+1}] = \mathbf{F}^{**} \mathbf{V} \mathbf{F}^{**'} + \Upsilon^* \Omega \Upsilon^{*'} + \Omega + \mathbf{G}^* \Omega \mathbf{G}^{*'} \quad (51)$$

5.5 Overview of unconditional variances

Table 1 shows the excess in the 1-step variances over Ω , for various values of α , the parameter change that most affects the variances, where $\delta_\alpha = \alpha^* - \alpha$, $\Lambda^* = \alpha^* \Lambda - \alpha$ and $\mathbf{B}^* = \mathbf{I}_n - \alpha^* \beta'$.

The unstarred matrices replace any starred parameter with its unstarred value: thus, $\mathbf{A} = \alpha \Lambda - \alpha = \alpha (\beta' \alpha)$. The pattern is clear, and the only exception is the DDV when $\alpha \neq \alpha^*$, although the fact that $\mathbf{V}^* = \Lambda^* \mathbf{V} \Lambda^{*'} + \beta' \Omega \beta$ allows considerable rearrangement without altering the substance.

Next, table 2 records the excess in the 2-step variances over Ω , where $\mathbf{C}^* = \alpha^* (\mathbf{I}_r + \Lambda^*)$, $\mathbf{D}^* = \mathbf{C}^* - \mathbf{C}$, $\mathbf{F}^* = \mathbf{C}^* \Lambda - 2\alpha$ and $\mathbf{G}^* = \mathbf{C}^* \beta' - 2\mathbf{I}_n$. As before, $\mathbf{C} = \alpha (\mathbf{I}_r + \Lambda)$ so $\mathbf{D} = \mathbf{0}$, and $\mathbf{F} = \alpha (\beta' \alpha) (2\mathbf{I}_r + \Lambda)$ and $\mathbf{G} = \alpha \beta' (2\mathbf{I}_n + \alpha \beta') - 2\mathbf{I}_n$. Also, $\mathbf{F}^{**} = \mathbf{C}^* \Lambda^* - 2\alpha^*$. These matrices also could be given different parameterizations, since (e.g.) $\mathbf{D}^* = 2\delta_\alpha + [\alpha^* (\beta' \alpha^*) - \alpha (\beta' \alpha)]$.

The pattern is similar to the 1-step outcomes, although the values are larger, and the formulae more complicated: the rapid increase in the DDV variance is especially noticable.

Table 1 1-step unconditional variances.

$T \rightarrow T + 1$	VEqCM	DV	DDV
$\alpha = 0$	0	0	Ω
$\alpha = \alpha^*$	0	$\alpha V \alpha'$	$B \Omega B' + A V A'$
$\alpha \neq \alpha^*$	$\delta_\alpha V \delta_\alpha'$	$\alpha^* V \alpha^{*'} $	$B^* \Omega B^{*'} + A^* V A^{*'} $
$T + 1 \rightarrow T + 2$	VEqCM	DV	DDV
$\alpha = 0$	0	0	Ω
$\alpha = \alpha^*$	0	$\alpha V \alpha'$	$B \Omega B' + A V A'$
$\alpha \neq \alpha^*$	$\delta_\alpha V^* \delta_\alpha'$	$\alpha^* V^* \alpha^{*'} $	$B^* \Omega B^{*'} + \alpha^* \beta' \alpha^* V \alpha^{*'} \beta \alpha^{*'} $

Here and in the text $\alpha = 0$ implicitly implies $\alpha^* = 0$.

Table 2 2-step unconditional variances.

$T \rightarrow T + 2$	VEqCM	DV	DDV
$\alpha = 0$	Ω	Ω	5Ω
$\alpha = \alpha^*$	$\Upsilon \Omega \Upsilon'$	$\Upsilon \Omega \Upsilon' + C V C'$	$\Upsilon \Omega \Upsilon' + G \Omega G' + F V F'$
$\alpha \neq \alpha^*$	$\Upsilon^* \Omega \Upsilon^{*'} + D^* V D^{*'} $	$\Upsilon^* \Omega \Upsilon^{*'} + C^* V C^{*'} $	$\Upsilon^* \Omega \Upsilon^{*'} + G^* \Omega G^{*'} + F^* V F^{*'} $
$T + 1 \rightarrow T + 3$	VEqCM	DV	DDV
$\alpha = 0$	Ω	Ω	5Ω
$\alpha = \alpha^*$	$\Upsilon \Omega \Upsilon'$	$\Upsilon \Omega \Upsilon' + C V C'$	$\Upsilon \Omega \Upsilon' + G \Omega G' + F V F'$
$\alpha \neq \alpha^*$	$\Upsilon^* \Omega \Upsilon^{*'} + D^* V^* D^{*'} $	$\Upsilon^* \Omega \Upsilon^{*'} + C^* V^* C^{*'} $	$\Upsilon^* \Omega \Upsilon^{*'} + G^* \Omega G^{*'} + F^{**} V F^{**'} $

5.6 Discussion

When forecasting before a break, all three models are susceptible to forecast failure, and there is little to choose between them, although the VEqCM has the smallest variance component when it is correctly specified and no break occurs. When forecasting after a break, the DDV has the greatest robustness to a deterministic shift, but the largest and most rapidly-increasing forecast-error variances in general. The DV lies between, depending on what deterministic terms change.

Nevertheless, the longer the multi-step evaluation horizon, the less well the DDV, and probably the DV will perform, partly from their variance terms, and partly because most breaks will be after forecasting, a case in which these models offer no gains. Conversely, the shorter the horizon, for a sequence of horizons, the more likely some breaks will precede forecasting, and consequently, DDV and DV may outperform the VEqCM, even when it is correctly-specified in-sample.

This behaviour is precisely what was observed by Eitrheim, 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, 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, the DDV did best more often than any other method. The empirical illustration confirms similar results for a small monetary model of the UK.

5.7 Correct empirical variances

For the DV and DDV models, the programmed formulae for the forecast-error variances assume that the errors are serially uncorrelated. This can be misleading when the DGP is a VEqCM. In this section, we

discuss the potential impact of the residual autocorrelation on the error bands shown in their forecast graphs. It is important not to model the residual autocorrelation if the robustification properties of the methods are to be preserved.

Consider the DV model first. The programmed formulae assume it is the DGP, so its ‘error’ ξ_t is treated as being serially uncorrelated, leading to the computation of:

$$V_{DV,p}[\tilde{\mathbf{e}}_{x,T+h}] = \mathbb{E} \left[\sum_{j=0}^{h-1} \sum_{i=0}^{h-1} \mathbf{I}^i \xi_{T+h-i} \xi'_{T+h-j} \mathbf{I}^{j'} \right] = h \mathbf{\Omega}_{\xi} \sim O(h), \quad (52)$$

when $\mathbf{\Omega}_{\xi}$ is replaced by an estimate of the in-sample error-covariance matrix. It is possible to obtain an ‘autocorrelation-corrected’ formula for the DV, denoted $V_{DV,c}$, which is also $O(h)$. The actual MSFE is:

$$V_{DV,a}[\tilde{\mathbf{e}}_{x,T+h}] = \sum_{i=0}^{h-1} \sum_{q=0}^{h-1} \alpha \mathbf{\Lambda}^i \mathbf{V} \mathbf{\Lambda}^{q'} \alpha' + \sum_{s=0}^{h-1} \mathbf{\Upsilon}^s \mathbf{\Omega} \mathbf{\Upsilon}^{s'} \sim O(h). \quad (53)$$

Similar problems afflict the DDV model forecasts, where the actual forecast error uncertainty is:

$$V_{DDV,a} = \sum_{i=0}^{h-1} \mathbf{\Upsilon}^i \mathbf{\Omega} \mathbf{\Upsilon}^{i'} + h^2 \mathbf{\Omega}_{\xi} + \alpha \mathbf{\Lambda}_h \mathbf{V} \mathbf{\Lambda}'_h \alpha' - h \alpha \mathbf{\Lambda}_h \mathbf{C}_{\beta x_T, \Delta x_T} - h \mathbf{C}'_{\beta x_T, \Delta x_T} \mathbf{\Lambda}'_h \alpha' \sim O(h^2)$$

where $\mathbf{\Lambda}_h = \sum_{i=0}^{h-1} \mathbf{\Lambda}^i$ and $\mathbf{C}_{\beta x_T, \Delta x_T} = \mathbf{\Lambda} \mathbf{V} \alpha' + \beta' \mathbf{\Omega}$. However:

$$V_{DDV,p} = \left(\sum_{i=0}^h i^2 \right) \mathbf{\Omega}_{\zeta} = \frac{1}{6} h(h+1)(2h+1) \mathbf{\Omega}_{\zeta} \sim O(h^3) \quad (54)$$

where:

$$\mathbf{\Omega}_{\zeta} = [\mathbf{\Omega} + (\alpha \beta' - \mathbf{I}_n) \mathbf{\Omega} (\beta \alpha' - \mathbf{I}_n) + \mathbf{A} \mathbf{V} \mathbf{A}'] .$$

Thus, a serious overestimation of the uncertainty results from using (54). An expression for $V_{DDV,c}$ is also obtainable, but for simplicity, we again only report numerical calculations.

5.7.1 Bivariate illustration

We now quantify the extent to which the formula in (52), calculated without allowing for the in-sample errors to be autocorrelated, is a misleading guide to the true forecast error uncertainty associated with the DV model forecasts, given by (53).

When $n = 2$, $\mathbf{\Lambda} = 1 + \beta' \alpha = \lambda$ is a scalar. Then the formulae for the DV model simplify to:

$$\begin{aligned} V_{DV,a}[\tilde{\mathbf{e}}_{x,T+h}] &= \sum_{s=0}^{h-1} \mathbf{\Upsilon}^s \mathbf{\Omega} \mathbf{\Upsilon}^{s'} + \alpha \mathbf{V} \alpha' \frac{(1-\lambda^h)^2}{(1-\lambda)^2} \\ V_{DV,p}[\tilde{\mathbf{e}}_{x,T+h}] &= h \mathbf{\Omega}_{\xi} \\ V_{DV,c}[\tilde{\mathbf{e}}_{x,T+h}] &= h \mathbf{\Omega}_{\xi} + \sum_{j=2}^h \sum_{i=1}^{j-1} (2\alpha \mathbf{V} \alpha' \lambda^i + \alpha \beta' \mathbf{\Omega} \lambda^{i-1} + \mathbf{\Omega} \beta \alpha' \lambda^{i-1}) \end{aligned}$$

Similar expressions hold for the DDV.

Table 3 gives the results of numerically evaluating the forecast-error variance formulae for DV and DDV models for the set of parameter values $\{\mathbf{\Omega} = \mathbf{I}_2, \alpha = [-0.2, 0.1], \beta = [1, -1]\}$. We report figures for predicting levels of the first variable at horizons up to 12: qualitatively similar results hold for the second variable.

Table 3 Illustrative forecast-error variances.

h	V_{VEqCM}	$V_{DV,a}$	$V_{DV,p}$	$V_{DV,c}$	$V_{DDV,a}$	$V_{DDV,p}$	$V_{DDV,c}$
1	1.00	1.16	1.16	1.16	2.49	2.49	2.49
2	1.68	2.13	2.31	2.53	7.37	12.47	7.47
3	2.23	2.98	3.47	4.06	14.58	34.92	14.98
4	2.74	3.75	4.63	5.70	24.08	74.82	25.09
5	3.24	4.45	5.78	7.41	35.87	137.18	37.88
6	3.74	5.10	6.94	9.18	49.93	226.96	53.46
7	4.26	5.73	8.10	10.98	66.28	349.18	71.92
8	4.79	6.33	9.25	12.81	84.91	508.80	93.37
9	5.32	6.92	10.41	14.66	105.82	710.82	117.92
10	5.86	7.50	11.57	16.52	129.03	960.24	145.66
11	6.40	8.08	12.73	18.39	154.54	1262.02	176.72
12	6.95	8.64	13.88	20.26	182.35	1621.18	211.18

DV results. Forecast confidence intervals calculated on the basis of V_p are far too wide, particularly at longer horizons, and overstate the true (V_a) uncertainty attached to the DV model forecasts. Correcting for the serial correlation in the DV model residuals (as in V_c) tends to inflate the bands further, since the omission of the equilibrium-correction term induces positive autocorrelation in the DV equations' errors. The first column records the DGP baseline forecast-error variance, as an aid to comparison. The degree to which V_p (and V_c) overstate the uncertainty depends on λ . As $\lambda \rightarrow 1$, so that there is no cointegration and the DV is correctly specified, then all the measures converge.

DDV results. The true forecast uncertainty associated with the DDV model increases much more quickly, but this is greatly exacerbated by the V_p measure. Now, the correction for autocorrelation (V_c) largely undoes the inflation in the V_p measure, by accounting for the substantial negative moving average induced in the DDV model errors. These results should be kept in mind when interpreting the empirical application below.

6 Empirical example: UK M1

We consider UK quarterly data (seasonally adjusted), to study the impact on forecast failure of a major financial innovation in 1984(2).² Hendry and Ericsson (1991) show that the resulting introduction of non-zero own interest rates (learning adjusted) on checking accounts was tantamount to a deterministic shift in the equilibrium demand for M1, and failure to model that effect induced very poor forecasts. The model we investigate is a descendant of that first proposed in Hendry (1979), and builds on Hendry and Mizon (1993) who embedded it in a 4-variable system. Hendry (1996) considered the forecast behaviour of the single-equation model of UK M1 estimated over the sample 1963(3) to 1983(2), and showed that its forecasts failed badly when the data period was extended to 1989(2). This study extends his analysis to the multivariate context. Finally, Hendry and Doornik (1994) embedded the equation from Hendry and Ericsson (1991) in a 4-variable system. Relative to these studies, we return to the system in Hendry and Mizon (1993), to illustrate the impact of an unmodelled deterministic shift, but over an extended

²Ericsson, Hendry and Tran (1994) consider the issue of seasonal adjustment of this data, and compare the results to those obtained using the unadjusted series, showing only small distortions result.

Table 4 System goodness of fit and evaluation.

statistic	$m - p$	i	Δp	R	VAR
$\hat{\sigma}$	1.76%	1.11%	0.62%	1.33	
$F_{ar}(5, 67)$	3.15*	1.02	1.68	1.02	
$F_{arch}(4, 64)$	0.40	0.24	0.58	2.32	
$F_{het}(18, 53)$	0.82	1.17	0.74	1.32	
$\chi^2_{nd}(2)$	3.08	1.62	3.76	1.99	
$F^v_{ar5}(80, 195)$					1.40*
$F^v_{het}(180, 407)$					0.94
$\chi^2_{nd^v}(8)$					4.95

sample. Further, we focus on the multi-period forecast performance of the alternative systems under analysis, to discuss which methods win in practice in this setting. When needed, we treat the model in Hendry and Doornik (1994) as if it were the DGP.

Let M denote nominal M1, I total final expenditure, P its deflator, and R the interest rate on three-month Local-Authority bills: lower case denotes logs, and $\Delta = (1 - L)$ is the first difference, when L is the lag operator. We consider forecasting over two distinct historical periods. For the first, the sample period is 1964(3)–1985(2), after initial values for lags, with the remaining observations for 1985(3)–1989(2) retained for out-of-sample forecasting. The four variables ($m - p$, Δp , i , R) appear to be $I(1)$, so we begin by developing a dynamic system, undertake a cointegration analysis, then simplify to a model in $I(0)$ space. As expected, the system's multi-step forecast performance is very poor. Adding a step-shift dummy to allow a separate intercept (autonomous growth) over the forecast period rescues the forecasts, similar to those from the 'correct' model: this is a form of intercept correction (see Hendry and Clements, 1994). We also develop 'time-series' models which do not fail on forecasting, as the test period commences after the structural break. Thus this choice of forecast period illustrates the efficacy of intercept corrections and 'time-series' models when a major break has occurred prior to forecasting. The second exercise selects the immediately preceding 16 observations as the forecast period, i.e., 1981(3)–1985(2), to assess the costs of these strategies over a period when the dynamic system remains a reasonably good approximation to the DGP. The specifications of the models for the second exercise are carried over directly from those estimated on the longer sample – given its in-sample constancy, re-specifying the model would probably make little difference. Below we emphasize the 'post-break' forecast period, where the hitherto well-specified simultaneous equation model exhibits spectacular forecast failure, but full results are also reported for the 'pre-break' forecast period for the empirical forecast comparisons of the forecasting methods.

6.1 A 4-equation VAR

The variables $((m - p)_t, \Delta p_t, i_t, R_t)$ were analyzed in a VAR with 2 lags, including a constant, linear deterministic trend, and two indicator variables for output ($dout$ equal to zero, except for unity in 1972(4), 1973(1), and 1979(2)) and the oil crises ($doil$, unity in 1973(3), 1973(4), 1974(2) and 1979(3)). These indicators adjust for the largest residuals in the system; the issues raised by how the dummies enter the system are discussed in Doornik, Hendry and Nielsen (1998). All computations and graphics were produced by GiveWin and PcFiml (see Doornik and Hendry, 1996, 1997).

Table 4 shows the individual-equation, and system, goodness-of-fit and evaluation statistics: $\hat{\sigma}$ is the standard deviation of the residuals, and diagnostic tests are of the form $F_j(k, T - l)$ which denotes an

Table 5 System correlations and residual cross correlations.

	$m - p$	i	Δp	R
$m - p$	0.986			
i	0.01	0.998		
Δp	-0.47	-0.15	0.911	
R	-0.48	0.04	0.29	0.924

F-test against the alternative hypothesis j for: 5^{th} -order serial correlation (F_{ar} : see Godfrey, 1978), 4^{th} -order autoregressive conditional heteroscedasticity (F_{arch} : see Engle, 1982), heteroscedasticity (F_{het} : see White, 1980); and a chi-square test for normality ($\chi_{nd}^2(2)$: see Doornik and Hansen, 1994): * and ** denote significance at the 5% and 1% levels respectively. Vector tests are shown as $F_j^v(k, T - l)$.

The money-demand equation reveals residual serial correlation, but otherwise the outcomes are consistent with a congruent system. Table 5 records the correlations of fitted and actual (on the diagonal), and the residual inter-correlations: those between $m - p$ and $(\Delta p, R)$ are much the largest.

The eigenvalues of the long-run matrix are -0.41 , -0.05 , and $-0.11 \pm 0.19\iota$ (using ι to denote $\sqrt{-1}$ to avoid confusion with income, i), so the rank seems non-zero, and is likely to be one or perhaps two. The eigenvalues of the companion matrix (denoted λ) are shown in table 6. Only one root is very close to unity, two have moduli near 0.9, and the remainder are small.

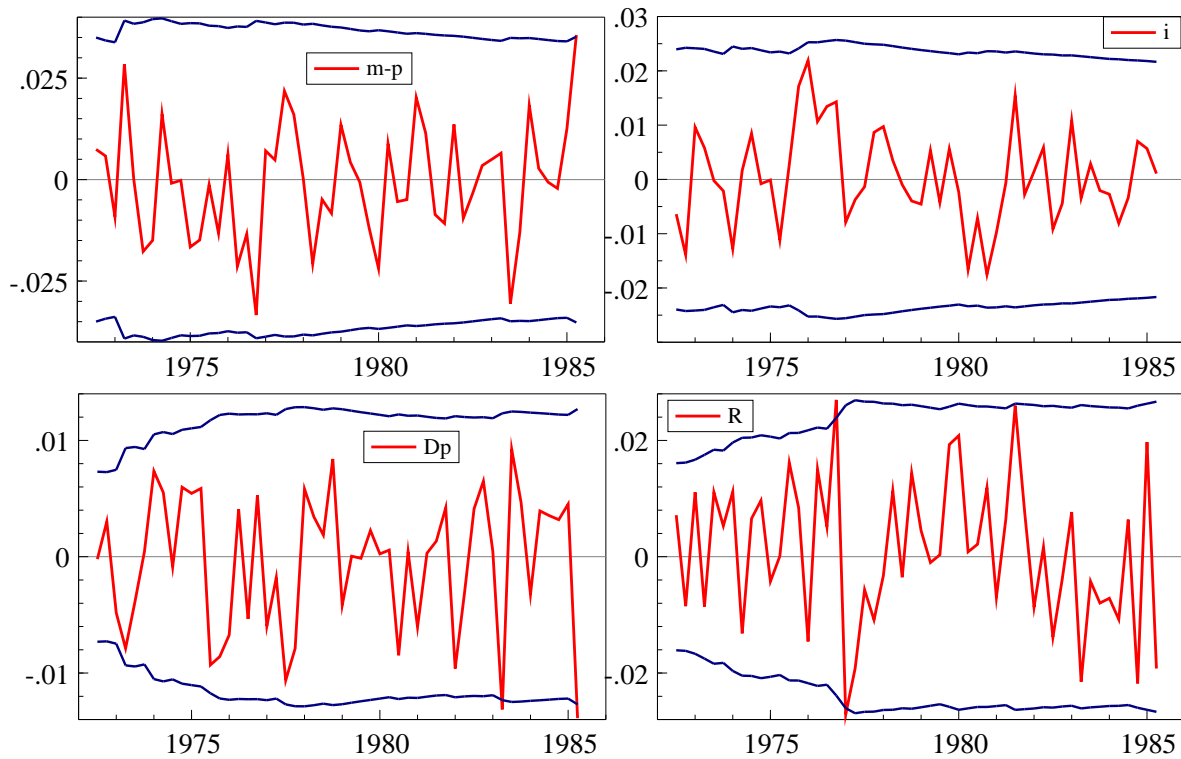


Figure 2 System recursive 1-step residuals.

All first lags were significant, but the second lags and the trend were insignificant (on $F(4, 72)$, at 5% or less). Figure 2 shows the in-sample recursively-computed system 1-step residuals with 95% confidence bands: the equations for Δp and R are somewhat non-constant, although the system break-

Table 6 System dynamics.

λ	0.97	$0.86 \pm 0.17t$	0.64	-0.32	$-0.22 \pm 0.12t$	0.19
$ \lambda $	0.97	0.87, 0.87	0.64	0.32	0.25, 0.25	0.19

point Chow (1960) test did not exceed the 1% critical value within sample.

Figure 3 reports the 1-step ahead out-of-sample forecasts with approximate 95% confidence intervals: there is some evidence of mis-forecasting in the money and interest-rate equations, but overall, the performance is respectable, consistent with the constancy-test outcome of $F(64, 72) = 0.87$. This outcome can be interpreted as a sequence of the form in §4.1 and §4.2, etc.

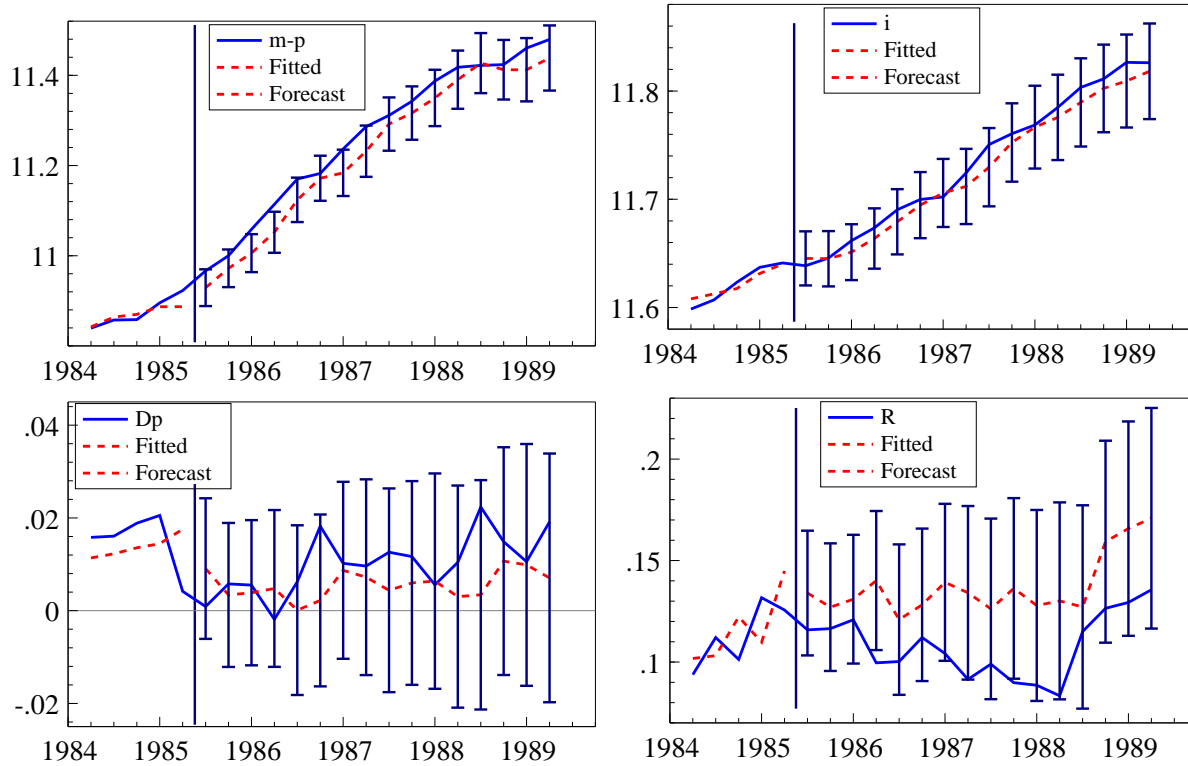


Figure 3 System 1-step ahead forecasts.

Finally, fig. 4 records the fitted and actual values for each variable together with the 16-steps ahead forecasts and their approximate 95% confidence intervals: the excellent fit but awful multi-step forecast performance of this unrestricted system is manifest. This outcome can be interpreted as a sequence of the form in §4.1 and §4.3, etc., so is the combination of the post-forecasting deterministic shift due to financial innovation, interacting with the $I(1)$ formulation and the over-parameterization. We address these last two issues in turn.

6.2 Cointegration

The cointegration analysis restricted the trend to the cointegration space, but the constant and dummies entered unrestrictedly, and yielded (see Doornik and Hendry, 1997, Banerjee, Dolado, Galbraith and Hendry, 1993, and Johansen, 1995). For each value of the rank $s \geq 1$ of the long-run matrix in the Johansen (1988) procedure, Table 7 reports the log-likelihood values (ℓ), eigenvalues (λ) and associated

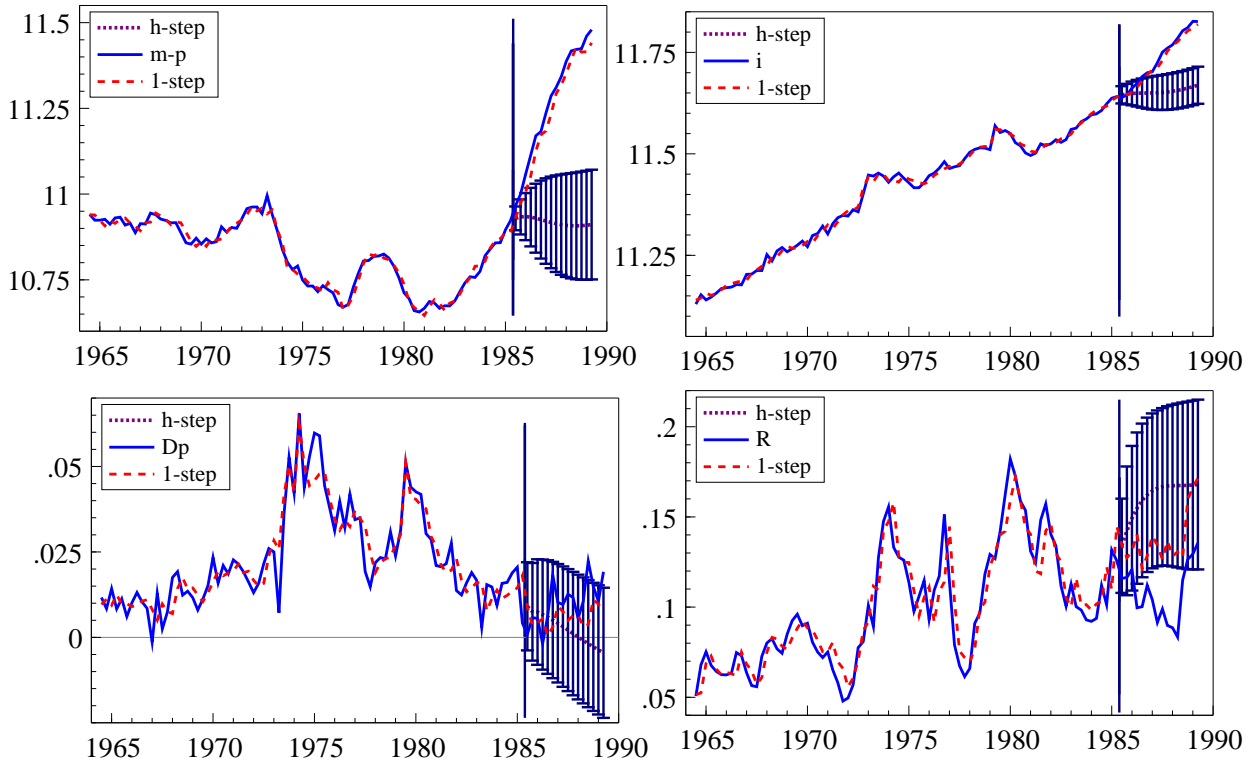


Figure 4 System graphical statistics with 16-step ahead forecasts.

maximum eigenvalue (Max) and trace (Tr) statistics together with the estimated cointegrating vectors ($\hat{\beta}$) and feedback coefficients ($\hat{\alpha}$). The null of no cointegration is strongly rejected at conventional $l(1)$ critical values, and although the second cointegrating vector is not very significant, we retain it following Hendry and Mizon (1993), given the interpretability of its coefficients after restrictions.

To uniquely determine and interpret the two possible cointegration vectors, we removed the trend from the first, and $m - p$ from the second. Then we restricted the income coefficient to -1 in the first vector, and the trend coefficient in the second to the mean value of Δi (namely, 0.0062 , approximately 2.5% p.a.), also eliminating inflation. Finally, we set the feedbacks to zero for the second vector on the first equation, and the first on the last three equations (related to long-run weak exogeneity) which yielded the results shown in Table 8, with the test of the restrictions being $\chi^2(7) = 9.58$.

The first cointegration vector relates the ratio of money to expenditure ($m - p - i$) negatively to inflation and interest rates, so it has the interpretation of an excess demand for transactions money. The second cointegration vector is interpretable as the excess demand for goods and services (the deviation of expenditure from trend, negatively related to interest rates), and its main influence is onto the i equation, so we retain these two long-run relations.

Figure 5 records the time series of the two unrestricted and restricted cointegration vectors, and the associated recursively-computed eigenvalues. Those for money are very similar, whereas the restrictions have noticeably altered the second; both eigenvalues seem constantly estimated (see Hansen and Johansen, 1992). Thus, the two, zero-mean, $l(0)$ linear combinations defining the equilibrium-correction mechanisms (EqCMs) are:

$$c_{1,t} = m_t - p_t - i_t + 6.67\Delta p_t + 6.53R_t - 0.223 \quad (55)$$

Table 7 Cointegration analysis.

s	1	2	3	4
ℓ	1541	1548	1553	1556
λ	0.52	0.16	0.11	0.06
Max	61.4**	14.8	9.9	5.3
Tr	91.4**	30.0	15.2	5.3

,

$\hat{\alpha}$	1	2	3	4
$m-p$	-0.03	-0.01	-0.16	-0.02
i	-0.01	0.00	0.37	-0.00
Δp	-0.00	0.02	-0.06	0.01
R	-0.01	-0.03	0.00	0.01

$\hat{\beta}'$	$m-p$	i	Δp	R	t
1	1	-0.74	13.4	16.6	-0.0125
2	-1.12	1	-10.1	2.56	-0.0108
3	0.07	-0.46	1	-0.11	0.0026
4	1.37	4.43	-6.54	1	-0.0271

and:

$$c_{2,t} = i_t - 0.0062t + 1.22R_t - 11.125. \quad (56)$$

The definitions in (55) and (56) are required for multi-step forecasts when formulating the model in terms of the differences $(\Delta(m-p)_t, \Delta i_t, \Delta^2 p_t, \Delta R_t)$ of the original variables.

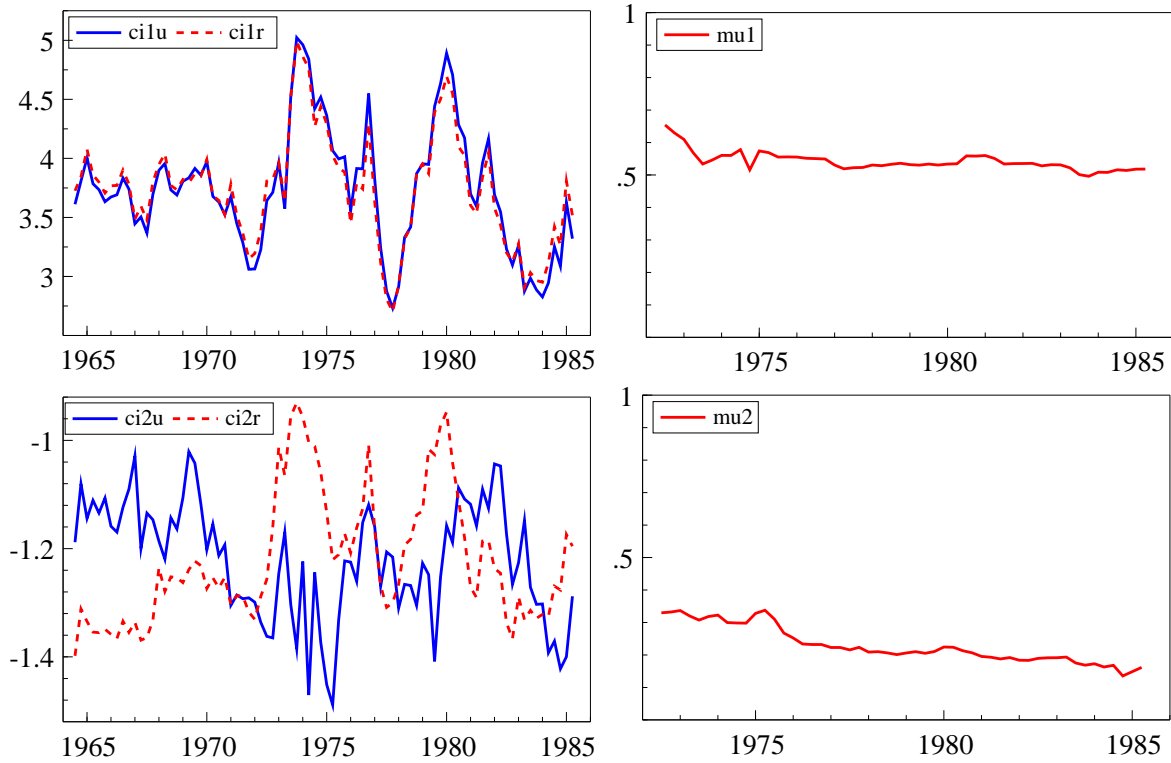


Figure 5 Cointegration vectors, and recursively-computed eigenvalues.

Table 8 Restricted Cointegration analysis.

$\hat{\alpha}$	1	2				
$m - p$	-0.098 (0.014)	0 (-)				
i	0 (-)	-0.124 (0.030)				
Δp	0 (-)	-0.019 (0.016)				
R	0 (-)	-0.007 (0.035)				

,

$\hat{\beta}'$	$m - p$	i	Δp	R	t
1	1 (-)	-1 (-)	6.67 (1.61)	6.53 (0.67)	0 (-)
2	0 (-)	1 (-)	0 (-)	1.22 (0.31)	-0.0062 (-)

6.3 The $l(0)$ system

Going from the second-order VAR in the levels of the variables $((m - p)_t, \Delta p_t, i_t, R_t)$ to a simultaneous-equations model involves a number of steps (see, e.g., Hendry and Mizon, 1993, Clements and Mizon, 1991), any of which might potentially affect forecast performance. We first considered the impact of imposing cointegration. The initial system in the levels of the variables is given an equivalent representation in terms of differences, cointegrating combinations, and (two) lagged level terms. The $l(1)$ level terms are then deleted from all four equations to assess the impact of imposing unit roots and cointegration, as studied by Clements and Hendry (1995) via Monte Carlo, and empirically in a simplified monetary model, as distinct from parsimony *per se*, so the insignificant $l(0)$ terms are retained. Imposing cointegration made little difference to forecasting performance relative to the unrestricted VAR. Full results are shown in the empirical forecast comparisons.

6.4 A simultaneous-equations model

A model of the $l(0)$ system was developed by sequential simplification, based on earlier findings, and delivered the estimates shown in Table 9, augmented by the definitions in (55) and (56). This resulted in only 13 estimated $l(0)$ parameters with the three $l(1)$ from Table 8 (so should avoid any over-parameterization problems that may have affected the initial system), but was an acceptable reduction as the likelihood-ratio test of all the restrictions yielded $\chi^2_{\text{or}}(23) = 14.1$ ($p > 0.92$), which does not reject. Table 10 records the model evaluation statistics (entries above the first diagonal are for the $l(0)$ -VAR; the diagonal shows $\hat{\sigma}$ values, and the lower triangle, the model outcomes).

Thus, the model is a valid reduction of the initial system. However, $F_{\text{Ch}}(64, 80) = 3.99^{**}$, so parameter constancy out-of-sample is strongly rejected, and the 1-step forecast performance is poor relative to the in-sample fit, as fig. 6 shows, especially for $\Delta(m - p)$.

6.5 Multi-step forecasts

The model's multi-step forecast performance is not nearly as poor as that of the system, suggesting some benefit from the $l(0)$ reduction and parsimony, perhaps because the explanatory variables in Table 9 have small mean values. Figure 7 shows both the multi-step and the 1-step forecasts for comparison, only over the forecast horizon. For the last three variables, the forecast bands for the multi-step forecasts hardly increase as the horizon increases, consistent with their nearly non-dynamic nature, so the forecasts quickly become the unconditional means of their respective growth rates. Further, the bands are not

Table 9 FIML model estimates.

$$\begin{aligned}
\Delta(m-p)_t &= - \underset{(0.06)}{0.16} \Delta(m-p-i)_{t-1} - \underset{(0.16)}{0.68} (\Delta^2 p_t + \Delta R_t) - \underset{(0.009)}{0.098} c_{1,t-1} \\
\Delta i_t &= \underset{(0.006)}{0.050} Dout - \underset{(0.022)}{0.11} c_{2,t-1} + \underset{(0.07)}{0.20} \Delta i_{t-1} + \underset{(-)}{0.0062} \\
\Delta^2 p_t &= - \underset{(0.08)}{0.34} \Delta^2 p_{t-1} + \underset{(0.004)}{0.028} Doil - \underset{(0.0007)}{0.0015} - \underset{(0.015)}{0.029} c_{2,t-1} \\
\Delta R_t &= \underset{(0.10)}{0.14} \Delta R_{t-1} + \underset{(0.007)}{0.014} Doil + \underset{(0.06)}{0.14} \Delta(m-p)_{t-1}
\end{aligned}$$

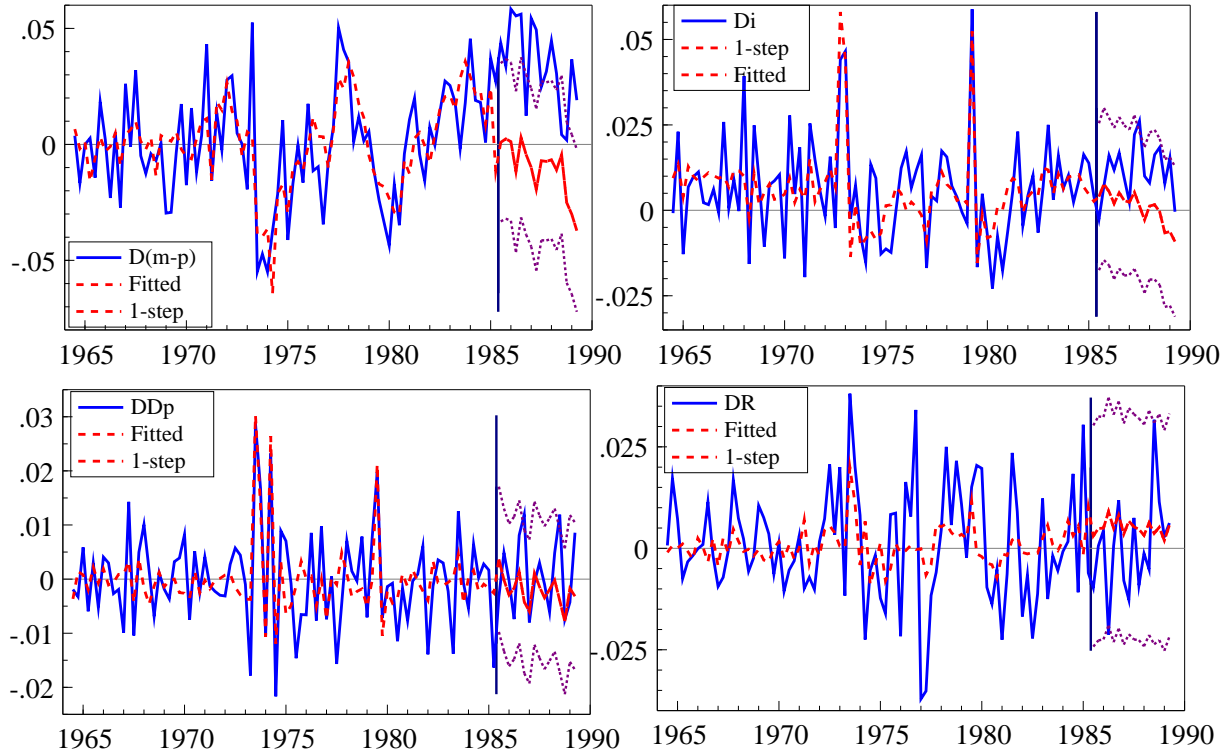


Figure 6 Fitted and actual values with 1-step forecasts, from the VEqCM.

Table 10 Model evaluation statistics.

	Residual correlations					Model diagnostic tests	
	$\Delta(m-p)_t$	Δi_t	$\Delta^2 p_t$	ΔR_t		statistic	
$\Delta(m-p)_t$	1.40%	0.03	-0.47	-0.45	,	$F_{ar}^v(80, 231)$	1.25
Δi_t	0.02	1.06%	-0.19	0.06		$F_{het}^v(160, 488)$	0.84
$\Delta^2 p_t$	-0.09	-0.17	0.65%	0.24		$\chi_{nd}^2 v(8)$	6.97
ΔR_t	0.18	0.07	0.27	1.35			

much larger than the corresponding 1-step bars. The bands increase at first for $m-p$, where the EqCM

plays a key role, but again the forecasts converge to the mean growth, although now the 1-step bars are distinctly narrower. Nevertheless, despite the structural break revealed by the parameter constancy test, multi-step predictive failure is not nearly so manifest.

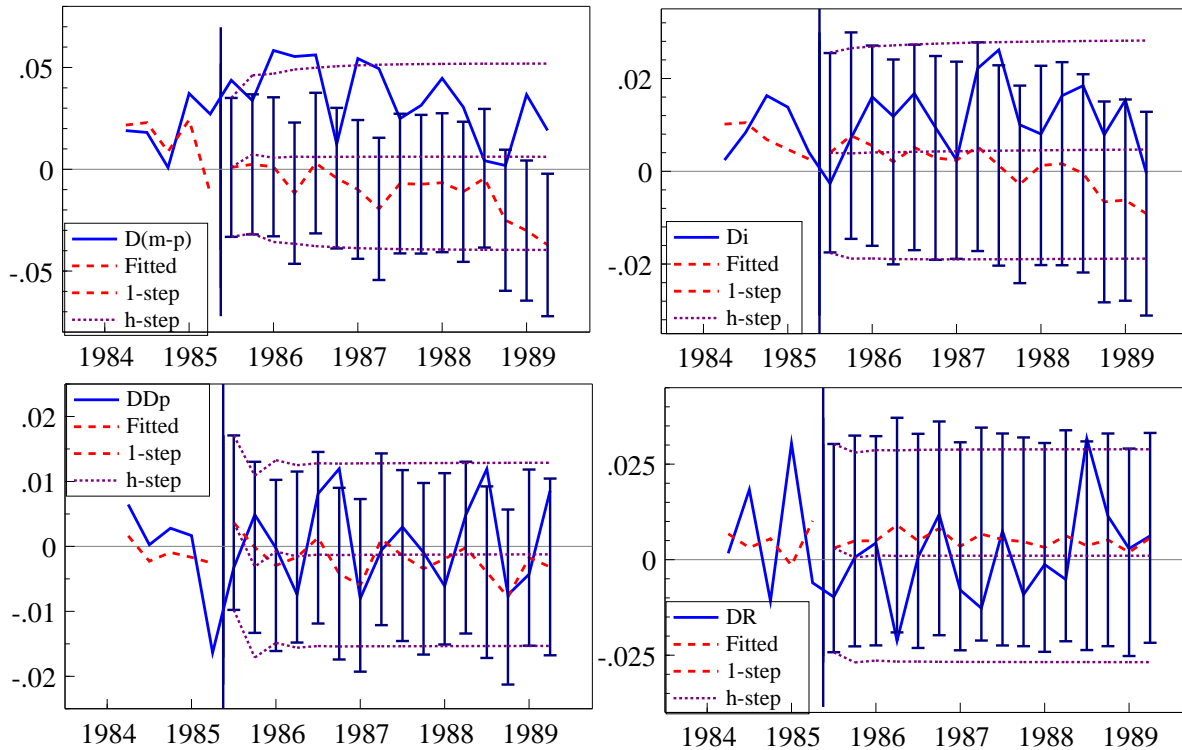


Figure 7 Multi-step and the 1-step forecasts, from the VEqCM.

7 Forecast comparisons

The theory above predicts that the DV and DDV should be less susceptible to a deterministic structural break in the equilibrium mean than the VEqCM, but have larger forecast standard errors. The former corresponds to dropping the EqCMs from the VEqCM, replacing the $\Delta(m-p)$ equation by its ‘reduced form’, and eliminating insignificant variables in the resulting model. Figure 8 shows the combined 1-step and multi-step forecasts. By eliminating the equilibrium-correction terms, the DV suffers from residual autocorrelation ($F_{ar}(80, 231) = 1.51^{**}$), so its confidence intervals calculated by the usual formulae are incorrect, probably understating the actual uncertainty. Nevertheless, the absence of bias in the forecasts conforms to the earlier theory, when forecasting after a break.

Figure 9 shows the same set of forecasts for the DDV. By double differencing, there is substantial negative residual autocorrelation ($F_{ar}(80, 211) = 2.01^{**}$), so the calculated confidence intervals are again incorrect, this time seriously overstating the uncertainty. Nevertheless, the bias performance seems good visually.

Next, fig. 10 compares all three multi-step forecasts, in the space of $(\Delta(m-p), \Delta i, \Delta R, \Delta^2 p)$. The actual multi-step forecasts are very similar for all three forecasting devices, namely zero $(\Delta R, \Delta^2 p)$ or the unconditional growth rate $(\Delta(m-p), \Delta i)$. In this representation, the DDV has easily the largest confidence intervals, and they increase rapidly in the horizon, matching the theoretical calculations.

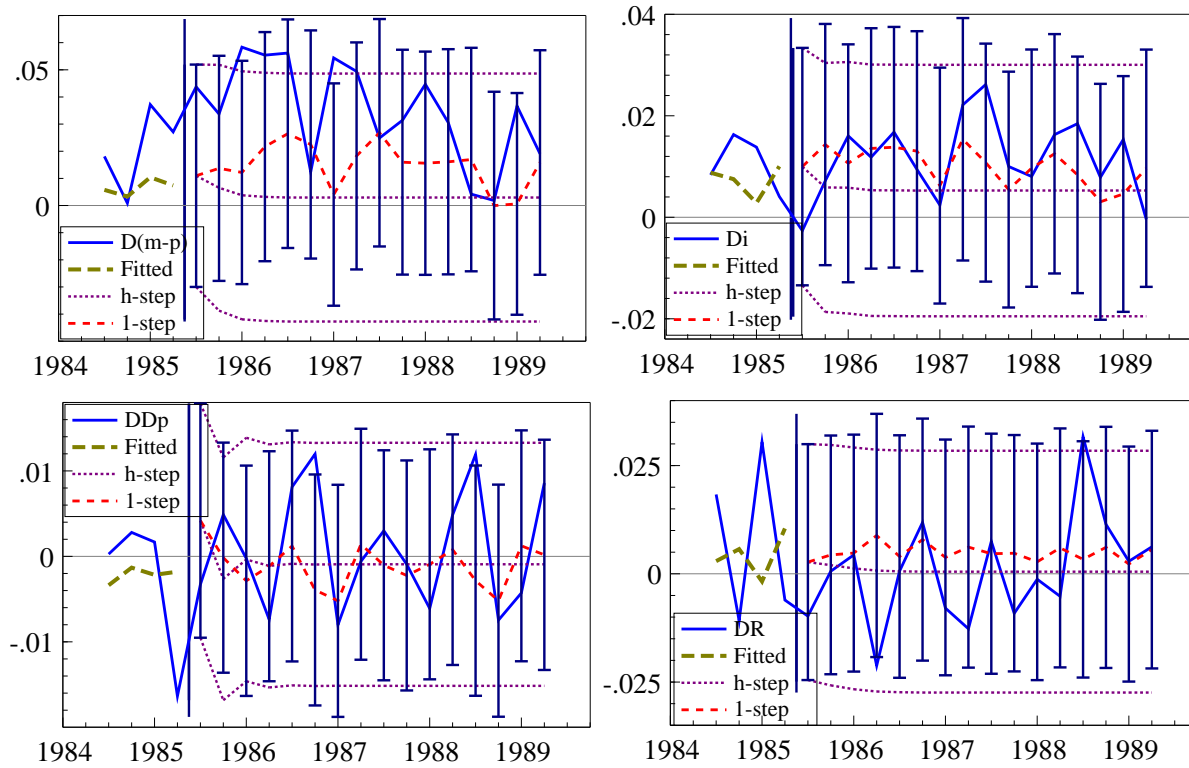


Figure 8 Multi-step and the 1-step forecasts, from the DV.

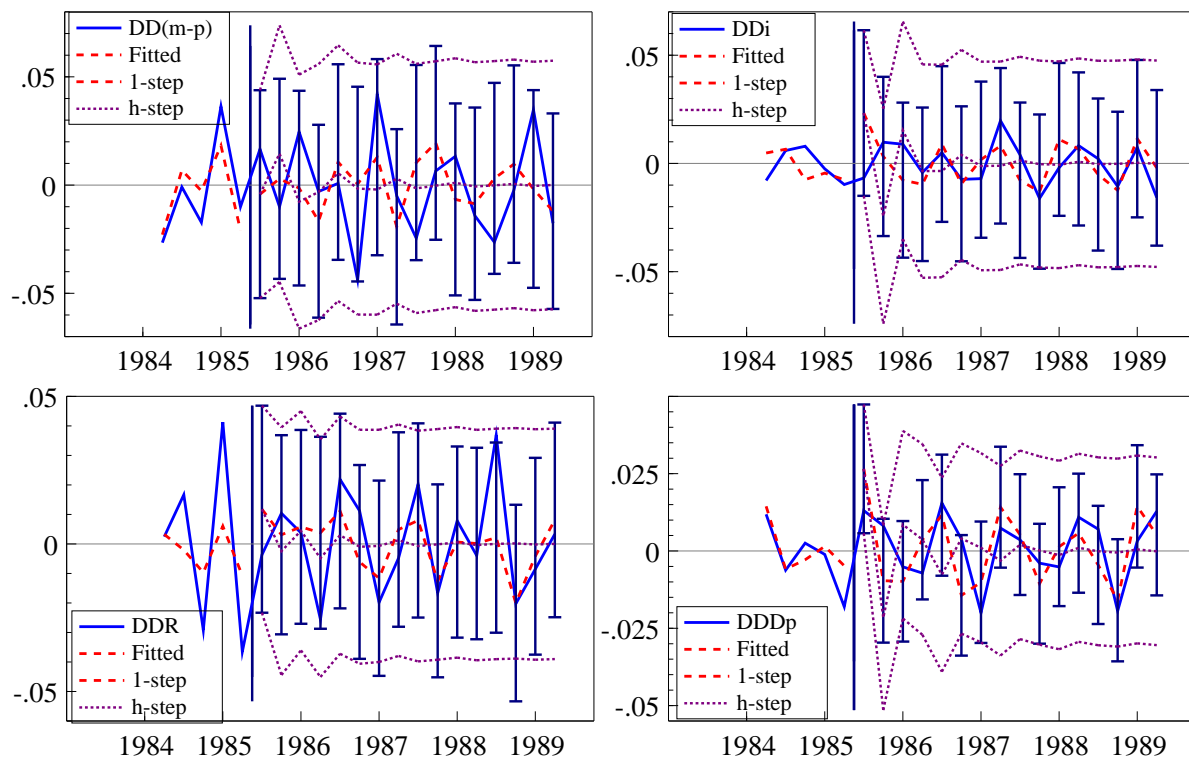


Figure 9 Multi-step and the 1-step forecasts, from the DDV.

Between the VEqCM and the DV, the VEqCM has the wider intervals for money demand where the EqCM is strongest, but they are closely similar for the other three variables.

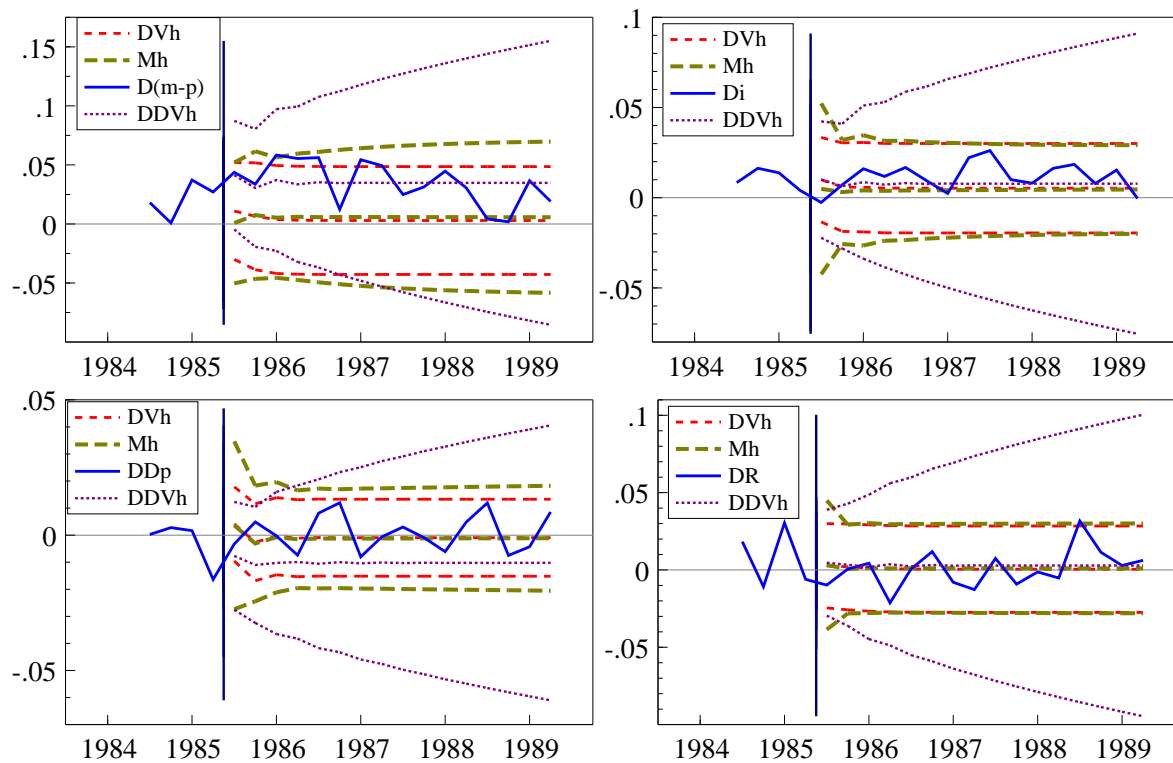


Figure 10 Multi-step forecasts in differences from the three forecasting devices.

Finally, we show the comparison between the VEqCM and the DDV in levels in fig. 11. This exemplifies the main theory predictions: the robustness to the break of the DDV, but its rapidly increasing – and large – confidence intervals as against the forecast failure of the VEqCM after the break, although the results in §5.7 show that the uncertainty is badly overestimated by the bands shown in the figures.

8 Intercept corrections

When the source of a model's mis-specification is known, it is usually corrected, but in many settings, mis-specifications are unknown, so are difficult to correct. One widely-used tool is intercept correction (denoted IC), which sets the model 'back on track' to start from the actual forecast origin x_T . Hendry and Clements (1994) develop a general theory of intercept corrections, and Clements and Hendry (1996) show that such corrections can robustify forecasts against breaks that have happened, but only at the cost of an often substantial increase in forecast-error variance. The form of correction envisaged in that analysis is such that the correction alters as the forecast origin moves through the sample – the correction is always based on the error(s) made at, or immediately prior to, the origin. However, those forms of correction require a steadily expanding information set, and to treat the intercept-correcting strategy on a par with the other forecasting models, in this section we consider a simpler correction. This form of intercept correction can be implemented by adding an indicator variable equal to unity from the last sample observation onwards, so that the same correction is applied at all forecast origins. Figure 7a shows why such an IC will work here: immediately prior to forecasting, the model is under-fitting by a substantial amount, and 'shifting' the forecast origin to the data will offset much of the later

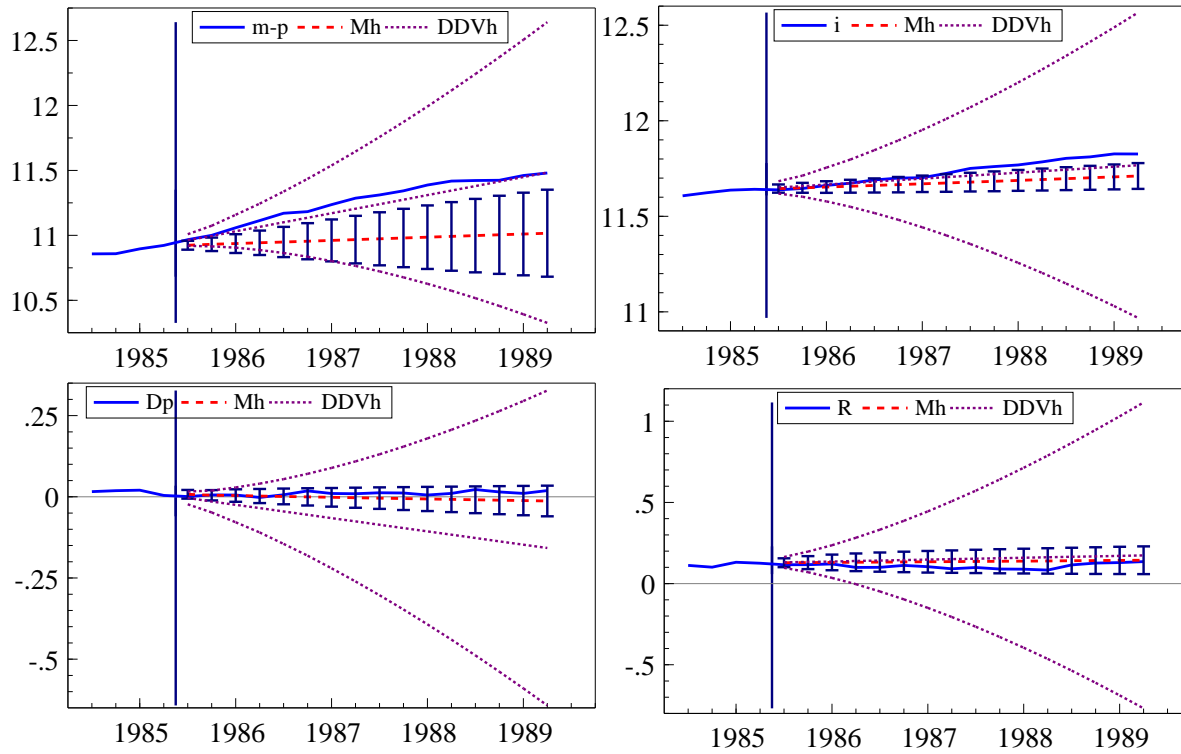


Figure 11 Multi-step forecasts in levels from the VEqCM and DV.

mis-forecasting. To reduce the forecast-error variance, the IC can be set to unity for the last few sample observations: here we chose two (namely 1985(1), (2)). Further, to highlight the effects, we only entered the IC in the first equation (where it was significant at the 5% level: it was insignificant if added to the remaining equations). Figure 12 shows the impact in the $l(0)$ representation, and fig. 13 in the $l(1)$. The IC shifts upward the sequence of forecasts of $\Delta(m - p)$, but still underestimates the resulting growth, and hence the level. Equally, the improvement in the UK's rate of output growth from 0.62% per quarter over the estimation sample to 1.17% over the forecast period leads to substantial under-prediction of the final level for the i equation with c_2 .

Thus, these outcomes are all in line with the theory.

9 Empirical forecast-accuracy comparison

The bias and variance formulae developed above for the various forecasting models in the presence of structural breaks suggest a number of implications that can be confronted with data. In the preceding section, several of these have already been borne out by the empirical example, and depicted graphically. In this section, we undertake an empirical forecast-accuracy comparison. We consider the pre- and post-break forecast periods for six forecasting methods – the unrestricted system (denoted SYS), the $l(0)$ -VAR (SYS $l(0)$), the simultaneous-equations model (SEM), the VAR in differences (DV), the VAR in second differences (DDV), and the intercept-corrected model (IC) – computing forecast biases, error variances and MSFEs for 1- and 4-step ahead forecasts.³ The IC strategy is that described in the previous section, but we also consider a strategy that more closely reflects the analysis in, e.g., Clements

³Notwithstanding that MSFEs can be problematic for evaluating multi-step forecasts from systems: see Clements and Hendry (1993).

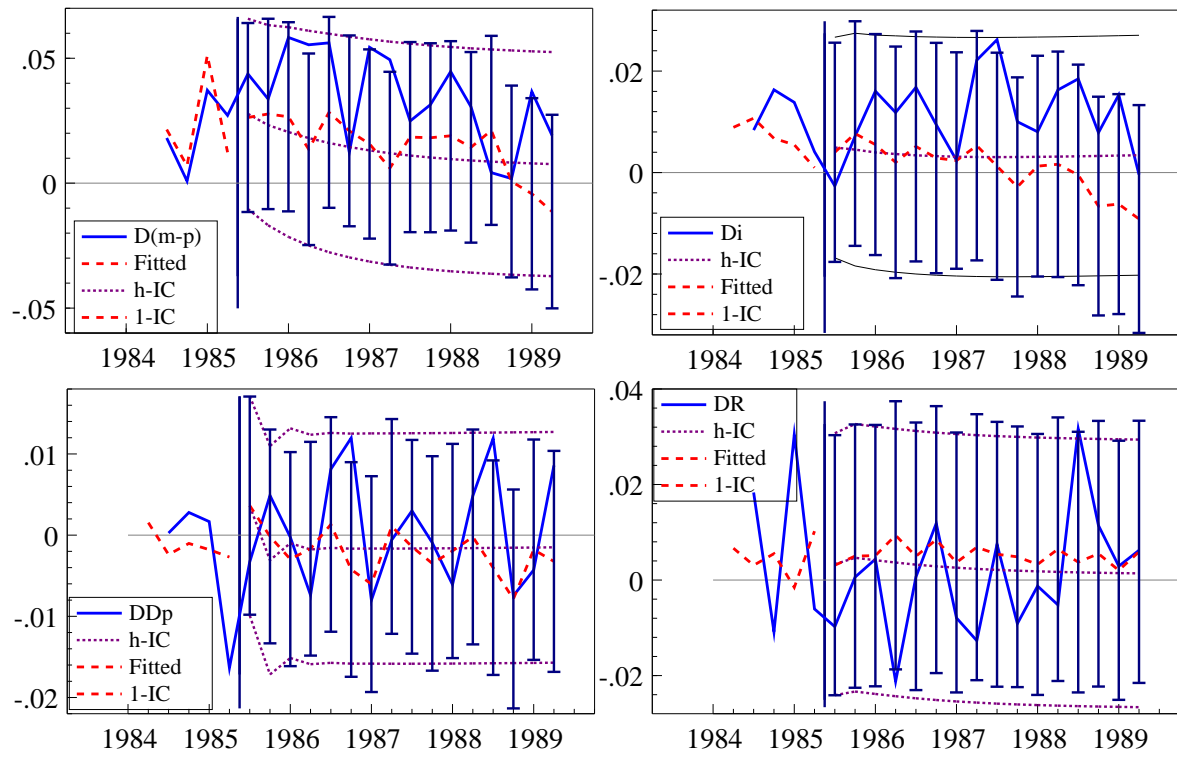


Figure 12 $l(0)$ forecasts from the VEqCM with IC.

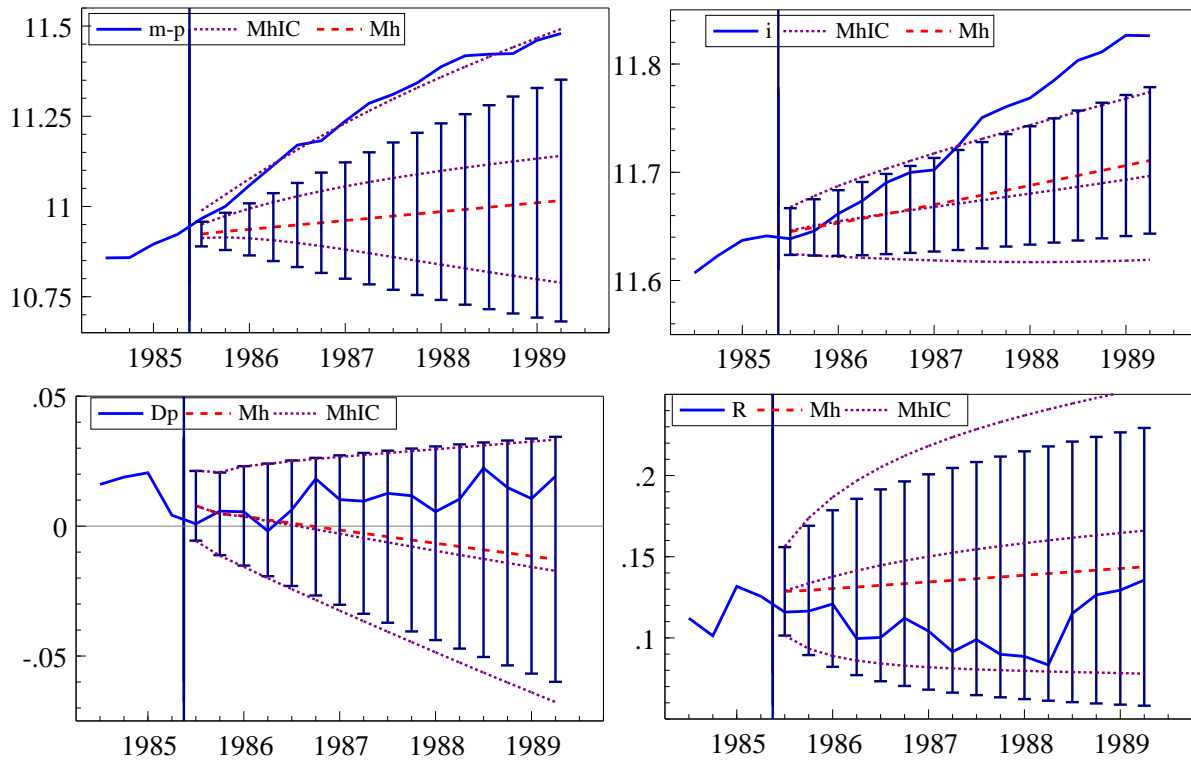


Figure 13 $l(1)$ forecasts from the VEqCM with IC.

and Hendry (1996). This is denoted IC^* . It is implemented by estimating the SEM on data up to each forecast origin, with an indicator variable equal to unity from the immediately prior period onwards. Thus the model estimates are updated as the origin moves through the sample,⁴ and for each origin the correction is an average of the error made at the origin and of the previous error. Alternatively, the correction could have been allowed to depend on the forecast origin, but with the parameter estimates held fixed at their sample period values. This would have disentangled the effects of parameter updating and changing the correction. Since the correction is being tuned to set the forecasts ‘back on track’ as we move through the sample, the IC^* strategy should result in greater bias reduction than the IC case, but at the cost of inflated forecast-error variances, for the reasons described in Clements and Hendry (1996). The IC^* results were only calculated for the post-break forecast period, and then only for $m - p$, since the other equations were not intercept-corrected, so the first-round effects apply only to $m - p$. The results are shown in table 11.

A number of features are worthy of comment.

- (1) In the absence of structural change, the SEM is generally at least as good as the other forecasting models on MSFE at 1- and 4-steps ahead. Moreover, on MSFE, imposing unit roots and cointegration also appears to pay dividends, particularly at 4-steps (compare SYS and SYSI(0)).
- (2) As expected in the absence of structural change, the ‘time-series’ models (DV and DDV) lose out relative to the SEM. Their forecast-error variances are larger, particularly that of the DDV, especially at 4-steps ahead.
- (3) On the post-break sample, however, many of these findings are reversed. In terms of forecasting $m - p$, the SEM is clearly dominated on MSFE by the DV, the DDV, and the IC^* -strategy, at both horizons.
- (4) The DDV and IC^* have the smallest biases for $m - p$.
- (5) However, the DDV and IC^* have the largest error variances for forecasting $m - p$, again at both horizons.
- (6) The IC has a relatively negligible impact on variance compared to the IC^* , but fails to reduce the SEM bias to the same degree as the IC^* strategy.

10 Conclusion

In earlier research, we obtained a taxonomy of forecast errors which suggested that deterministic shifts were a primary determinant of serious forecast failure. We therefore conjectured that a major factor determining the outcome of forecasting competitions on economic time series was the occurrence of deterministic shifts immediately prior to some of the forecast-evaluation periods. Thus, the length of the forecast horizon also should matter. Our theory predicted that differencing several times would be most effective, when multiple breaks have occurred, for evaluation based on many short horizons. A scalar empirical case was investigated in Clements and Hendry (1998) and found to be consistent with that theory.

The present paper extends the analysis to a multivariate context. Taking the data generation process as a vector equilibrium-correction mechanism subject to deterministic shifts, we are able to derive formulae for the conditional forecast biases and unconditional variances. In the four-variable empirical example of forecasting using a small monetary model of the UK when there was a known major financial innovation, the theoretical results matched the outcomes over both pre- and post-break forecast

⁴The model specification was held constant over time, and the parameter constraints of the SEM were imposed except for the forecast origins of 88(2) to 89(1), when the model failed to solve with the constraints imposed.

Table 11 Empirical forecast accuracy comparisons.

Model	m-p		i		Δp		R	
	1-step	4-step	1-step	4-step	1-step	4-step	1-step	4-step
Bias ($\times 100$)								
Post-break forecast period								
SYS	3.288	17.395	0.740	5.097	0.417	1.178	-2.974	-8.621
SYSI(0)	4.749	14.255	1.294	3.927	0.126	0.809	-0.360	-0.139
SEM	4.522	14.010	1.069	3.696	0.296	0.941	-0.445	-0.369
DV	1.995	10.889	0.149	2.312	0.197	0.766	-0.425	-0.857
DDV	-0.091	-2.141	-0.022	0.715	0.134	0.735	0.013	1.383
IC	1.973	7.221	1.070	3.832	0.300	0.983	-0.461	-1.335
IC*	0.259	2.057						
Pre-break forecast period								
SYS	-0.079	3.538	0.642	2.990	0.240	0.469	-1.353	-5.213
SYSI(0)	0.274	3.052	0.165	0.784	-0.109	-0.208	-0.549	-2.886
SEM	-0.043	1.105	0.165	0.539	-0.018	0.108	-0.260	-1.952
DV	0.724	4.573	0.241	1.173	0.048	0.289	-0.249	-1.312
DDV	0.194	1.195	0.069	0.317	-0.096	0.130	0.050	0.516
IC	-0.989	-1.194	0.164	0.595	-0.017	0.124	-0.261	-2.396
Standard error ($\times 100$)								
Post-break forecast period								
SYS	1.826	2.243	0.695	1.334	0.685	0.778	1.261	2.070
SYSI(0)	1.728	3.990	0.874	1.074	0.661	0.655	1.218	2.657
SEM	1.794	3.957	0.805	0.970	0.686	0.681	1.248	2.585
DV	1.877	4.645	0.753	0.933	0.690	0.680	1.245	2.537
DDV	1.921	5.099	1.035	2.617	0.995	2.001	1.561	3.704
IC	1.791	3.967	0.806	0.971	0.686	0.682	1.250	2.588
IC*	2.093	5.386						
Pre-break forecast period								
SYS	1.783	3.502	0.719	1.014	0.815	0.580	1.711	2.175
SYSI(0)	1.781	2.431	0.742	0.627	0.873	0.612	1.526	1.802
SEM	1.660	2.747	0.714	0.675	0.700	0.631	1.541	2.063
DV	1.499	3.043	0.777	0.970	0.725	0.674	1.542	2.431
DDV	1.810	5.270	1.222	3.051	1.106	2.042	1.951	5.064
IC	1.666	2.745	0.714	0.675	0.700	0.631	1.540	2.064
MSFE ($\times 1000$)								
Post-break forecast period								
SYS	1.415	30.763	0.103	2.776	0.064	0.199	1.043	7.860
SYSI(0)	2.553	21.914	0.244	1.657	0.045	0.108	0.161	0.708
SEM	2.367	21.194	0.179	1.460	0.056	0.135	0.176	0.682
DV	0.750	14.014	0.059	0.621	0.052	0.105	0.173	0.717
DDV	0.370	3.058	0.107	0.736	0.101	0.454	0.244	1.563
IC	0.710	6.788	0.179	1.562	0.056	0.143	0.178	0.848
IC*	0.445	3.324						
Pre-break forecast period								
SYS	0.318	2.478	0.093	0.997	0.072	0.056	0.476	3.190
SYSI(0)	0.325	1.522	0.058	0.101	0.077	0.042	0.263	1.158
SEM	0.276	0.877	0.054	0.075	0.049	0.041	0.244	0.807
DV	0.277	3.017	0.066	0.232	0.053	0.054	0.244	0.763
DDV	0.331	2.920	0.150	0.941	0.123	0.419	0.381	2.591
IC	0.375	0.896	0.054	0.081	0.049	0.041	0.244	1.000

periods by the three methods considered, and were supported by the effectiveness of the two intercept-correction strategies investigated. These findings are also consistent with the results of using a large macro-econometric system as reported in Eitrheim *et al.* (1997). Thus, although other forms of structural break, model mis-specification, a lack of parsimony including failing to impose restrictions such as unit roots and cointegration, inaccurate forecast-origin data, and inefficient estimation may all exacerbate forecast failure, we believe they play supporting roles, as in Hendry and Doornik (1997).

One consequence of these findings is to caution against automatic adoption of a method that happens to win a particular forecasting competition: careful analysis as to why is strongly recommended. If a deterministic shift is suspected, or confirmed, then methods that are not robust to such a shift are likely to have performed poorly. That would not necessarily preclude the continued use of a model that suffered forecast failure, subject to appropriate intercept corrections being used. In the policy context of forecasting after a structural break, where policy is then altered, Hendry and Mizon (1998) show the benefits of using a ‘time-series’ model (such as the DDV) to intercept correct the econometric system, when the latter is retained for policy analysis despite forecast failure.

References

- Banerjee, A., Dolado, J. J., Galbraith, J. W., and Hendry, D. F. (1993). *Co-integration, Error Correction and the Econometric Analysis of Non-Stationary Data*. Oxford: Oxford University Press.
- 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. (1995). Forecasting in cointegrated systems. *Journal of Applied Econometrics*, **10**, 127–146.
- 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. (1998). Forecasting economic processes. *International Journal of Forecasting*, **14**, 111–131.
- Clements, M. P., and Mizon, G. E. (1991). Empirical analysis of macroeconomic time series: VAR and structural models. *European Economic Review*, **35**, 887–932.
- Doornik, J. A., and Hansen, H. (1994). A practical test for univariate and multivariate normality. Discussion paper, Nuffield College.
- Doornik, J. A., and Hendry, D. F. (1996). *GiveWin: An Interactive Empirical Modelling Program*. London: Timberlake Consultants Press.
- Doornik, J. A., and Hendry, D. F. (1997). *Modelling Dynamic Systems using PcFiml 9 for Windows*. London: Timberlake Consultants Press.
- Doornik, J. A., Hendry, D. F., and Nielsen, B. (1998). Inference in cointegrated models: UK M1 revisited. Mimeo, Nuffield College, University of Oxford. forthcoming, *Journal of Economic Surveys*.
- Eitrheim, Ø., Husebø, T. A., and Nymoen, R. (1997). Error-correction versus differencing in macro-econometric forecasting. Mimeo, Department of Economics, University of Oslo.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflations. *Econometrica*, **50**, 987–1007.

- Ericsson, N. R., Hendry, D. F., and Tran, H.-A. (1994). Cointegration, seasonality, encompassing and the demand for money in the United Kingdom. In Hargreaves, C. (ed.), *Non-stationary Time-series Analysis and Cointegration*, pp. 179–224. Oxford: Oxford University Press.
- Godfrey, L. G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica*, **46**, 1303–1313.
- Hansen, H., and Johansen, S. (1992). Recursive estimation in cointegrated VAR-models. Discussion paper, Institute of Mathematical Statistics, University of Copenhagen.
- 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. (1996). On the constancy of time-series econometric equations. *Economic and Social Review*, **27**, 401–422.
- Hendry, D. F., and Clements, M. P. (1994). On a theory of intercept corrections in macro-economic forecasting. In Holly, S. (ed.), *Money, Inflation and Employment: Essays in Honour of James Ball*, pp. 160–182. Aldershot: Edward Elgar.
- Hendry, D. F., and Clements, M. P. (1998). 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 Doornik, J. A. (1994). Modelling linear dynamic econometric systems. *Scottish Journal of Political Economy*, **41**, 1–33.
- Hendry, D. F., and Doornik, J. A. (1997). The implications for econometric modelling of forecast failure. *Scottish Journal of Political Economy*, **44**, 437–461. Special Issue.
- Hendry, D. F., and Ericsson, N. R. (1991). Modeling the demand for narrow money in the United Kingdom and the United States. *European Economic Review*, **35**, 833–886.
- 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. (1998). Selecting econometric models for policy analysis by forecast accuracy. Mimeo, Nuffield College, University of Oxford.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, **12**, 231–254.
- Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.
- 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.
- White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817–838.

11 Appendix A: Chow test derivation

Imposing an intercept of zero, so $\phi_1 = \phi$, and defining the $T \times T$ matrix \mathbf{D} :

$$\mathbf{D} = \begin{pmatrix} \mathbf{0}_{T_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{T_2} \end{pmatrix} \text{ so } \mathbf{D}\mathbf{z} = \begin{pmatrix} \mathbf{0} \\ \mathbf{z}_{T_2} \end{pmatrix},$$

as $T_1 + T_2 = T$, the DGP becomes:

$$\mathbf{y} = \mathbf{z}\phi + \mathbf{D}\mathbf{z}\delta + \boldsymbol{\epsilon}$$

whereas the estimate of ϕ from (8) (again without an intercept) is:

$$\hat{\phi} = (\mathbf{z}'\mathbf{z})^{-1} \mathbf{z}'\mathbf{y} = \phi + (\mathbf{z}'\mathbf{z})^{-1} \mathbf{z}'\boldsymbol{\epsilon} + \delta (\mathbf{z}'\mathbf{z})^{-1} (\mathbf{z}'_{T_2} \mathbf{z}_{T_2})$$

so:

$$\mathbb{E} [\hat{\phi}] = \phi + \delta k.$$

Next, using:

$$\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \hat{\mathbf{y}} = \boldsymbol{\epsilon} + \mathbf{D}\mathbf{z}\delta - \mathbf{z} \left(\hat{\phi} - \phi \right) = \mathbf{Q}_z \boldsymbol{\epsilon} + \delta \left[\mathbf{D} - (\mathbf{z}'\mathbf{z})^{-1} (\mathbf{z}'_{T_2} \mathbf{z}_{T_2}) \right] \mathbf{z}$$

with $\mathbf{Q}_z = (\mathbf{I}_T - \mathbf{z} (\mathbf{z}'\mathbf{z})^{-1} \mathbf{z}')$, then:

$$\mathbb{E} [\hat{\boldsymbol{\epsilon}}] = \delta \eta (\mathbf{D} - k \mathbf{I}_T)$$

and hence:

$$\mathbb{E} [\hat{\boldsymbol{\epsilon}}' \hat{\boldsymbol{\epsilon}}] \simeq (T-1) \sigma_{\epsilon}^2 + \delta^2 (\sigma_z^2 + \eta^2) (1-k) T_2.$$

Sub-sample values can be obtained by specialization.

12 Appendix B: Bias derivations

12.1 From T to $T+1$

From (9), (10), and (11) respectively:

$$\hat{\mathbf{e}}_{T+1|T} = \gamma^* - \gamma + \boldsymbol{\alpha}^* (\boldsymbol{\beta}' \mathbf{x}_T - \boldsymbol{\mu}^*) - \boldsymbol{\alpha} (\boldsymbol{\beta}' \mathbf{x}_T - \boldsymbol{\mu}) + \boldsymbol{\nu}_{T+1}; \quad (57)$$

$$\tilde{\mathbf{e}}_{T+1|T} = \mathbf{x}_{T+1} - \tilde{\mathbf{x}}_{T+1|T} = \gamma^* - \gamma + \boldsymbol{\alpha}^* (\boldsymbol{\beta}' \mathbf{x}_T - \boldsymbol{\mu}^*) + \boldsymbol{\nu}_{T+1}, \quad (58)$$

and:

$$\bar{\mathbf{e}}_{T+1|T} = \mathbf{x}_{T+1} - \bar{\mathbf{x}}_{T+1|T} = \gamma^* + \boldsymbol{\alpha}^* (\boldsymbol{\beta}' \mathbf{x}_T - \boldsymbol{\mu}^*) + \boldsymbol{\nu}_{T+1} - \Delta \mathbf{x}_T. \quad (59)$$

Let $\boldsymbol{\beta}' \mathbf{x}_T - \boldsymbol{\mu} = \boldsymbol{\kappa}_T$, then the *conditional* biases are:

$$\mathbb{E} [\hat{\mathbf{e}}_{T+1|T} \mid I_T] = (\gamma^* - \gamma) - \boldsymbol{\alpha}^* (\boldsymbol{\mu}^* - \boldsymbol{\mu}) + (\boldsymbol{\alpha}^* - \boldsymbol{\alpha}) \boldsymbol{\kappa}_T, \quad (60)$$

$$\mathbb{E} [\tilde{\mathbf{e}}_{T+1|T} \mid I_T] = (\gamma^* - \gamma) - \boldsymbol{\alpha}^* (\boldsymbol{\mu}^* - \boldsymbol{\mu}) + \boldsymbol{\alpha}^* \boldsymbol{\kappa}_T, \quad (61)$$

and:

$$\mathbb{E} [\bar{\mathbf{e}}_{T+1|T} \mid I_T] = (\gamma^* - \Delta \mathbf{x}_T) - \boldsymbol{\alpha}^* (\boldsymbol{\mu}^* - \boldsymbol{\mu}) + \boldsymbol{\alpha}^* \boldsymbol{\kappa}_T. \quad (62)$$

12.2 From $T + 1$ to $T + 2$

The three 1-step forecasts become:

$$\Delta \hat{\mathbf{x}}_{T+2|T+1} = \gamma + \alpha (\beta' \mathbf{x}_{T+1} - \mu) \quad (63)$$

$$\Delta \tilde{\mathbf{x}}_{T+2|T+1} = \gamma, \quad (64)$$

and:

$$\Delta \bar{\mathbf{x}}_{T+2|T+1} = \Delta \mathbf{x}_{T+1}, \quad (65)$$

with corresponding 1-step forecast errors:

$$\hat{\mathbf{e}}_{T+2|T+1} = \gamma^* - \gamma + \alpha^* (\beta' \mathbf{x}_{T+1} - \mu^*) - \alpha (\beta' \mathbf{x}_{T+1} - \mu) + \nu_{T+2} \quad (66)$$

$$\tilde{\mathbf{e}}_{T+2|T+1} = \gamma^* - \gamma + \alpha^* (\beta' \mathbf{x}_{T+1} - \mu^*) + \nu_{T+2} \quad (67)$$

and:

$$\bar{\mathbf{e}}_{T+2|T+1} = \gamma^* + \alpha^* (\beta' \mathbf{x}_{T+1} - \mu^*) + \nu_{T+2} - \Delta \mathbf{x}_{T+1}, \quad (68)$$

from which the conditional biases follow as cited in the text.

12.3 From T to $T + 2$

The data are generated by:

$$\mathbf{x}_{T+2} = (\Upsilon^*)^2 \mathbf{x}_T + (\mathbf{I}_n + \Upsilon^*) (\gamma^* - \alpha^* \mu^*) + \Upsilon^* \nu_{T+1} + \nu_{T+2} \quad (69)$$

as:

$$\mathbf{x}_{T+1} = \Upsilon^* \mathbf{x}_T + (\gamma^* - \alpha^* \mu^*) + \nu_{T+1}. \quad (70)$$

Also

$$\hat{\mathbf{x}}_{T+2|T} = \Upsilon \hat{\mathbf{x}}_{T+1|T} + \gamma - \alpha \mu = \Upsilon^2 \mathbf{x}_T + (\mathbf{I}_n + \Upsilon) (\gamma - \alpha \mu); \quad (71)$$

$$\tilde{\mathbf{x}}_{T+2|T} = \mathbf{x}_T + 2\gamma; \quad (72)$$

and:

$$\bar{\mathbf{x}}_{T+2|T} = \bar{\mathbf{x}}_{T+1|T} + \Delta \bar{\mathbf{x}}_{T+1|T} = \mathbf{x}_T + 2\Delta \mathbf{x}_T. \quad (73)$$

After some simplification, using $\Upsilon \gamma = \gamma$ and $\Upsilon^* \gamma^* = \gamma^*$, and:

$$\begin{aligned} (\Upsilon^*)^2 - \Upsilon^2 &= ((\mathbf{I}_n + \Upsilon^*) \alpha^* - (\mathbf{I}_n + \Upsilon) \alpha) \beta' \\ &= [\alpha^* (\mathbf{I}_r + \Lambda^*) - \alpha (\mathbf{I}_r + \Lambda)] \beta' \end{aligned} \quad (74)$$

the corresponding 2-step ahead forecast errors are:

$$\begin{aligned} \hat{\mathbf{e}}_{T+2|T} &= \left[(\Upsilon^*)^2 - \Upsilon^2 \right] \mathbf{x}_T + 2(\gamma^* - \gamma) - (\mathbf{I}_n + \Upsilon^*) \alpha^* \mu^* - (\mathbf{I}_n + \Upsilon) \alpha \mu + \Upsilon^* \nu_{T+1} + \nu_{T+2} \\ &= \mathbf{E} [\hat{\mathbf{e}}_{T+2|T}] + (\alpha^* [\mathbf{I}_r + \Lambda^*] - \alpha [\mathbf{I}_r + \Lambda]) \kappa_T + \Upsilon^* \nu_{T+1} + \nu_{T+2} \end{aligned} \quad (75)$$

where:

$$\mathbf{E} [\hat{\mathbf{e}}_{T+2|T}] = 2(\gamma^* - \gamma) - \alpha^* (\mathbf{I}_r + \Lambda^*) (\mu^* - \mu),$$

and as $(\Upsilon^*)^2 \mathbf{x}_T - \mathbf{x}_T = \alpha^* (2\mathbf{I}_r + \beta' \alpha^*) \beta' \mathbf{x}_T = \alpha^* (\mathbf{I}_r + \Lambda^*) \beta' \mathbf{x}_T$ then:

$$\begin{aligned} \tilde{\mathbf{e}}_{T+2|T} &= \alpha^* (2\mathbf{I}_r + \beta' \alpha^*) \kappa_T + 2(\gamma^* - \gamma) - \alpha^* (\mathbf{I}_r + \Lambda^*) (\mu^* - \mu) + \Upsilon^* \nu_{T+1} + \nu_{T+2} \\ &= \mathbf{E} [\tilde{\mathbf{e}}_{T+2|T}] + \alpha^* (\mathbf{I}_r + \Lambda^*) \kappa_T + \Upsilon^* \nu_{T+1} + \nu_{T+2}, \end{aligned} \quad (76)$$

and:

$$\begin{aligned}\bar{\mathbf{e}}_{T+2|T} &= \boldsymbol{\alpha}^* (2\mathbf{I}_r + \boldsymbol{\beta}' \boldsymbol{\alpha}^*) \boldsymbol{\kappa}_T + 2(\gamma^* - \gamma) - \boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) (\boldsymbol{\mu}^* - \boldsymbol{\mu}) \\ &\quad + \boldsymbol{\Upsilon}^* \boldsymbol{\nu}_{T+1} + \boldsymbol{\nu}_{T+2} - 2(\Delta \mathbf{x}_T - \gamma) \\ &= \mathbb{E} [\bar{\mathbf{e}}_{T+2|T}] + \boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) \boldsymbol{\kappa}_T - 2\boldsymbol{\alpha} \boldsymbol{\kappa}_{T-1} - 2\boldsymbol{\nu}_T + \boldsymbol{\Upsilon}^* \boldsymbol{\nu}_{T+1} + \boldsymbol{\nu}_{T+2},\end{aligned}\quad (77)$$

as

$$\Delta \mathbf{x}_T - \gamma = \boldsymbol{\alpha} \boldsymbol{\kappa}_{T-1} + \boldsymbol{\nu}_T.$$

12.4 From $T + 1$ to $T + 3$

First, the two-step derivations closely match those from T to $T + 2$ shifted forward one period:

$$\mathbf{x}_{T+3} = (\boldsymbol{\Upsilon}^*)^2 \mathbf{x}_{T+1} + (\mathbf{I}_n + \boldsymbol{\Upsilon}^*) (\gamma^* - \boldsymbol{\alpha}^* \boldsymbol{\mu}^*) + \boldsymbol{\Upsilon}^* \boldsymbol{\nu}_{T+2} + \boldsymbol{\nu}_{T+3} \quad (78)$$

so that:

$$\boldsymbol{\beta}' \mathbf{x}_{T+3} = (\mathbf{I}_r + 2\boldsymbol{\beta}' \boldsymbol{\alpha}^* + \boldsymbol{\beta}' \boldsymbol{\alpha}^* \boldsymbol{\beta}' \boldsymbol{\alpha}^*) \boldsymbol{\beta}' \mathbf{x}_{T+1} + \boldsymbol{\beta}' (\mathbf{I}_n + \boldsymbol{\Upsilon}^*) (\gamma^* - \boldsymbol{\alpha}^* \boldsymbol{\mu}^*) \quad (79)$$

with:

$$\boldsymbol{\beta}' \mathbf{x}_{T+2} = (\mathbf{I}_r + 2\boldsymbol{\beta}' \boldsymbol{\alpha}^* + \boldsymbol{\beta}' \boldsymbol{\alpha}^* \boldsymbol{\beta}' \boldsymbol{\alpha}^*) \boldsymbol{\beta}' \mathbf{x}_T + \boldsymbol{\beta}' (\mathbf{I}_n + \boldsymbol{\Upsilon}^*) (\gamma^* - \boldsymbol{\alpha}^* \boldsymbol{\mu}^*). \quad (80)$$

Also:

$$\hat{\mathbf{x}}_{T+3} = \boldsymbol{\Upsilon}^2 \mathbf{x}_{T+1} + (\mathbf{I}_n + \boldsymbol{\Upsilon}) (\gamma - \boldsymbol{\alpha} \boldsymbol{\mu}) \quad (81)$$

$$\tilde{\mathbf{x}}_{T+3} = \mathbf{x}_{T+1} + 2\gamma, \quad (82)$$

and:

$$\bar{\mathbf{x}}_{T+3} = \mathbf{x}_{T+1} + 2\Delta \mathbf{x}_{T+1}. \quad (83)$$

Using (74):

$$\begin{aligned}\hat{\mathbf{e}}_{T+3|T+1} &= \left[(\boldsymbol{\Upsilon}^*)^2 - \boldsymbol{\Upsilon}^2 \right] \mathbf{x}_{T+1} + (\mathbf{I}_n + \boldsymbol{\Upsilon}^*) (\gamma^* - \boldsymbol{\alpha}^* \boldsymbol{\mu}^*) - (\mathbf{I}_n + \boldsymbol{\Upsilon}) (\gamma - \boldsymbol{\alpha} \boldsymbol{\mu}) + \boldsymbol{\Upsilon}^* \boldsymbol{\nu}_{T+2} + \boldsymbol{\nu}_{T+3} \\ &= [\boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) - \boldsymbol{\alpha} (\mathbf{I}_r + \boldsymbol{\Lambda})] \boldsymbol{\beta}' \mathbf{x}_{T+1} + 2(\gamma^* - \gamma) - \boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) \boldsymbol{\mu}^* + \boldsymbol{\alpha} (\mathbf{I}_r + \boldsymbol{\Lambda}) \boldsymbol{\mu} \\ &\quad + \boldsymbol{\Upsilon}^* \boldsymbol{\nu}_{T+2} + \boldsymbol{\nu}_{T+3} \\ &= 2(\gamma^* - \gamma) - \boldsymbol{\alpha} (\mathbf{I}_r + \boldsymbol{\Lambda}) (\boldsymbol{\mu}^* - \boldsymbol{\mu}) + [\boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) - \boldsymbol{\alpha} (\mathbf{I}_r + \boldsymbol{\Lambda})] \boldsymbol{\kappa}_{T+1}^* + \boldsymbol{\Upsilon}^* \boldsymbol{\nu}_{T+2} + \boldsymbol{\nu}_{T+3} \\ &= \mathbb{E} [\hat{\mathbf{e}}_{T+3|T+1}] + [\boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) - \boldsymbol{\alpha} (\mathbf{I}_r + \boldsymbol{\Lambda})] (\boldsymbol{\kappa}_{T+1}^* - \mathbb{E} [\boldsymbol{\kappa}_{T+1}^*]) + \boldsymbol{\Upsilon}^* \boldsymbol{\nu}_{T+2} + \boldsymbol{\nu}_{T+3},\end{aligned}$$

as:

$$\mathbb{E} [\boldsymbol{\kappa}_{T+1}^*] = -\boldsymbol{\Lambda}^* (\boldsymbol{\mu}^* - \boldsymbol{\mu}),$$

where:

$$\mathbb{E} [\hat{\mathbf{e}}_{T+3|T+1}] = 2(\gamma^* - \gamma) - [\boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) \boldsymbol{\Lambda}^* - \boldsymbol{\alpha} (\mathbf{I}_r + \boldsymbol{\Lambda}) \boldsymbol{\beta}' \boldsymbol{\alpha}^*] (\boldsymbol{\mu}^* - \boldsymbol{\mu}).$$

Next:

$$\begin{aligned}\tilde{\mathbf{e}}_{T+3|T+1} &= \left[(\boldsymbol{\Upsilon}^*)^2 - \mathbf{I}_n \right] \mathbf{x}_{T+1} + (\mathbf{I}_n + \boldsymbol{\Upsilon}^*) (\gamma^* - \boldsymbol{\alpha}^* \boldsymbol{\mu}^*) + \boldsymbol{\Upsilon}^* \boldsymbol{\nu}_{T+2} + \boldsymbol{\nu}_{T+3} - 2\gamma \\ &= \boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) \boldsymbol{\beta}' \mathbf{x}_{T+1} + 2(\gamma^* - \gamma) - \boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) \boldsymbol{\mu}^* + \boldsymbol{\Upsilon}^* \boldsymbol{\nu}_{T+2} + \boldsymbol{\nu}_{T+3} \\ &= 2(\gamma^* - \gamma) + \boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) \boldsymbol{\kappa}_{T+1}^* + \boldsymbol{\Upsilon}^* \boldsymbol{\nu}_{T+2} + \boldsymbol{\nu}_{T+3} \\ &= \mathbb{E} [\tilde{\mathbf{e}}_{T+3|T+1}] + \boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) (\boldsymbol{\kappa}_{T+1}^* - \mathbb{E} [\boldsymbol{\kappa}_{T+1}^*]) + \boldsymbol{\Upsilon}^* \boldsymbol{\nu}_{T+2} + \boldsymbol{\nu}_{T+3}\end{aligned}$$

as:

$$\mathbb{E} [\tilde{\mathbf{e}}_{T+3|T+1}] = 2(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}) - \boldsymbol{\alpha}^* \boldsymbol{\Lambda}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) (\boldsymbol{\mu}^* - \boldsymbol{\mu}).$$

Finally, the DDV forecast-error is:

$$\begin{aligned} \bar{\mathbf{e}}_{T+3|T+1} &= \left[(\boldsymbol{\Upsilon}^*)^2 - \mathbf{I}_n \right] \mathbf{x}_{T+1} + (\mathbf{I}_n + \boldsymbol{\Upsilon}^*) (\boldsymbol{\gamma}^* - \boldsymbol{\alpha}^* \boldsymbol{\mu}^*) + \boldsymbol{\Upsilon}^* \boldsymbol{\nu}_{T+2} + \boldsymbol{\nu}_{T+3} - 2\Delta \mathbf{x}_{T+1} \\ &= 2(\boldsymbol{\gamma}^* - \Delta \mathbf{x}_{T+1}) + \boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) \boldsymbol{\kappa}_{T+1}^* + \boldsymbol{\Upsilon}^* \boldsymbol{\nu}_{T+2} + \boldsymbol{\nu}_{T+3} \\ &= \mathbb{E} [\bar{\mathbf{e}}_{T+3|T+1}] + 2(\mathbb{E} [\Delta \mathbf{x}_{T+1}] - \Delta \mathbf{x}_{T+1}) + \boldsymbol{\alpha}^* (\mathbf{I}_r + \boldsymbol{\Lambda}^*) (\boldsymbol{\kappa}_{T+1}^* - \mathbb{E} [\boldsymbol{\kappa}_{T+1}^*]) \\ &\quad + \boldsymbol{\Upsilon}^* \boldsymbol{\nu}_{T+2} + \boldsymbol{\nu}_{T+3}, \end{aligned}$$

where:

$$\mathbb{E} [\bar{\mathbf{e}}_{T+3|T+1}] = -\boldsymbol{\alpha}^* (2\mathbf{I}_r + \boldsymbol{\Lambda}^*) \boldsymbol{\beta}' \boldsymbol{\alpha}^* (\boldsymbol{\mu}^* - \boldsymbol{\mu}).$$