Robustifying Forecasts from Equilibrium-Correction Models

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

April 26, 2004

Abstract

In a non-stationary world subject to structural breaks, where model and mechanism differ, equilibrium-correction models are a risky device from which to forecast. Equilibrium shifts entail systematic forecast failure, and indeed forecasts will tend to move in the opposite direction to the data. A new explanation for the empirical success of second differencing is proposed. We consider model transformations based on additional differencing to reduce forecast-error biases, as usual at some cost in increased forecast-error variances. The analysis is illustrated by an empirical application to narrow money holdings in the UK.

Contents

1 Introduction . . . 1
2 A cointegrated DGP . . . 2
2.1 Forecasting properties . . . 3
3 Location shifts . . . 5
4 Forecasting by $\Delta \mathbf{x}_T$ . . . 6
4.1 Constant-parameter case . . . 7
4.1.1 Scalar illustration 1 . . . 8
4.2 Changed-parameter case . . . 9
4.2.1 Scalar illustration 2 . . . 10
4.3 Longer-period differences . . . 11
5 Forecasting from a transformed VEqCM . . . 11
5.1 Differencing the VEqCM . . . 12
5.1.1 Forecast-error variances . . . 13
6 Empirical illustration: UK M1 . . . 13
6.1 Single-equation results . . . 14
6.2 System behaviour . . . 16
7 Conclusions . . . 17
References . . . 18

1 Introduction

Developments in cointegration analysis from Granger (1981), through Granger and Weiss (1983) and Engle and Granger (1987), to Johansen (1988) have led to equilibrium-correction econometric systems

*Financial support from the ESRC under grant RES 015 270035, and helpful comments from Jennifer Castle, Mike Clements, Neil Ericsson, Katarina Juselius, and Grayham Mizon are gratefully acknowledged.

being ubiquitous for modelling, forecasting and economic policy analysis. In fact, most econometric models are members of the equilibrium-correction class: this includes not only explicit equilibrium-correction models (denoted EqCMs) based on cointegration, and almost all regression equations and simultaneous models, but also most other econometric systems, including vector autoregressions (VARs), dynamic stochastic general-equilibrium models (DSGEs) and many variance models (such as GARCH). The forecasting properties of this huge class are essentially generic, and are well represented by those of standard vector EqCMs (VEqCMs: see e.g., Hendry, 2003).

Initially, theory and Monte Carlo simulations suggested VEqCMs should outperform when forecasting, especially for cointegrated combinations of variables: see e.g., Engle and Yoo (1987), Lütkepohl (1991) and Clements and Hendry (1995). However, the findings of forecasting competitions (see e.g., Makridakis and Hibon, 2000, Clements and Hendry, 2001, and Fildes and Ord, 2002), extensive applications to forecasting macro time series as in Stock and Watson (1999), and empirical mis-forecasting of events, such as money demand in the UK (see Hendry and Mizon, 1993) and UK consumers' expenditure (see e.g., Clements and Hendry, 1998a) suggested that all was not well. The theory of forecasting from mis-specified models of non-stationary processes subject to structural breaks in Clements and Hendry (1998b, 1999) highlighted that VEqCMs were not robust to shifts in the underlying equilibrium. The results in Hendry and Doornik (1997) and Hendry (2000) showed that location shifts (such as changes in equilibria) were the most pernicious problem for forecasting in this class. Indeed, following an equilibrium shift, forecasts from VEqCMs tended to move in the opposite direction to the data, thereby inducing forecast failure, defined as a significant deterioration in forecast performance relative to in-sample behaviour. Finally, the prevalence of structural changes in macro time series confirmed in Stock and Watson (1996) helped account for such outcomes, including those reported in Stock and Watson (1999).

Moreover, Clements and Hendry (1998b, 1999) show that an economic theory causal basis for forecasting models is of no avail in a world of location shifts. Thus, while VEqCMs are excellent when the process is stationary after differencing and cointegration reductions, they are unreliable if breaks occur. Consequently, we consider model transformations which can reduce forecast-error biases from systematic mis-forecasting by VEqCMs, as usual at some cost in increased forecast-error variances (other adaptive approaches, and the basis for these, are discussed in Hendry, 2003). We also thereby discover a new explanation for why some so-called 'naive' forecasting devices may be hard to outperform even when they are apparently poor approximations to the in-sample data generation process (DGP).

Section 2 specifies a cointegrated DGP and its properties as a forecasting device, then section 3 considers the effects thereon when breaks occur. Section 4 discusses why 'second-differenced' forecasting devices may perform well in processes subject to structural breaks; and section 5 examines a transformation which might improve the robustness of VEqCMs when forecasting in such a context. Section 6 illustrates the ideas for the much-used empirical example of the behaviour of UK M1. Section 7 concludes.

2 A cointegrated DGP

We consider a first-order VAR for simplicity, where the vector of n variables of interest is denoted by $\mathbf{x}_t$ (often taken to be the logs of the original variables), and its in-sample DGP is:

$$\mathbf{x}_t = \boldsymbol{\tau} + \boldsymbol{\Gamma}\mathbf{x}_{t-1} + \boldsymbol{\epsilon}_t \text{ where } \boldsymbol{\epsilon}_t \sim \mathsf{IN}_n\left[\mathbf{0}, \boldsymbol{\Omega}_{\epsilon}\right]. \tag{1}$$

$\boldsymbol{\Gamma}$ is an $n \times n$ matrix of coefficients and $\boldsymbol{\tau}$ is an n dimensional vector of intercepts. The specification in (1) is assumed constant in-sample, and the system is taken to be $\mathsf{I}(1)$, satisfying the $r < n$ cointegration relations:

$$\mathbf{\Gamma} = \mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}'. \tag{2}$$

In (2), $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $n \times r$ full-rank matrices, no roots of $|\mathbf{I} - \mathbf{\Gamma} L| = 0$ lie inside unit circle (where $L^k \mathbf{x}_t = \mathbf{x}_{t-k}$), and $\boldsymbol{\alpha}'_\perp \mathbf{\Gamma} \boldsymbol{\beta}_\perp$ is full rank $(n-r)$, where $\boldsymbol{\alpha}_\perp$ and $\boldsymbol{\beta}_\perp$ are full column rank $n \times (n-r)$ matrices, with $\boldsymbol{\alpha}'\boldsymbol{\alpha}_\perp = \boldsymbol{\beta}'\boldsymbol{\beta}_\perp = \mathbf{0}$ (see e.g., Johansen, 1992). Additional lags do not materially affect the analysis below. Then (1) is reparametrized as the vector equilibrium-correction model (VEqCM):

$$\Delta \mathbf{x}_t = \boldsymbol{\tau} + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \boldsymbol{\epsilon}_t. \tag{3}$$

Both $\Delta \mathbf{x}_t$ and $\boldsymbol{\beta}'\mathbf{x}_t$ are $\mathsf{I}(0)$ but may have non-zero means. Let:

$$\boldsymbol{\tau} = \boldsymbol{\gamma} - \boldsymbol{\alpha}\boldsymbol{\mu} \tag{4}$$

then:

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

When $\boldsymbol{\beta}'\boldsymbol{\alpha}$ is non-singular, the variables grow at the unconditional rate:

$$\mathsf{E}\left[\Delta \mathbf{x}_t\right] = \boldsymbol{\gamma} = \left(\mathbf{I}_n - \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)^{-1}\boldsymbol{\beta}'\right)\boldsymbol{\tau} = \mathbf{K}\boldsymbol{\tau},$$

where $\mathbf{K}$ is non-symmetric idempotent with $\boldsymbol{\beta}'\mathbf{K} = \mathbf{0}'$ and $\mathbf{K}\boldsymbol{\alpha} = \mathbf{0}$ so $\mathbf{\Gamma}\mathbf{K} = \mathbf{K}$ which implies that $\boldsymbol{\beta}'\boldsymbol{\gamma} = \mathbf{0}$ so $\mathbf{\Gamma}\boldsymbol{\gamma} = \boldsymbol{\gamma}$; and the long-run equilibrium mean is:

$$\mathsf{E}\left[\boldsymbol{\beta}'\mathbf{x}_t\right] = \boldsymbol{\mu}. \tag{6}$$

Thus, in (5), both $\Delta \mathbf{x}_t$ and $\boldsymbol{\beta}'\mathbf{x}_t$ are expressed as deviations about their means. Note that $\boldsymbol{\gamma}$ is $n \times 1$, but subject to r restrictions, and $\boldsymbol{\mu}$ is $r \times 1$, leaving n unrestricted intercepts in total in (5). Also, $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\mu}$ are assumed to be variation free, although in principle, $\boldsymbol{\mu}$ could depend on $\boldsymbol{\gamma}$: see Hendry and von Ungern-Sternberg (1981). Then $(\boldsymbol{\tau}, \mathbf{\Gamma})$ are not variation free, as seems reasonable when $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ are the 'deep' parameters: for a more extensive analysis, see Clements and Hendry (1996).

2.1 Forecasting properties

When the parameters are constant in-sample, sampling variations in estimates thereof have only a small effect on the analysis, so we consider the case of known parameters to focus on the issue of forecast failure. In that case, 1-step ahead forecasts from (5) coincide with the conditional expectation $\mathsf{E}_T\left[\Delta \mathbf{x}_{T+1} | \mathbf{x}_T\right]$, and are given by:

$$\widehat{\Delta \mathbf{x}}_{T+1|T} = \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}\right). \tag{7}$$

The h-step forecast errors for the growth rate are $\widehat{\boldsymbol{\epsilon}}_{T+h} = \Delta \mathbf{x}_{T+h} - \Delta \widehat{\mathbf{x}}_{T+h|T}$ where $\widehat{\boldsymbol{\epsilon}}_{T+1} = \boldsymbol{\epsilon}_{T+1}$.

It is easiest to first derive forecast errors $\widetilde{\boldsymbol{\epsilon}}_{T+h} = \mathbf{x}_{T+h} - \widehat{\mathbf{x}}_{T+h|T}$ for the levels:

$$\widehat{\mathbf{x}}_{T+1|T} = \mathbf{x}_T + \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}\right) = \boldsymbol{\tau} + \mathbf{\Gamma}\mathbf{x}_T, \tag{8}$$

so $\widetilde{\boldsymbol{\epsilon}}_{T+1} = \widehat{\boldsymbol{\epsilon}}_{T+1}$. However, the h-step forecast errors from (8) are then generated recursively by:

$$\widehat{\mathbf{x}}_{T+h|T} = \boldsymbol{\tau} + \mathbf{\Gamma}\widehat{\mathbf{x}}_{T+h-1|T} = \sum_{i=0}^{h-1} \mathbf{\Gamma}^i \boldsymbol{\tau} + \mathbf{\Gamma}^h \mathbf{x}_T. \tag{9}$$

As:

$$\mathbf{x}_{T+h} = \boldsymbol{\tau} + \boldsymbol{\Gamma}\mathbf{x}_{T+h-1} + \boldsymbol{\epsilon}_{T+h} = \sum_{i=0}^{h-1}\boldsymbol{\Gamma}^i\boldsymbol{\tau} + \boldsymbol{\Gamma}^h\mathbf{x}_T + \sum_{i=0}^{h-1}\boldsymbol{\Gamma}^i\boldsymbol{\epsilon}_{T+h-i},$$

for known parameters:

$$\widetilde{\boldsymbol{\epsilon}}_{T+h} = \sum_{i=0}^{h-1}\boldsymbol{\Gamma}^i\boldsymbol{\epsilon}_{T+h-i},$$

with:

$$\mathsf{E}\left[\widetilde{\boldsymbol{\epsilon}}_{T+h}\right] = 0 \;\text{ and }\; \mathsf{V}\left[\widetilde{\boldsymbol{\epsilon}}_{T+h}\right] = \sum_{i=0}^{h-1}\boldsymbol{\Gamma}^i\boldsymbol{\Omega}_\epsilon\boldsymbol{\Gamma}^{i\prime} \tag{10}$$

where $\mathsf{V}[\cdot]$ denotes the variance, and is $\mathsf{O}(h)$ in (10) because $\boldsymbol{\Gamma}^i$ increases in i.

Returning to growth rates, since $\Delta\mathbf{x}_{T+h} = \mathbf{x}_{T+h} - \mathbf{x}_{T+h-1}$:

$$\begin{aligned}\Delta\mathbf{x}_{T+h} &= \boldsymbol{\Gamma}^{h-1}\boldsymbol{\tau} + \boldsymbol{\Gamma}^{h-1}\left(\boldsymbol{\Gamma} - \mathbf{I}_n\right)\mathbf{x}_T + \boldsymbol{\epsilon}_{T+h} + \left(\boldsymbol{\Gamma} - \mathbf{I}_n\right)\sum_{i=0}^{h-2}\boldsymbol{\Gamma}^i\boldsymbol{\epsilon}_{T+h-i-1} \\ &= \boldsymbol{\gamma} + \boldsymbol{\alpha}\boldsymbol{\Psi}^{h-1}\left(\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}\right) + \boldsymbol{\epsilon}_{T+h} - \boldsymbol{\alpha}\sum_{i=0}^{h-2}\boldsymbol{\Psi}^i\boldsymbol{\beta}'\boldsymbol{\epsilon}_{T+h-1-i},\end{aligned}$$

where we use the well-known results that (see e.g., Clements and Hendry, 1995):

$$\boldsymbol{\beta}'\boldsymbol{\Gamma} = \boldsymbol{\beta}'\left(\mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}'\right) = \left(\mathbf{I}_n + \boldsymbol{\beta}'\boldsymbol{\alpha}\right)\boldsymbol{\beta}' \mathrel{\hat{=}} \boldsymbol{\Psi}\boldsymbol{\beta}',$$

and:

$$\boldsymbol{\Gamma}\boldsymbol{\alpha} = \left(\mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}'\right)\boldsymbol{\alpha} = \boldsymbol{\alpha}\boldsymbol{\Psi}.$$

Thus:

$$\mathsf{E}\left[\widehat{\boldsymbol{\epsilon}}_{T+h}\right] = 0 \;\text{ and }\; \mathsf{V}\left[\widehat{\boldsymbol{\epsilon}}_{T+h}\right] = \boldsymbol{\Omega}_\epsilon + \sum_{i=0}^{h-2}\boldsymbol{\alpha}\boldsymbol{\Psi}^i\boldsymbol{\beta}'\boldsymbol{\Omega}_\epsilon\boldsymbol{\beta}\boldsymbol{\Psi}^{i\prime}\boldsymbol{\alpha}', \tag{11}$$

where $\mathsf{V}[\cdot]$ denotes the variance, and is $\mathsf{O}(1)$ in h in (11) because $\boldsymbol{\Psi}^i \to \mathbf{0}$ as i increases. Parameter estimation adds terms of $\mathsf{O}(T^{-1})$ to $\mathsf{V}[\cdot]$ for a sample of size T. Note that:

$$\boldsymbol{\Gamma}^h = \mathbf{I}_n + \boldsymbol{\alpha}\sum_{i=0}^{h-1}\boldsymbol{\Psi}^i\boldsymbol{\beta}' = \mathbf{I}_n - \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)^{-1}\left(\mathbf{I}_n - \boldsymbol{\Psi}^h\right)\boldsymbol{\beta}' = \mathbf{K} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)^{-1}\boldsymbol{\Psi}^h\boldsymbol{\beta}',$$

and therefore $\boldsymbol{\Gamma}^h \to \mathbf{K}$ with:

$$\mathbf{x}_{T+h} = \mathbf{x}_T + h\boldsymbol{\gamma} - \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)^{-1}\left(\mathbf{I}_n - \boldsymbol{\Psi}^h\right)\left(\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}\right) + \sum_{i=0}^{h-1}\boldsymbol{\Gamma}^i\boldsymbol{\epsilon}_{T+h-i},$$

so any disequilibrium at the forecast origin has an increasing impact over time on the level of the series, albeit possibly ‘hidden’ in practice by the increased noise from the cumulative error term. Given this background, we now introduce location shifts into the DGP.

3 Location shifts

The shift of interest here is $\nabla\boldsymbol{\mu}^* = \boldsymbol{\mu}^* - \boldsymbol{\mu}$, where $\boldsymbol{\mu}^*$ denotes the post-break equilibrium mean. Although $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\Omega}_\epsilon$ could alter also, equivalent magnitude changes to these do not entail the same degree of forecast failure—see Hendry (2000). Importantly, being the unconditional growth rate, the sizes of changes to $\boldsymbol{\gamma}$ are generally limited for real variables (e.g., 2.5% pa growth yields $\boldsymbol{\gamma} \simeq 0.006$ in quarterly data, so even a change as large as 0.006 would double real growth). However, $\boldsymbol{\mu}$ need not have any 'natural units' (e.g., as in money demand), and even in cases where it does (consumption-income equations where 0.05–0.2 would be a feasible range), changes could be very large relative to the error standard deviation. In any case, shifts in $\boldsymbol{\gamma}$ are easily incorporated in the following analysis if they are of interest (e.g., as they would be for changes in China's growth rate over the last half century).

Following a change to $\boldsymbol{\mu}^*$ at the forecast origin at time T:

$$\Delta\mathbf{x}_{T+1} = \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}^*\right) + \boldsymbol{\epsilon}_{T+1} \tag{12}$$

so adding and subtracting $\boldsymbol{\alpha\mu}$ in (12):

$$\Delta\mathbf{x}_{T+1} = \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}\right) + \boldsymbol{\epsilon}_{T+1} - \boldsymbol{\alpha}\nabla\boldsymbol{\mu}^* \tag{13}$$

or:

$$\Delta\mathbf{x}_{T+1} = \widehat{\Delta\mathbf{x}}_{T+1|T} - \boldsymbol{\alpha}\nabla\boldsymbol{\mu}^*. \tag{14}$$

The first right-hand side term in (14) (namely $\widehat{\Delta\mathbf{x}}_{T+1|T}$) is the constant-parameter forecast of $\Delta\mathbf{x}_{T+1}$ given by (7); the second is the shift with:

$$\mathsf{E}\left[\Delta\mathbf{x}_{T+1} - \widehat{\Delta\mathbf{x}}_{T+1|T}\right] = -\boldsymbol{\alpha}\nabla\boldsymbol{\mu}^*. \tag{15}$$

Since $\mathsf{E}[\boldsymbol{\beta}'\mathbf{x}_T] = \boldsymbol{\mu}$, then $-\boldsymbol{\alpha}\nabla\boldsymbol{\mu}^*$ is indeed the unanticipated increase in $\Delta\mathbf{x}_{T+1}$ relative to the constant-parameter setting.

For h-steps ahead:

$$\mathsf{E}\left[\Delta\mathbf{x}_{T+h} - \widehat{\Delta\mathbf{x}}_{T+h|T}\right] = -\boldsymbol{\alpha}\boldsymbol{\Psi}^{h-1}\nabla\boldsymbol{\mu}^* \tag{16}$$

which tends to zero as h increases. Thus, following an equilibrium shift in an $\mathsf{I}(1)$ system, further ahead growth rates are forecast more accurately then 1-step. This occurs because adjustment following the change in the level of $\mathbf{x}_t$ induced by the shift in $\boldsymbol{\mu}$ acts like a change in growth which dies out as the new equilibrium mean is attained. Of course, such an outcome is very different from that obtaining in a time-invariant process. As before, the increased variance of multi-period forecasts will entail reduced precision.

Importantly, recommencing the h-steps ahead forecast sequence at $T+j$ using an unchanged model does not alter these results: (15) and (16) continue to hold with (e.g.) $\mathsf{E}[\Delta\mathbf{x}_{T+h+j} - \widehat{\Delta\mathbf{x}}_{T+h+j|T+j}] = -\boldsymbol{\alpha}\boldsymbol{\Psi}^{h-1}\nabla\boldsymbol{\mu}^*$.

However, for levels forecasts after the break:

$$\mathbf{x}_{T+h} = h\boldsymbol{\gamma} - \boldsymbol{\alpha}\sum_{i=0}^{h-1}\boldsymbol{\Psi}^i\boldsymbol{\mu}^* + \sum_{i=0}^{h-1}\boldsymbol{\Gamma}^i\boldsymbol{\epsilon}_{T+h-i} + \boldsymbol{\Gamma}^h\mathbf{x}_T,$$

yielding a forecast error of:

$$\mathsf{E}\left[\mathbf{x}_{T+h} - \widehat{\mathbf{x}}_{T+h|T}\right] = -\boldsymbol{\alpha}\sum_{i=0}^{h-1}\boldsymbol{\Psi}^i\nabla\boldsymbol{\mu}^* = \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)^{-1}\left(\mathbf{I}_n - \boldsymbol{\Psi}^h\right)\nabla\boldsymbol{\mu}^* \tag{17}$$

which increases over the forecast horizon. As with (16), (17) persists for a forecast origin of $T+j$. In both cases, forecast error variance formulae are unchanged from the constant-parameter setting.

A scalar numerical illustration based on the empirical example of UK money demand in section 6 helps highlight some possible magnitudes. Using inverse velocity adjusted for the foregone interest cost of holding money, we have approximately, $\alpha = -0.1$, and $\beta = 1$ with $\nabla\mu^* = 0.5$ and $\sigma_\epsilon = 0.015$ (1.5%) so (15) is initially $0.05 > 3\sigma_\epsilon$ and tends to zero, whereas (17) also starts at 0.05 but increases to 0.5, which is interpretable as 50% of the money stock...

Section 4 now examines possible solutions which avoid such massive forecast failures. Two closely related approaches are considered to improving forecasting robustness in the face of location shifts:

- forecasting from a double-differenced device (denoted DDV) which adjusts quickly to breaks;
- differencing the VEqCM term in (5) to eliminate the equilibrium mean.

We take these two transformations in turn. It should be noted that VEqCMs and DDVs perform equally badly in terms of forecast biases when a break occurs after forecasts are announced (see Clements and Hendry, 1999), so they do not differ in that regard for such a setting, although the latter will have a larger error variance, offset in part by smaller parameter estimation uncertainty. The key difference lies in their performance when forecasting after a break, in which case the VEqCM continues to perform just as badly, as seen above , but the DDV becomes relatively immune to the earlier break. As we will show below, differencing the VEqCM achieves a similar objective for shifts in $\boldsymbol{\mu}$. Updating the parameter estimates is considered in Hendry (2003) as an additional adaptation to change, but in the present context would simply drive the estimated $\boldsymbol{\alpha}$ to zero, and hence end as a model in differences.

4 Forecasting by $\Delta \mathbf{x}_T$

Most economic time series do not continuously accelerate, entailing a zero unconditional expectation of the second difference:

$$\mathsf{E}\left[\Delta^2 \mathbf{x}_t\right] = \mathbf{0}, \tag{18}$$

and suggesting the forecasting rule:

$$\widetilde{\Delta \mathbf{x}}_{T+1|T} = \Delta \mathbf{x}_T. \tag{19}$$

This will deliver unconditionally unbiased, but noisy, forecasts when the DGP has the form (5), even if that DGP is augmented by additional lagged differences. One key to the success of double differencing is that no deterministic terms remain. Indeed, second differencing not only removes two unit roots, any intercepts and linear trends, it also changes location shifts to 'blips', and converts breaks in trends to impulses. Figure 1 illustrates. Thus, while (19) will suffer forecast failure for large changes in $\boldsymbol{\mu}$ in the period of change, it adjusts quickly to breaks, and need not fail even one period later.

For example, from (12) for $\widetilde{\Delta \mathbf{x}}_{T+2|T+1} = \Delta \mathbf{x}_{T+1}$:

$$\Delta \mathbf{x}_{T+2} - \widetilde{\Delta \mathbf{x}}_{T+2|T+1} = \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T+1} - \boldsymbol{\mu}^*\right) + \boldsymbol{\epsilon}_{T+2} - \Delta \mathbf{x}_{T+1} = \boldsymbol{\alpha}\boldsymbol{\beta}'\Delta \mathbf{x}_{T+1} + \Delta\boldsymbol{\epsilon}_{T+2},$$

so for $\Delta \mathbf{x}_{T+2} - \widetilde{\Delta \mathbf{x}}_{T+2|T+1} = \widetilde{\mathbf{u}}_{T+2}$:

$$\mathsf{E}\left[\widetilde{\mathbf{u}}_{T+2}\right] = \mathsf{E}\left[\boldsymbol{\alpha}\boldsymbol{\beta}'\Delta \mathbf{x}_{T+1} + \Delta\boldsymbol{\epsilon}_{T+2}\right] = \mathsf{E}\left[\boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}^*\right)\right] = -\boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)\nabla\boldsymbol{\mu}^*. \tag{20}$$

Compared to (14), which will remain the 1-step error of the VEqCM from a forecast origin of $T+1$, (20) must be smaller. This pattern persists for 1-step errors h-periods after the shift:

$$\mathsf{E}\left[\Delta \mathbf{x}_{T+h} - \widetilde{\Delta \mathbf{x}}_{T+h|T+h-1}\right] = -\boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)\boldsymbol{\Psi}^{h-2}\nabla\boldsymbol{\mu}^*,$$

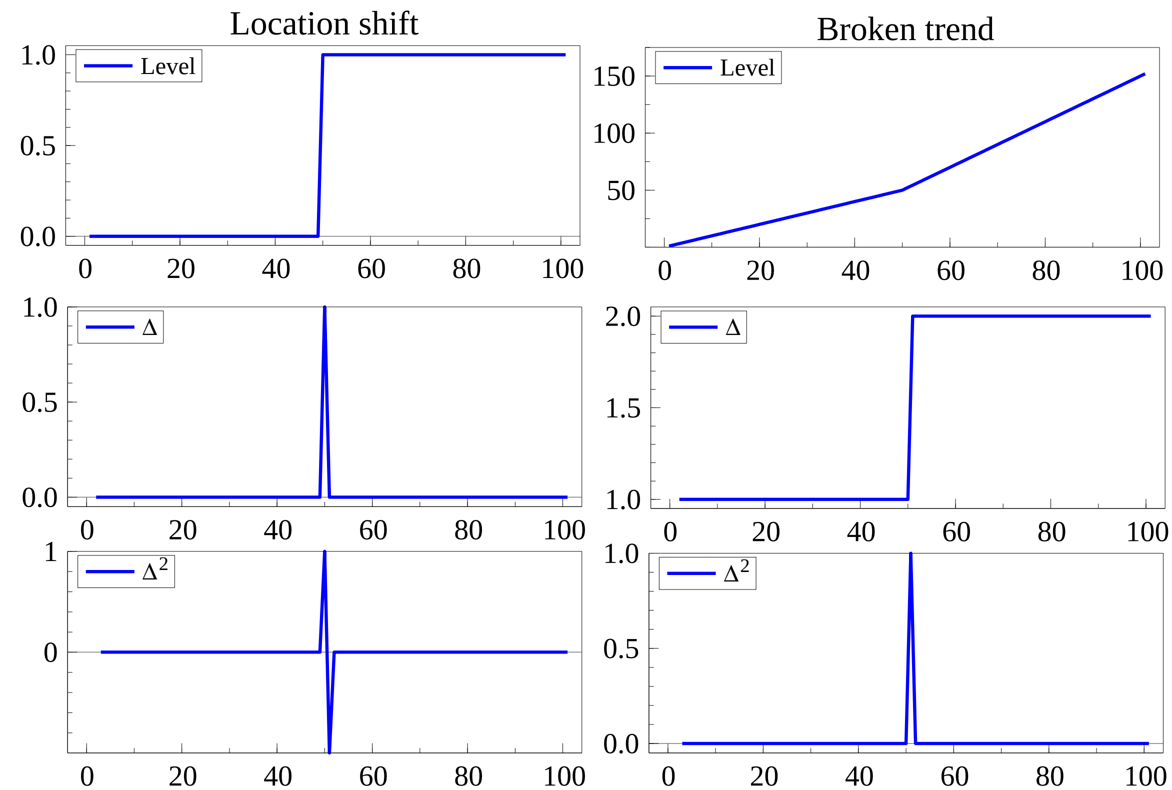

Figure 1 Location shifts and broken trends.

whereas $\mathsf{E}[\Delta \mathbf{x}_{T+h} - \widehat{\Delta \mathbf{x}}_{T+h|T+h-1}] = -\boldsymbol{\alpha} \nabla \boldsymbol{\mu}^*$. For the numerical example above, (20) delivers a bias of -0.005, so has already become negligible.

In addition to the properties just noted, there is a deeper reason why a forecast of the form (19) may generally perform well. Consider an extended in-sample DGP:

$$\Delta \mathbf{x}_t = \boldsymbol{\gamma}_0 + \boldsymbol{\alpha}_0 \left(\boldsymbol{\beta}_0' \mathbf{x}_{t-1} - \boldsymbol{\mu}_0 \right) + \boldsymbol{\Upsilon}_0 \mathbf{z}_t + \boldsymbol{\epsilon}_t, \tag{21}$$

where $\boldsymbol{\epsilon}_t \sim \mathsf{IN}_n \left[\mathbf{0}, \boldsymbol{\Sigma}_\epsilon \right]$ independently of all the included variables and their history, with population parameter values denoted by the subscript 0. In (21), $\{\mathbf{z}_t\}$ denotes potentially many omitted effects, possibly all lagged, but which are $\mathsf{I}(0)$ for consistency with $\mathbf{x}_t$ being $\mathsf{I}(1)$, perhaps because of 'internal' cointegration, differencing, or intrinsic stationarity. We assume $\mathbf{z}_t$ is the mean-zero VAR:

$$\mathbf{z}_t = \boldsymbol{\Phi} \mathbf{z}_{t-1} + \boldsymbol{\eta}_t \;\text{ where }\; \boldsymbol{\eta}_t \sim \mathsf{IN}_k \left[\mathbf{0}, \boldsymbol{\Omega}_\eta \right] \tag{22}$$

and, although it is unrealistic, take $\mathbf{z}_t$ to be orthogonal to $\boldsymbol{\beta}_0' \mathbf{x}_{t-1}$, so need $\boldsymbol{\beta}_0' \boldsymbol{\Upsilon}_0 = \mathbf{0}$. Then the parameter estimates in the original VEqCM are consistent: non-orthogonality would exacerbate the mis-specification problem, so this is probably the most favourable case for the VEqCM, and allows us to work with known parameters to focus on forecast failure comparisons close to those of the previous section. Now the VEqCM (7) is mis-specified by omitting $\boldsymbol{\Upsilon}_0 \mathbf{z}_t$ as well as confronting a location shift. Both effects favour $\widetilde{\Delta \mathbf{x}}_{T+h|T+h-1}$ as we now show.

4.1 Constant-parameter case

We first consider the constant-parameter DGP (21), where we contrast the forecasts from a VEqCM with a DDV for $\mathbf{x}_t$, so both models are mis-specified, but in different ways. Then the 1-step forecast error from the VEqCM is $\boldsymbol{\Upsilon}_0 \mathbf{z}_{T+1} + \boldsymbol{\epsilon}_{T+1}$ where from (22):

$$\mathsf{E} \left[\boldsymbol{\Upsilon}_0 \mathbf{z}_{T+1} + \boldsymbol{\epsilon}_{T+1} \right] = \mathbf{0} \tag{23}$$

and:

$$\mathsf{V}\left[\boldsymbol{\Upsilon}_0 \mathbf{z}_{T+1} + \boldsymbol{\epsilon}_{T+1}\right] = \boldsymbol{\Upsilon}_0 \mathsf{V}\left[\mathbf{z}_t\right] \boldsymbol{\Upsilon}_0' + \boldsymbol{\Omega}_\epsilon \tag{24}$$

where

$$\mathsf{V}\left[\mathbf{z}_t\right] = \boldsymbol{\Phi}\mathbf{V}_z\boldsymbol{\Phi}' + \boldsymbol{\Omega}_\eta. \tag{25}$$

The DDV 1-step forecast error is $\Delta\mathbf{x}_{T+1} - \Delta\mathbf{x}_T = \mathbf{u}_{T+1}$ so is the difference of the right-hand side of (21) at $T+1$:

$$\Delta\mathbf{x}_{T+1} - \Delta\mathbf{x}_T = \boldsymbol{\alpha}_0\boldsymbol{\beta}_0'\Delta\mathbf{x}_T + \boldsymbol{\Upsilon}_0\Delta\mathbf{z}_{T+1} + \Delta\boldsymbol{\epsilon}_{T+1}, \tag{26}$$

which has mean zero and variance:

$$\mathsf{V}\left[\mathbf{u}_{T+1}\right] = \boldsymbol{\alpha}_0\boldsymbol{\beta}_0'\mathsf{V}\left[\Delta\mathbf{x}_T\right]\boldsymbol{\beta}_0\boldsymbol{\alpha}_0' + \boldsymbol{\Upsilon}_0\mathsf{V}\left[\Delta\mathbf{z}_{T+1}\right]\boldsymbol{\Upsilon}_0' + 2\boldsymbol{\Omega}_\epsilon \tag{27}$$

as the covariance $\mathsf{C}\left[\Delta\mathbf{x}_T\Delta\mathbf{z}_{T+1}'\right]$ vanishes when $\boldsymbol{\beta}_0'\boldsymbol{\Upsilon}_0 = \mathbf{0}$, where:

$$\mathsf{V}\left[\Delta\mathbf{z}_t\right] = \left(\boldsymbol{\Phi} - \mathbf{I}_k\right)\mathsf{V}\left[\mathbf{z}_t\right]\left(\boldsymbol{\Phi} - \mathbf{I}_k\right)' + \boldsymbol{\Omega}_\eta. \tag{28}$$

Using (25) and (28), the difference between (24) and (27) is:

$$\boldsymbol{\Upsilon}_0\left(\boldsymbol{\Phi}\mathsf{V}\left[\mathbf{z}_t\right] + \mathsf{V}\left[\mathbf{z}_t\right]\boldsymbol{\Phi}' - \mathsf{V}\left[\mathbf{z}_t\right]\right)\boldsymbol{\Upsilon}_0' - \boldsymbol{\alpha}_0\boldsymbol{\beta}_0'\mathsf{V}\left[\Delta\mathbf{x}_T\right]\boldsymbol{\beta}_0\boldsymbol{\alpha}_0' - \boldsymbol{\Omega}_\epsilon. \tag{29}$$

When $\boldsymbol{\Phi} = \mathbf{0}$ (or, of course, $\boldsymbol{\Upsilon}_0 = \mathbf{0}$), then the VEqCM forecast-error variance dominates that of the DDV, since (29) is negative semi-definite. However, if $\boldsymbol{\Phi} \simeq \mathbf{I}_k$, and the omitted variables are important in explaining $\mathbf{x}_t$, then the difference is:

$$\boldsymbol{\Upsilon}_0\mathsf{V}\left[\mathbf{z}_t\right]\boldsymbol{\Upsilon}_0' - \boldsymbol{\alpha}_0\boldsymbol{\beta}_0'\mathsf{V}\left[\Delta\mathbf{x}_T\right]\boldsymbol{\beta}_0\boldsymbol{\alpha}_0' - \boldsymbol{\Omega}_\epsilon,$$

which could be positive semi-definite, albeit that serious mis-specification is required. Nevertheless, the usual argument that differencing doubles the error variance applies only to the innovation component of the error, and is attenuated by omitted variables.

4.1.1 Scalar illustration 1

When $n = k = 1$, explicitly comparable formulae are readily obtained for the scalar DGP:

$$\Delta x_{T+1} = \gamma_0 + \alpha_0\left(x_T - \mu\right) + \nu_0 z_{T+1} + \epsilon_{T+1}.$$

Then (24) becomes:

$$\sigma_\eta^2\frac{\nu_0^2}{1-\phi^2} + \sigma_\epsilon^2 \tag{30}$$

since $\sigma_z^2 = \sigma_\eta^2/\left(1-\phi^2\right)$; and (27) becomes:

$$2\alpha_0^2\frac{\left(\sigma_\eta^2\nu_0^2 + \sigma_\epsilon^2\right)}{2+\alpha_0} + 2\sigma_\eta^2\nu_0^2\frac{1}{1+\phi} + 2\sigma_\epsilon^2 \tag{31}$$

so the difference between (24) and (27) is:

$$\sigma_\eta^2\nu_0^2\left(\frac{2\phi-1}{1-\phi^2} - \frac{2\alpha_0^2}{2+\alpha_0}\right) - \sigma_\epsilon^2\left(1 + \frac{2\alpha_0^2}{2+\alpha_0}\right),$$

which will be positive only if $\phi > 0.5$, but can certainly be positive (e.g., $\alpha_0 = -0.1$, $\nu_0 = 1$, $\sigma_\eta^2 = \sigma_\epsilon^2$, $\phi > 0.75$ would suffice). Thus, even in a constant parameter world, the 'naive' predictor $\widetilde{\Delta x}_{T+1|T}$ could outperform a (mis-specified) VEqCM.

4.2 Changed-parameter case

However, the more relevant case for our analysis is when the DGP changes over the forecast horizon, and for generality we let all parameters shift to:

$$\Delta \mathbf{x}_{T+i} = \boldsymbol{\gamma}_0^* + \boldsymbol{\alpha}_0^* \left((\boldsymbol{\beta}_0^*)' \mathbf{x}_{T+i-1} - \boldsymbol{\mu}_0^* \right) + \boldsymbol{\Upsilon}_0^* \mathbf{z}_{T+i} + \boldsymbol{\epsilon}_{T+i}. \tag{32}$$

If $\Delta \mathbf{x}_{T+i} - \Delta \widehat{\mathbf{x}}_{T+i|T+i-1} = \mathbf{w}_{T+i}$ when the postulated econometric model is the estimated VEqCM in $\mathbf{x}_t$:

$$\Delta \widehat{\mathbf{x}}_{T+i|T+i-1} = \widehat{\boldsymbol{\gamma}} + \widehat{\boldsymbol{\alpha}} \left(\widehat{\boldsymbol{\beta}}' \mathbf{x}_{T+i-1} - \widehat{\boldsymbol{\mu}} \right) \tag{33}$$

then:

$$\mathbf{w}_{T+i} = \boldsymbol{\gamma}_0^* + \boldsymbol{\alpha}_0^* \left((\boldsymbol{\beta}_0^*)' \mathbf{x}_{T+i-1} - \boldsymbol{\mu}_0^* \right) + \boldsymbol{\Upsilon}_0^* \mathbf{z}_{T+i} + \boldsymbol{\epsilon}_{T+i} - \widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\alpha}} \left(\widehat{\boldsymbol{\beta}}' \mathbf{x}_{T+i-1} - \widehat{\boldsymbol{\mu}} \right). \tag{34}$$

All the main sources of forecast error occur, given (32): stochastic and deterministic breaks, omitted variables, inconsistent parameter estimates, estimation uncertainty, and innovation errors: data measurement errors could be added. Replacing in-sample estimates by the corresponding in-sample population parameter (pseudo-true) values will reduce the forecast-error variances but not otherwise affect the analysis, so is again imposed, leading to (using $\mathsf{E}[\widehat{\boldsymbol{\gamma}}] = \boldsymbol{\gamma}_p$ etc., for in-sample average values):

$$\mathbf{w}_{T+i} = \boldsymbol{\gamma}_0^* + \boldsymbol{\alpha}_0^* \left((\boldsymbol{\beta}_0^*)' \mathbf{x}_{T+i-1} - \boldsymbol{\mu}_0^* \right) + \boldsymbol{\Upsilon}_0^* \mathbf{z}_{T+i} + \boldsymbol{\epsilon}_{T+i} - \boldsymbol{\gamma}_p - \boldsymbol{\alpha}_p \left(\boldsymbol{\beta}_p' \mathbf{x}_{T+i-1} - \boldsymbol{\mu}_p \right). \tag{35}$$

Notice that (35) constitutes a sequence of 1-step ahead forecast errors as the forecast origin increases after the break. Even so, it is difficult to analyze (35) unconditionally as its terms are not necessarily I(0). However, conditional on $(\mathbf{x}_{T+i-1}, \mathbf{z}_{T+i-1})$, $\mathbf{w}_{T+i}$ has an approximate mean forecast error relative to the relevant post-break distribution at $T+i$ of:

$$\begin{aligned} \mathsf{E}_{T+i}\left[\mathbf{w}_{T+i} \mid \mathbf{x}_{T+i-1}, \mathbf{z}_{T+i-1}\right] \quad = \quad & \left(\boldsymbol{\gamma}_0^* - \boldsymbol{\gamma}_p\right) - \left(\boldsymbol{\alpha}_0^* \boldsymbol{\mu}_0^* - \boldsymbol{\alpha}_p \boldsymbol{\mu}_p\right) + \left[\boldsymbol{\alpha}_0^* (\boldsymbol{\beta}_0^*)' - \boldsymbol{\alpha}_p \boldsymbol{\beta}_p'\right] \mathbf{x}_{T+i-1} \\ & + \boldsymbol{\Upsilon}_0^* \mathsf{E}_{T+i}\left[\mathbf{z}_{T+i} \mid \mathbf{x}_{T+i-1}, \mathbf{z}_{T+i-1}\right]. \end{aligned} \tag{36}$$

In general, ignoring chance cancellations, this will be considerably worse than either (15) or (23). Also, neglecting parameter estimation variance uncertainty as $\mathsf{O}_p(T^{-1})$, $\mathbf{w}_{T+i}$ has an approximate conditional forecast-error variance matrix:

$$\mathsf{V}_{T+i}\left[\mathbf{w}_{T+i} \mid \mathbf{x}_{T+i-1}, \mathbf{z}_{T+i-1}\right] = \boldsymbol{\Upsilon}_0^* \mathsf{V}_{T+i}\left[\mathbf{z}_{T+i} \mid \mathbf{x}_{T+i-1}, \mathbf{z}_{T+i-1}\right] \boldsymbol{\Upsilon}_0^{*\prime} + \boldsymbol{\Omega}_\epsilon, \tag{37}$$

and its conditional mean-square forecast error (MSFE) matrix is the sum of (37) and the outer product of (36).

Contrast using the sequence of $\Delta \mathbf{x}_{T+i-1}$ to forecast $\Delta \mathbf{x}_{T+i}$, as in an extension of (19):

$$\widetilde{\Delta \mathbf{x}}_{T+i|T+i-1} = \Delta \mathbf{x}_{T+i-1}. \tag{38}$$

Because of (32), $\Delta \mathbf{x}_{T+i-1}$ is in fact (for $i > 1$):

$$\Delta \mathbf{x}_{T+i-1} = \boldsymbol{\gamma}_0^* + \boldsymbol{\alpha}_0^* \left((\boldsymbol{\beta}_0^*)' \mathbf{x}_{T+i-2} - \boldsymbol{\mu}_0^* \right) + \boldsymbol{\Upsilon}_0^* \mathbf{z}_{T+i-1} + \boldsymbol{\epsilon}_{T+i-1}. \tag{39}$$

Thus, (39) shows that, without the economist needing to know the causal variables or the structure of the economy, $\Delta \mathbf{x}_{T+i-1}$ actually reflects all the desired effects in the DGP, including all the unknown influences and all their changes, with no omitted variables, and no estimation required at all.

Let $\Delta \mathbf{x}_{T+i} - \widetilde{\Delta \mathbf{x}}_{T+i|T+i-1} = \mathbf{u}_{T+i}$, then commencing the analysis at least two periods after the break occurred, so using (39) for $\Delta \mathbf{x}_{T+i-1}$:

$$\begin{aligned} \mathbf{u}_{T+i} &= \boldsymbol{\gamma}_0^* + \boldsymbol{\alpha}_0^* \left((\boldsymbol{\beta}_0^*)' \mathbf{x}_{T+i-1} - \boldsymbol{\mu}_0^* \right) + \boldsymbol{\Upsilon}_0^* \mathbf{z}_{T+i-1} + \boldsymbol{\epsilon}_{T+i} \\ &\quad - \left[\boldsymbol{\gamma}_0^* + \boldsymbol{\alpha}_0^* \left((\boldsymbol{\beta}_0^*)' \mathbf{x}_{T+i-2} - \boldsymbol{\mu}_0^* \right) + \boldsymbol{\Upsilon}_0^* \mathbf{z}_{T+i-1} + \boldsymbol{\epsilon}_{T+i-1} \right] \\ &= \boldsymbol{\alpha}_0^* (\boldsymbol{\beta}_0^*)' \Delta \mathbf{x}_{T+i-1} + \boldsymbol{\Upsilon}_0^* \Delta \mathbf{z}_{T+i} + \Delta \boldsymbol{\epsilon}_{T+i}. \end{aligned} \tag{40}$$

Thus, the outcome is the same as (26), but for the post-break parameters. All terms in the last line must be $\mathsf{I}(-1)$, so will be very 'noisy', but systematic failure should not result.

There are two drawbacks to using (38) which partially offset its advantages: the unwanted presence of $\boldsymbol{\epsilon}_{T+i-1}$ in (39), which doubles the innovation error variance; and all variables in the DGP enter lagged one extra period, which adds the 'noise' of many $\mathsf{I}(-1)$ effects. There is a clear trade-off between using a carefully modelled VEqCM like (33) which might nevertheless be both mis-specified and subject to breaks, and the 'naive' predictor (38). In forecasting competitions across many states of nature with structural breaks and complicated DGPs, it is easy to see why $\Delta \mathbf{x}_{T+i-1}$ could win. Indeed, sufficiently far after the break:

$$\mathsf{E}\left[\mathbf{u}_{T+i}\right] = \boldsymbol{\alpha}_0^* \mathsf{E}\left[(\boldsymbol{\beta}_0^*)' \Delta \mathbf{x}_{T+i-1} \right] + \boldsymbol{\Upsilon}_0^* \mathsf{E}\left[\Delta \mathbf{z}_{T+i}\right] + \mathsf{E}\left[\Delta \boldsymbol{\epsilon}_{T+i}\right] = \boldsymbol{\alpha}_0^* (\boldsymbol{\beta}_0^*)' \boldsymbol{\gamma}_0^{*\prime} = \mathbf{0}.$$

Consequently, (38) will not suffer forecast failure well after breaks, and will fail to win all the time only because of variance effects. Neglecting covariances, we have for variances:

$$\begin{aligned} \mathsf{V}\left[\mathbf{u}_{T+i}\right] &= \mathsf{V}\left[\boldsymbol{\alpha}_0^* (\boldsymbol{\beta}_0^*)' \Delta \mathbf{x}_{T+i-1} \right] + \mathsf{V}\left[\boldsymbol{\Upsilon}_0^* \Delta \mathbf{z}_{T+i} \right] + \mathsf{V}\left[\Delta \boldsymbol{\epsilon}_{T+i} \right] \\ &= \boldsymbol{\alpha}_0^* (\boldsymbol{\beta}_0^*)' \mathsf{V}\left[\Delta \mathbf{x}_{T+i-1}\right] \boldsymbol{\beta}_0^* \boldsymbol{\alpha}_0^{*\prime} + \boldsymbol{\Upsilon}_0^* \mathsf{V}\left[\Delta \mathbf{z}_{T+i}\right] \boldsymbol{\Upsilon}_0^{*\prime} + 2\boldsymbol{\Omega}_\epsilon \end{aligned} \tag{41}$$

which is the MSFE matrix when $\mathsf{E}\left[\mathbf{u}_{T+i}\right] = \mathbf{0}$. Conventional analysis argues for the doubling of $\boldsymbol{\Omega}_\epsilon$ in (41) relative to (37). However, as before, only the innovation error variance component is doubled, so the variance component could even be smaller as in section 4.1, clearly guaranteeing that the combined MSFE would be smaller than from the VEqCM.

4.2.1 Scalar illustration 2

Reverting to a change in μ only for illustrative purposes, with all other parameters constant, and no omitted variables,

$$w_{T+i} = -\alpha_0 \left(\mu_0^* - \mu_0 \right) + \epsilon_{T+i} \tag{42}$$

for which we can calculate the unconditional outcome, namely:

$$\mathsf{E}\left[w_{T+i}\right] = -\alpha_0 \left(\mu_0^* - \mu_0 \right) \;\text{ and }\; \mathsf{V}\left[w_{T+i}\right] = \sigma_\epsilon^2 \tag{43}$$

so the 1-step sequence of MSFEs is approximately:

$$\mathsf{M}\left[w_{T+i}\right] = \alpha_0^2 \left(\mu_0^* - \mu_0 \right)^2 + \sigma_\epsilon^2. \tag{44}$$

In comparison (41) is:

$$\mathsf{M}\left[u_{T+i}\right] = 2\sigma_\epsilon^2 \left(1 + \frac{\alpha_0^2}{2 + \alpha_0} \right). \tag{45}$$

Using the same values $\alpha_0 = -0.1$ with $\nabla \mu_0^* = 0.5$ and $\sigma_\epsilon = 0.015$ related to the empirical example below, then (44) is approximately 6-fold larger than (45). Additional parameter shifts, estimation uncertainty, or specification mistakes would compound that effect.

4.3 Longer-period differences

Instead of (38), one might consider the past annual change to forecast quarterly, say:

$$\Delta\breve{\mathbf{x}}_{T+i|T+i-1} = \frac{1}{4}\sum_{j=1}^{4}\Delta\mathbf{x}_{T+i-j} = \frac{1}{4}\Delta_4\mathbf{x}_{T+i-1}. \tag{46}$$

While *ad hoc*, $\Delta_4\mathbf{x}_{T+i-1}/4$ is an adaptive estimator of $\boldsymbol{\gamma}$ which is slower to reflect breaks than $\Delta\mathbf{x}_{T+i-1}$ but much smoother, so its empirical behaviour is noted below.

5 Forecasting from a transformed VEqCM

We first consider replacing only the equilibrium-correction term in the VEqCM by its first difference, retaining all the other parameters unaltered, namely:

$$\Delta\mathbf{x}_t = \boldsymbol{\gamma} + \boldsymbol{\alpha}\Delta\left(\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\mu}\right) + \boldsymbol{\xi}_t = \boldsymbol{\gamma} + \boldsymbol{\alpha}\boldsymbol{\beta}'\Delta\mathbf{x}_{t-1} + \boldsymbol{\xi}_t. \tag{47}$$

In this simple setting, the effect in (47) is to produce an autoregression in $\Delta\mathbf{x}_t$, albeit not what would be found on estimation: if there is already a lagged $\Delta\mathbf{x}_t$ in the VEqCM, with coefficient $\mathbf{\Pi}_1$ say, then $\mathbf{\Pi}_1$ must be added to $\boldsymbol{\alpha}\boldsymbol{\beta}'$. Since shifts in $\boldsymbol{\mu}$ are the most pernicious for forecasting, (47) might be more robust to such breaks than the original VEqCM (5). On the other hand, there will be a loss of information during periods where no breaks occur.

To examine the behaviour of (47) forecasting $\Delta\mathbf{x}_{T+2}$ from $T+1$ after a break in $\boldsymbol{\mu}$ at time T, let:[1]

$$\overline{\Delta\mathbf{x}}_{T+2|T+1} = \boldsymbol{\gamma} + \boldsymbol{\alpha}\boldsymbol{\beta}'\Delta\mathbf{x}_{T+1} \tag{48}$$

so the forecast error is:

$$\Delta\mathbf{x}_{T+2} - \overline{\Delta\mathbf{x}}_{T+2|T+1} = \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T+1} - \boldsymbol{\mu}^*\right) + \boldsymbol{\epsilon}_{T+2} - \boldsymbol{\gamma} - \boldsymbol{\alpha}\boldsymbol{\beta}'\Delta\mathbf{x}_{T+1}. \tag{49}$$

Since:

$$\mathsf{E}\left[\Delta\mathbf{x}_{T+2}\right] = \boldsymbol{\gamma} - \boldsymbol{\alpha}\boldsymbol{\Psi}\nabla\boldsymbol{\mu}^* \text{ and } \mathsf{E}\left[\overline{\Delta\mathbf{x}}_{T+2|T+1}\right] = \boldsymbol{\gamma} - \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)\nabla\boldsymbol{\mu}^*,$$

then:

$$\mathsf{E}\left[\Delta\mathbf{x}_{T+2} - \overline{\Delta\mathbf{x}}_{T+2|T+1}\right] = -\boldsymbol{\alpha}\boldsymbol{\Psi}\nabla\boldsymbol{\mu}^* + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)\nabla\boldsymbol{\mu}^* = -\boldsymbol{\alpha}\nabla\boldsymbol{\mu}^*,$$

which is the same as the mean forecast error from the original VEqCM, delivering no benefit. Intuitively, the source of the forecast error can be seen in (49), which depends on $\boldsymbol{\mu}^*$ only through the EqCM term, yet $\mathsf{E}\left[\boldsymbol{\beta}'\mathbf{x}_{T+1}\right] = \boldsymbol{\mu}^* - \boldsymbol{\Psi}\nabla\boldsymbol{\mu}^*$ does not fully reflect $\boldsymbol{\mu}^*$.

However, later-period forecasts will benefit. For forecasting $\Delta\mathbf{x}_{T+3}$ from an origin at $T+2$, say:

$$\mathsf{E}\left[\Delta\mathbf{x}_{T+3} - \overline{\Delta\mathbf{x}}_{T+3|T+2}\right] = -\boldsymbol{\alpha}\boldsymbol{\Psi}\nabla\boldsymbol{\mu}^*,$$

so the mean forecast error will gradually decline. Although (48) will induce a smaller increase in the error variance than (38), namely $\boldsymbol{\Omega}_\epsilon + \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{\Omega}_\epsilon\boldsymbol{\beta}\boldsymbol{\alpha}'$ rather than $2\boldsymbol{\Omega}_\epsilon$, merely eliminating the equilibrium mean by differencing does not seem advantageous. Moreover, (48) remains vulnerable to shifts in $\boldsymbol{\gamma}$.

[1] Forecasting one period after the break serves to confirm the absence of a gain from this approach.

5.1 Differencing the VEqCM

Since shifts in γ are the next most pernicious for forecasting, we consider forecasting not from (5) itself, but from a variant thereof which has been differenced *after a congruent representation has been estimated*, namely:

$$\Delta \mathbf{x}_t = \Delta \mathbf{x}_{t-1} + \boldsymbol{\alpha}\boldsymbol{\beta}' \Delta \mathbf{x}_{t-1} + \Delta \boldsymbol{\epsilon}_t = \left(\mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}'\right) \Delta \mathbf{x}_{t-1} + \boldsymbol{\zeta}_t \tag{50}$$

or:

$$\Delta^2 \mathbf{x}_t = \boldsymbol{\alpha}\boldsymbol{\beta}' \Delta \mathbf{x}_{t-1} + \boldsymbol{\zeta}_t. \tag{51}$$

(50) is just the first difference of the original VAR, since $\left(\mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}'\right) = \boldsymbol{\Gamma}$, but with the rank restriction from cointegration imposed. Alternatively, $\Delta \mathbf{x}_{t-1}$ could be interpreted as a highly adaptive estimator of $\boldsymbol{\gamma}$ in (38). The second representation in (51) can be interpreted as augmenting the DDV forecast by $\boldsymbol{\alpha}\boldsymbol{\beta}'\Delta \mathbf{x}_{t-1}$, ‘adding back’ to the DDV the main observable component omitted by using just the lagged first difference as in (38). Thus, a DDV is not only the difference of a DVAR, but is also obtained by dropping the mean-zero term $\boldsymbol{\alpha}\boldsymbol{\beta}'\Delta \mathbf{x}_{t-1}$ from the simplest differenced VEqCM.

To trace the behaviour of (50) after a break in $\boldsymbol{\mu}$, let:

$$\widetilde{\Delta \mathbf{x}}_{T+1|T} = \left(\mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}'\right) \Delta \mathbf{x}_T \tag{52}$$

where from (13):

$$\Delta \mathbf{x}_{T+1} = \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}\right) + \boldsymbol{\epsilon}_{T+1} - \boldsymbol{\alpha}\nabla\boldsymbol{\mu}^*.$$

At time T, $\Delta \boldsymbol{\mu}^* = \nabla \boldsymbol{\mu}^*$, so:

$$\mathsf{E}\left[\Delta \mathbf{x}_{T+1}\right] = \boldsymbol{\gamma} - \boldsymbol{\alpha}\nabla\boldsymbol{\mu}^*,$$

and hence:

$$\mathsf{E}\left[\Delta \mathbf{x}_{T+1} - \widetilde{\Delta \mathbf{x}}_{T+1|T}\right] = \boldsymbol{\gamma} - \boldsymbol{\alpha}\nabla\boldsymbol{\mu}^* - \boldsymbol{\gamma} = -\boldsymbol{\alpha}\nabla\boldsymbol{\mu}^*.$$

As before, there is no gain when the break is after forecasts are announced.

However, $\Delta \boldsymbol{\mu}^* = \nabla \boldsymbol{\mu}^*$ only at time T, so one period later:

$$\mathsf{E}\left[\Delta \mathbf{x}_{T+2}\right] = \mathsf{E}\left[\boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T+1} - \boldsymbol{\mu}^*\right) + \boldsymbol{\epsilon}_{T+2}\right] = \boldsymbol{\gamma} - \boldsymbol{\alpha}\boldsymbol{\Psi}\nabla\boldsymbol{\mu}^*,$$

as:

$$\mathsf{E}\left[\boldsymbol{\beta}'\mathbf{x}_{T+1}\right] = \boldsymbol{\mu} - \boldsymbol{\beta}'\boldsymbol{\alpha}\nabla\boldsymbol{\mu}^* = \boldsymbol{\mu}^* - \boldsymbol{\Psi}\nabla\boldsymbol{\mu}^*$$

so:

$$\mathsf{E}\left[\Delta \mathbf{x}_{T+2} - \widetilde{\Delta \mathbf{x}}_{T+2|T+1}\right] = \boldsymbol{\gamma} - \boldsymbol{\alpha}\boldsymbol{\Psi}\nabla\boldsymbol{\mu}^* - \left(\boldsymbol{\gamma} - \boldsymbol{\alpha}\nabla\boldsymbol{\mu}^*\right) + \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{\alpha}\nabla\boldsymbol{\mu}^* = \mathbf{0}.$$

Thus, the differenced VEqCM ‘misses’ only for 1 period, then does not make systematic, and increasing, errors. Notice that while $\mathsf{E}\left[\boldsymbol{\beta}'\Delta \mathbf{x}_t\right] = \mathbf{0}$ when the process is in equilibrium, 1-step after a break, $\mathsf{E}\left[\boldsymbol{\beta}'\Delta \mathbf{x}_{T+1}\right] = -\boldsymbol{\beta}'\boldsymbol{\alpha}\nabla\boldsymbol{\mu}^*$ so contains important information about the recent forecast-error bias. When breaks occur in $\boldsymbol{\mu}$, (51) should outperform, especially if $\boldsymbol{\gamma}$ also alters.

If all parameters are constant, (52) remains unbiased but inefficient. The next sub-section considers the impact of unnecessary differencing on forecast-error variances, in the context of 1-step ahead forecasts.

5.1.1 Forecast-error variances

Let $\mathbf{e}_{T+h} = \Delta\mathbf{x}_{T+h} - \widetilde{\Delta\mathbf{x}}_{T+h|T+h-1}$ be the sequence of 1-step forecast errors from updating (52), then, ignoring parameter estimation uncertainty as $\mathsf{O}_p\left(T^{-1/2}\right)$:

$$\mathbf{e}_{T+1} = -\boldsymbol{\alpha}\nabla\boldsymbol{\mu}^* + \Delta\boldsymbol{\epsilon}_{T+1},$$

whereas:

$$\mathbf{e}_{T+2} = \Delta\boldsymbol{\epsilon}_{T+2}.$$

Since the system error is $\{\boldsymbol{\epsilon}_t\}$, then in the absence of other mis-specifications, the additional differencing doubles the 1-step error variance. Relative to a DDV, however, there is a gain from the DVEqCM, since the former has the component from the variance of the omitted variable $\boldsymbol{\alpha}\boldsymbol{\beta}'\Delta\mathbf{x}_{T+1}$ (namely $\boldsymbol{\alpha}\boldsymbol{\beta}'\mathsf{V}\left[\Delta\mathbf{x}_{T+1}\right]\boldsymbol{\beta}\boldsymbol{\alpha}'$ in (41)), as well as the same innovation errors. Thus, both central tendency and variability should be better for the DVEqCM than a DDV in the absence of parameter estimation uncertainty.

6 Empirical illustration: UK M1

The two 'forecasting' models of UK M1 in Hendry and Mizon (1993) and Hendry and Doornik (1994) respectively illustrate several of the above phenomena (related studies include Hendry, 1979; Hendry and Ericsson, 1991; Boswijk, 1992; Johansen, 1992; Paruolo, 1996; and Rahbek, Kongsted and Jørgensen, 1999). The data are quarterly, seasonally-adjusted, time series over 1963(1)–1989(2), defined as:

M	nominal M1,
I	real total final expenditure (*TFE*) at 1985 prices,
P	the *TFE* deflator,
R_{la}	the three-month local authority interest rate,
R_o	learning-adjusted own interest rate,
R_{net}	$R_{la} - R_o$.

The first model was based on using the competitive interest rate R_{la}, and the second on the opportunity-cost measure R_{net} appropriate after the Banking Act of 1984 legalized interest payments on chequing accounts. To simplify the results, we first consider only the money-demand equation, then turn briefly to system behaviour. In both cases, 'forecasts' are over the five years 1984(3)–1989(2), or subsets thereof, from an origin shortly after the Act.[2]

Figure 2 (panel a) shows the time series for $v = p + i - m$ (log velocity, using lower case for logs) and R_{la}, with a marked divergence apparent at the end of the sample. Panel b graphs the computed EqCMs for 'excess money' from the two earlier studies, defined respectively by:

$$\begin{aligned} \widehat{\boldsymbol{\beta}}'\mathbf{x}_t &= m - p - i + 7.3R_{la} + 0R_o + 5.6\Delta p \\ \widetilde{\boldsymbol{\beta}}'\mathbf{x}_t &= \widehat{\boldsymbol{\beta}}'\mathbf{x}_t - 7.3R_o \end{aligned}$$

These coincided till 1984(2), after which the former behaves as in earlier cycles, whereas the latter appears to plumb new depths: by the end of the sample, they have diverged by more than 50% of the money stock. That the correct EqCM is discrepant, may, at first sight, seem counter-intuitive, but it

[2] M1 data ceased to be collected after 1989 when Building Societies (in M4, but not M1) started converting to banks, which led to large jumps in the value of M1 on conversion days.

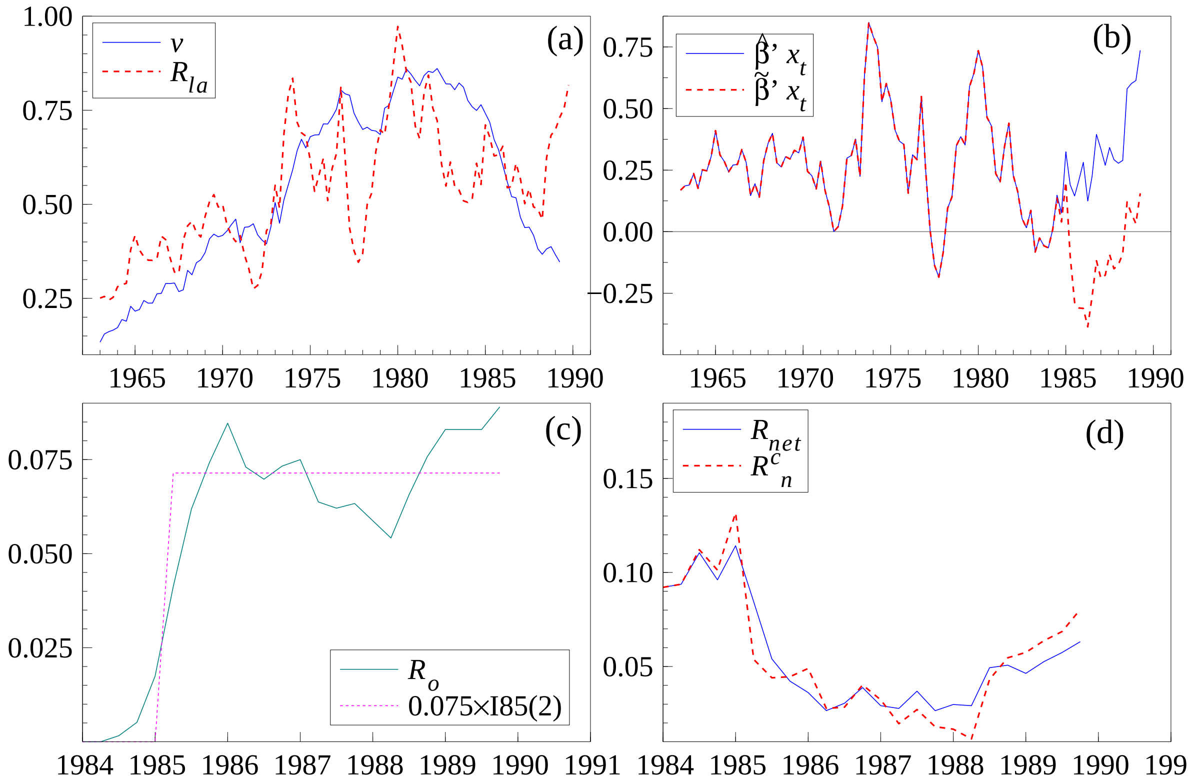

Figure 2 Effects of the 1984 Banking Act on UK M1 .

occurs precisely because the opportunity cost has shifted dramatically, yet $\widehat{\boldsymbol{\beta}}'\mathbf{x}_t$ does not reflect that shift: not doing so causes the forecast failure shown in figure 3 below. Figure 2c illustrates that the Banking Act corresponded to an equilibrium-mean shift relative to the model based on R_{la}.[3] The own rate, R_o has a mean of approximately 0.072 over the forecast horizon, and a shift indicator $1_{\{t>1985(2)\}}$ times that mean closely approximates the actual time path of R_o, so $R^c_n = R_{la} - 0.072 \times 1_{\{t>1985(2)\}}$ in figure 2d is close to R_{net}. Consequently:

$$\widetilde{\boldsymbol{\beta}}'\mathbf{x}_t \simeq \widehat{\boldsymbol{\beta}}'\mathbf{x}_t - 0.525 \times 1_{\{t>1985(2)\}},$$

yielding $\nabla\mu^* = 0.525$ as noted above. On this basis, the legislative change acts like a massive step shift in μ, so the earlier theory should be relevant to explaining this episode of forecast failure. Indeed, if real money and R_{net} co-break, as illustrated in Clements and Hendry (1999, Ch. 9), then $\widetilde{\boldsymbol{\beta}}'\mathbf{x}_t$ should also be an appropriate EqCM post the legislative change.

6.1 Single-equation results

Figure 3a shows the dismal performance on 20 1-step 'forecasts' of the Hendry and Mizon (1993) model for the growth rate of real money, $\Delta(m-p)$, based on $\widehat{\boldsymbol{\beta}}'\mathbf{x}_t$: this model uses current-dated values of R_{net} and Δp, yet almost none of the $\pm 2\widehat{\sigma}_f$ error bars includes the associated outcome. In fact, a large fall in money demand is forecast during what was the largest sustained rise ever experienced historically. The mean forecast error is 4.4% with a root mean squared forecast error (RMSFE) of 4.9%.

For comparison, the 20 1-step forecasts from the first differences of that original model are shown in figure 3b: there is a very substantial improvement, with no systematic under-forecasting, suggesting that the adaptation proposed in section 5.1 can be effective in the face of equilibrium-mean shifts. All

[3] The figure also shows why an intercept correction might perform well after 1985(4).

the panels are on the same scale, so the increase in the conventionally-calculated interval forecasts due to the differencing is also clear (although these error bars no longer correctly represent the uncertainty). The corresponding mean forecast error is 0.4% with an RMSFE of 1.8%: these are clearly a dramatic improvement, especially noting that the in-sample $\widehat{\sigma}$ is 1.3%. Figure 5a below shows the two sets of forecast errors (all panels on the same scale).

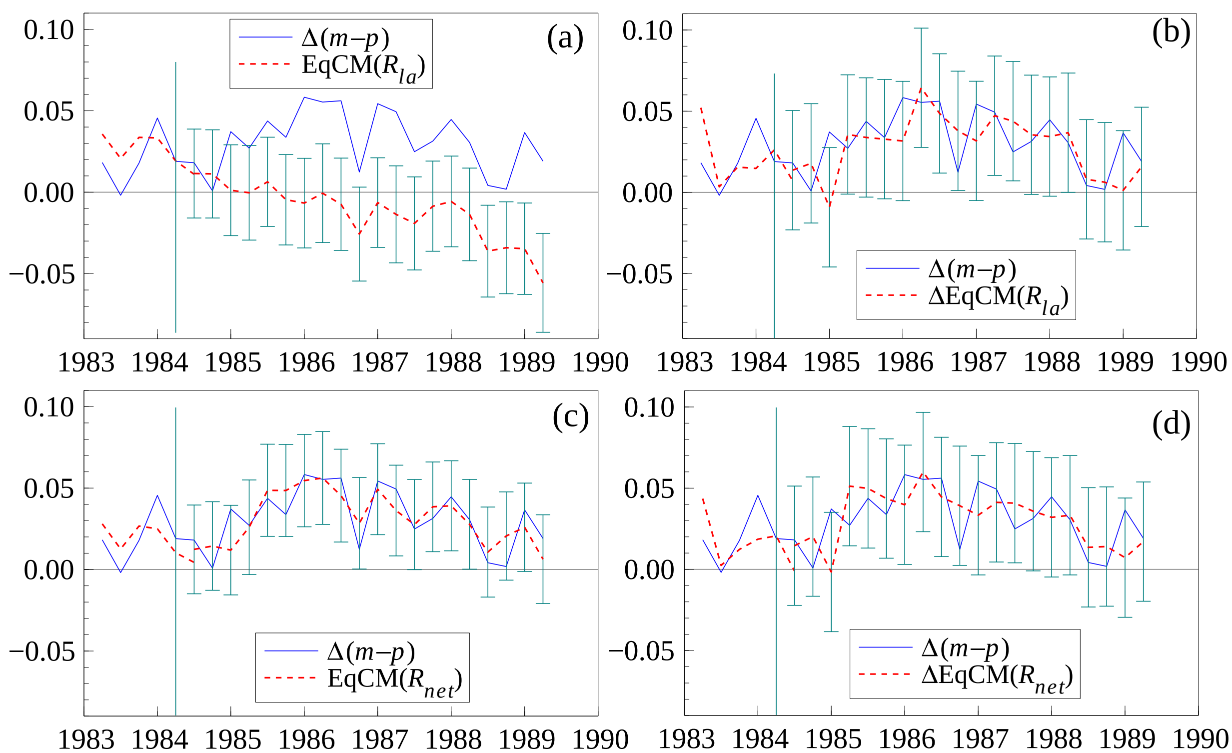

Figure 3 1-step forecasts of UK M1 from conditional models.

Figure 3c shows the good performance on 20 1-step forecasts of the 'correct' model (i.e., that based on R_{net}), which is identical in-sample to the failed model. The mean forecast error is negligible at 0.06% with an RMSFE of 1.14%. Thus, these forecasts are better than the fit.

Since one cannot know in advance whether or not a given model is 'correct' and hence robust to an apparent break, the effects of differencing applied to the R_{net} based-model are also worth investigating. These produce similar forecasts to the EqCM, as shown in figure 3d, but again with larger (conventional) error bars. Now the mean forecast error is 0.05% (the smallest of the four) with an RMSFE of 1.79%, which is essentially the same as from differencing the incorrect model: in fact, their forecast errors are correlated 0.94. Thus, the costs of the differencing strategy do not seem to be too high for the 'correct specification', but the benefits are substantial when differencing is needed.

For comparison, forecasts based on the other adaptive device, the DDV from section 4, are shown in figure 4 panel a. The DDV actually has a smaller mean error than the 'correct' model (less than 0.001%), but a much larger RMSFE of 2.25%, so there are definite benefits from correct causal information.[4] Moreover, the benefits from using either differenced EqCM are marked, consistent with the earlier theory that including $\boldsymbol{\alpha\beta}'\Delta\mathbf{x}_{t-1}$ would improve performance. Finally, that the RMSFE has doubled relative to the EqCM based on R_{net} suggests that omitted effects, other parameter changes, and estimation uncertainty must be minimal. Figure 5b shows the comparative forecast errors, and reveals how much smaller they are than those in panel a.

[4]Subject to the *caveats* that the 'correct' model uses current-dated variables in its 'forecasts'.

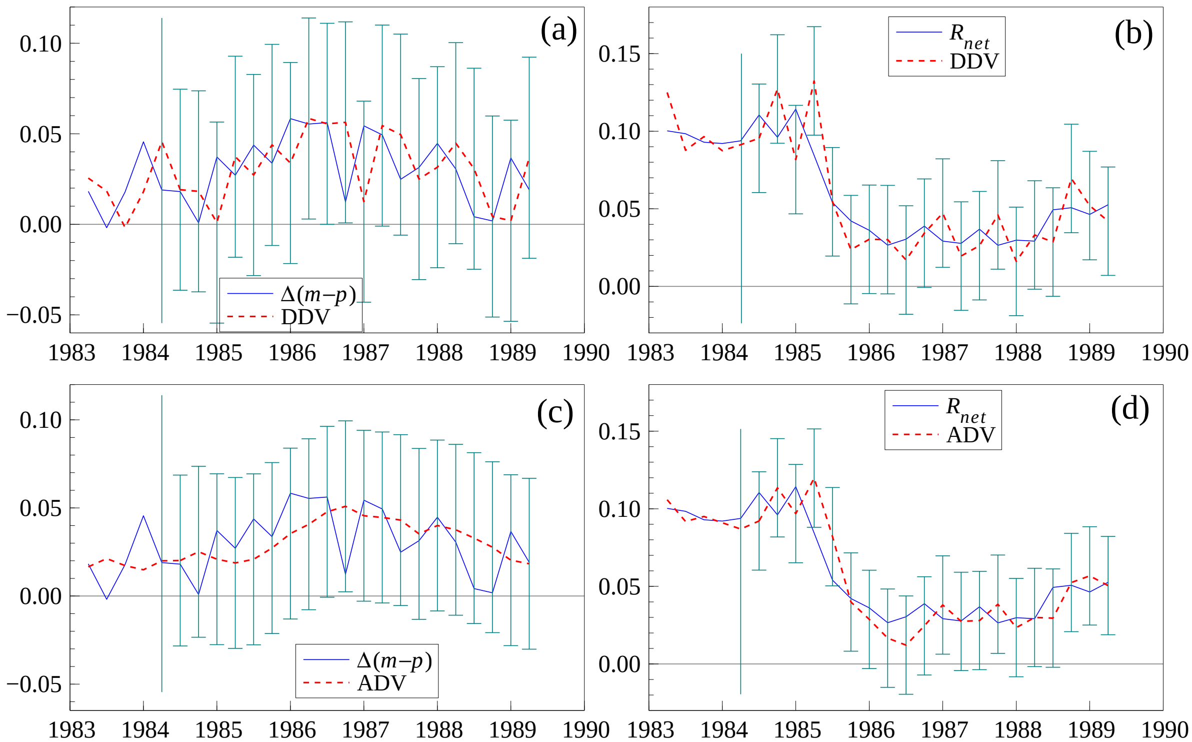

Figure 4 DDV and ADV 1-step forecasts of UK M1.

The ADV forecasts shown in figure 4c are distinctly better than the DDV, having a mean forecast error of -0.07% and an RMSFE of 1.8%. Hence some degree of smoothing seems to pay. This is also true of the ADV and DDV forecasts for R_{net} shown in figure 4 panels b and d (ADV RMSFE of 1.5% as against 1.9%). Thus, while double differencing is highly adaptive when a break occurs, the additional error variance at all points seems to more than offset its advantage in comparison to the smoother adaptation used here. Figure 5c shows that the resulting forecast errors are more volatile than those in panel b, but less biased than the EqCM(R_{la})-based forecasts.

6.2 System behaviour

In a system context, there are three major changes to most of the methods, although the DDV and ADV devices are unaltered. First, the contemporaneous variables in the money-demand model must be forecast, even for 1-step ahead. There is a smaller loss from doing so here than might be anticipated, with a mean forecast error of 0.7% and an RMSFE of 1.59%. Figure 5d records the VEqCM forecast errors for $\Delta\left(m-p\right)$ from the R_{net} system for comparison with the conditional single-equation forecast errors. It also shows the DVEqCM forecast errors to highlight the small loss from the additional differencing of the correct specification. The forecasts from the VEqCM based on R_{la} are as poor as the single equation ones for $\Delta\left(m-p\right)$, but differencing that VEqCM again corrects the main forecast error bias, delivering errors similar to those of the DEqCM.

Secondly, multi-step forecasts can be calculated. These serve to confirm the above results, and while more realistic of the operational setting confronting forecasters, add little to our understanding of the properties of the alternative devices under consideration here. Since the two VEqCMs are identical in-sample, so are their multi-step forecasts for any horizon h. Conversely, the DDV class has a rapidly increasing variance as the horizon grows due to its additional unit root.

Thirdly, the break which occurred in the money-demand equation in the VEqCM based on R_{la}

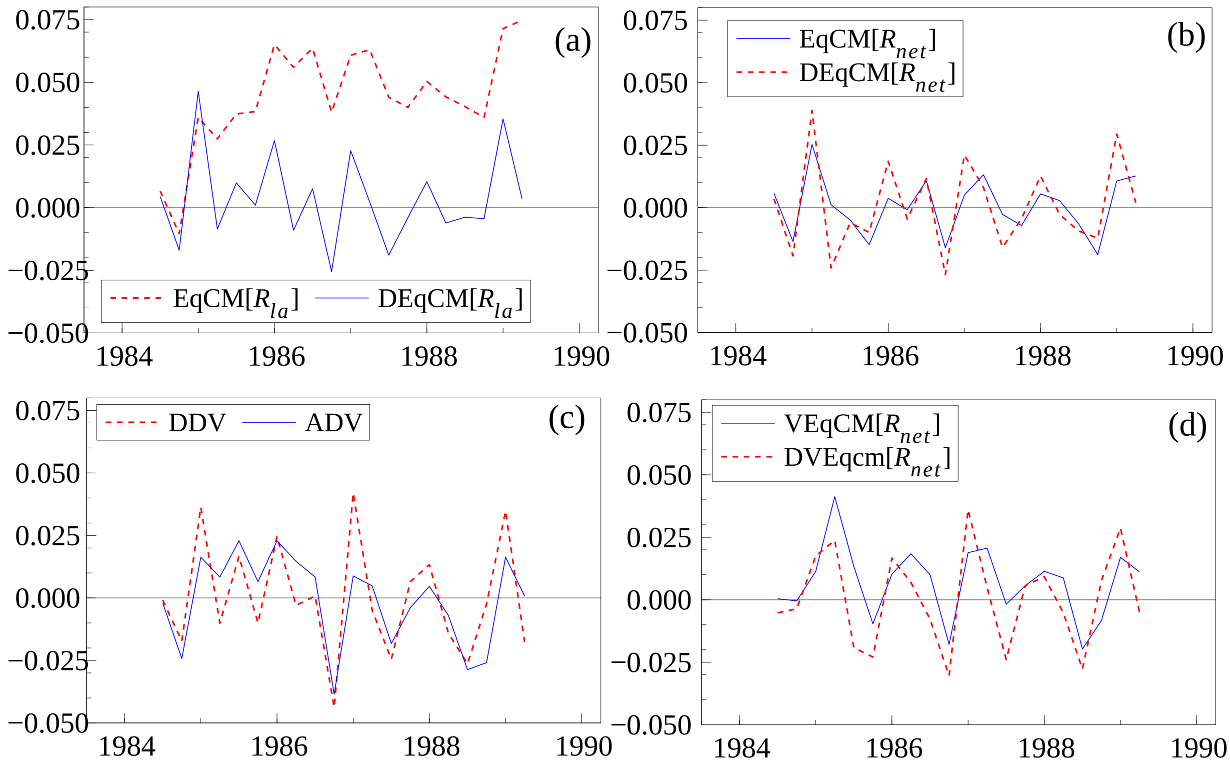

Figure 5 1-step forecast errors for different models of UK M1.

becomes a shift in the R_{net} equation in the second VEqCM—which in turn could not be forecast accurately. The problem for forecasters is that the most difficult variable to predict can unduly worsen the overall outcome. This is an aspect that multi-step forecasts of the levels highlight best, as can be seen in figure 6, for $(m-p)$ and R_{net} (the outcomes for *i* and Δp are omitted). Figure 6 is based on $h=4$, so the first four forecasts match for the corresponding variables, after which the correct VEqCM does noticeably better for $(m-p)$ but is unable to forecast R_{net} very well.

Other aspects of adaptive forecasting could be incorporated with any of the above devices, including intercept corrections, recursive updating of parameter estimates, and reselecting the relevant variables (see e.g., Phillips, 1994). The first of these would clearly be beneficial, given the systematic departures visible in figure 6. When implemented following a large location shift, the second often leads to estimates closer to a DDV than a VEqCM, as the additional differencing eliminates some of the adverse effects of the shift. The third accelerates the tendency just noted.

7 Conclusions

Using a cointegrated linear dynamic system with breaks over the forecast horizon as the illustrative DGP, two adaptations were considered. The first was using second differences to forecast; the second was forecasting from a differenced VEqCM. A new explanation for the relative success of the former was proposed, and the second related to that as also retaining one of the key observable components, namely the change in the equilibrium correction.

The empirical example of the behaviour of M1 in the UK following the Banking Act of 1984 illustrated these two adaptations in action, for mis-specified and 'correct' variants, respectively dependent on the pre and post Act opportunity-cost measures. All four approaches behaved as anticipated from the theory, and demonstrated the difficulty of out-performing 'naive extrapolative devices' when these

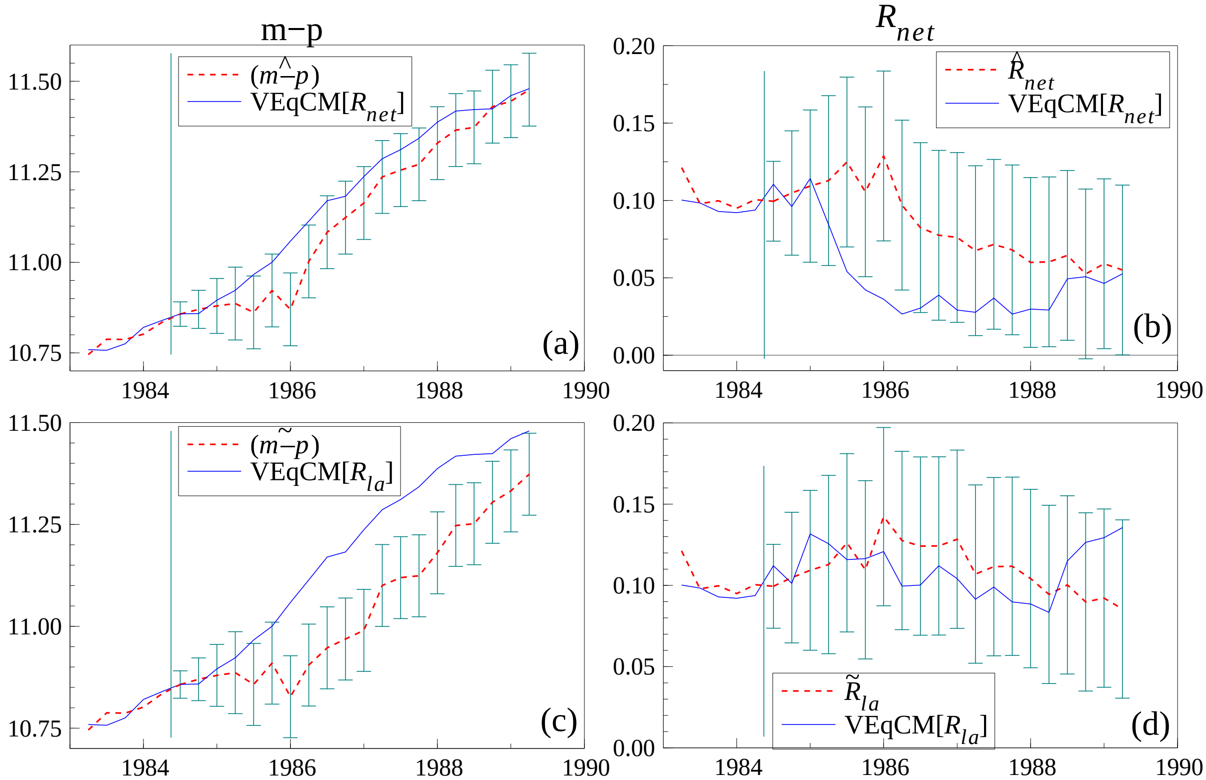

Figure 6 System 4-step forecasts from two VEqCMs of UK M1.

are adaptive to precisely those location shifts which are inherently inimical to econometric systems. Overall, the outcomes suggest that, to retain causal information when the forecast-horizon 'goodness' of the model in use is unknown, model transformations based on differencing may prove a worthwhile route.

References

Boswijk, H. P. (1992). *Cointegration, Identification and Exogeneity*, Vol. 37 of *Tinbergen Institute Research Series*. Amsterdam: Thesis Publishers.

Clements, M. P., and Hendry, D. F. (1995). Forecasting in cointegrated systems. *Journal of Applied Econometrics*, **10**, 127–146. Reprinted in T.C. Mills (ed.), *Economic Forecasting*. Edward Elgar, 1999.

Clements, M. P., and Hendry, D. F. (1996). Forecasting in macro-economics. In Cox, D. R., Hinkley, D. V., and Barndorff-Nielsen, O. E. (eds.), *Time Series Models: In econometrics, finance and other fields*, pp. 101–141. London: Chapman and Hall.

Clements, M. P., and Hendry, D. F. (1998a). Forecasting economic processes. *International Journal of Forecasting*, **14**, 111–131.

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

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

Clements, M. P., and Hendry, D. F. (2001). Explaining the results of the M3 forecasting competition. *International Journal of Forecasting*, **17**, 550–554.

Engle, R. F., and Granger, C. W. J. (1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica*, **55**, 251–276.

Engle, R. F., and Yoo, B. S. (1987). Forecasting and testing in co-integrated systems. *Journal of Econometrics*, **35**, 143–159.

Fildes, R., and Ord, K. (2002). Forecasting competitions – their role in improving forecasting practice and research. In Clements, M. P., and Hendry, D. F. (eds.), *A Companion to Economic Forecasting*, pp. 322–253. Oxford: Blackwells.

Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, **16**, 121–130.

Granger, C. W. J., and Weiss, A. A. (1983). Time series analysis of error-correction models. In Karlin, S., Amemiya, T., and Goodman, L. A. (eds.), *Studies in Econometrics, Time Series, and Multivariate Statistics*, pp. 255–278. New York: Academic Press.

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., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000.

Hendry, D. F. (2000). On detectable and non-detectable structural change. *Structural Change and Economic Dynamics*, **11**, 45–65. Reprinted in *The Economics of Structural Change*, Hagemann, H. Landesman, M. and Scazzieri (eds.), Edward Elgar, Cheltenham, 2002.

Hendry, D. F. (2003). Unpredictability and the foundations of economic forecasting. Working paper, Economics Department, Oxford University.

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 von Ungern-Sternberg, T. (1981). Liquidity and inflation effects on consumers' expenditure. In Deaton, A. S. (ed.), *Essays in the Theory and Measurement of Consumers' Behaviour*, pp. 237–261. Cambridge: Cambridge University Press. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, **12**, 231–254. Reprinted in R.F. Engle and C.W.J. Granger (eds), *Long-Run Economic Relationships*, Oxford: Oxford University Press, 1991, 131–52.

Johansen, S. (1992). A representation of vector autoregressive processes integrated of order 2. *Econometric Theory*, **8**, 188–202.

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. New York: Springer-Verlag.

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

Paruolo, P. (1996). On the determination of integration indices in I(2) systems. *Journal of Econometrics*, **72**, 313–356.

Phillips, P. C. B. (1994). Bayes models and forecasts of Australian macroeconomic time series. In Hargreaves, C. (ed.), *Non-stationary Time-series Analysis and Cointegration*, pp. 53–86. Oxford: Oxford University Press.

Rahbek, A., Kongsted, H. C., and Jørgensen, C. (1999). Trend-stationarity in the I(2) cointegration model. *Journal of Econometrics*, **90**, 265–289.

Stock, J. H., and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics*, **14**, 11–30.

Stock, J. H., and Watson, M. W. (1999). A comparison of linear and nonlinear models for forecasting macroeconomic time series. In Engle, R. F., and White, H. (eds.), *Cointegration, Causality and Forecasting*, pp. 1–44. Oxford: Oxford University Press.