Robustifying Forecasts from Equilibrium-correction Systems

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

August 8, 2005

Abstract

Cointegration analysis has led to equilibrium-correction econometric systems being ubiquitous. But 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, as forecasts will tend to move in the opposite direction to data. We explain the empirical success of second-differenced devices and of model transformations based on additional differencing as reducing forecast-error biases, at some cost in increased forecast-error variances. The analysis is illustrated by an empirical application to narrow money holdings in the UK.

JEL Classification: E41, C52

Keywords: *Cointegration, Non-stationarity, Equilibrium correction, Robust forecasts, UK money.*

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 being ubiquitous for modelling, forecasting and economic policy analysis. In fact, most econometric models are members of the equilibrium-correction class, which includes not only explicit vector equilibrium-correction models (denoted VEqCMs) 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 ARCH, GARCH etc.). The forecasting properties of this huge class are essentially generic, and seem well represented by those of VEqCMs. Hence we focus on those below.

Initially, both theory and Monte Carlo simulations suggested VEqCMs should outperform when forecasting, especially for cointegrated combinations of variables: see e.g., Engle and Yoo (1987),

*Department of Economics, Manor Road Building, Manor Road, Oxford, OX1 3UQ, UK. Tel: (44) (0)1865 278554. email: david.hendry@economics.ox.ac.uk

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 equilibria, and also showed that a causal economic theory basis for forecasting models is of no avail in a world of location shifts: following such breaks, methods using no variables from the data generation process (DGP) could outperform well-specified models, both theoretically and empirically (see e.g., Allen and Fildes, 2001, 2004, for evidence). 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 tend 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. Thus, the prevalence of structural changes in macroeconomic time series confirmed in Stock and Watson (1996) helped account for such outcomes.

Since VEqCMs should forecast well when a process is difference stationary, but are unreliable if location shifts occur, we consider model transformations which retain their causal basis yet reduce forecast-error biases, at some cost in increased forecast-error variances (other adaptive approaches, and the basis for these, are discussed in Hendry, 2003). We also present a new explanation for why some so-called 'naive' forecasting devices based on double differencing may be hard to outperform, even if they are apparently poor approximations to the in-sample DGP. Whereas a VEqCM will perform badly when forecasting after a location shift has occurred, a double-differenced device (DDD) can deliver near unbiased forecasts in that setting. On the basis of that finding, we can then show that the differenced VEqCM, denoted DVEqCM, should in turn outperform the DDD.

Section 2 provides some background to the present approach. Next, section 3 specifies the cointegrated DGP and its properties as a forecasting device, then section 4 considers the effects thereon of location shifts. Section 5 discusses why DDDs may forecast well in DGPs subject to such structural breaks, for which even the best VEqCM is an incomplete description. This leads in section 6 to a differencing transformation which improves the robustness of VEqCMs when forecasting in such a context, and helps them outperform DDDs. Section 7 illustrates these ideas for the much-studied empirical example of UK M1. Section 8 concludes.

2 Background

The background to this paper lies in a theory of economic forecasting based on the properties of unpredictable processes, considered in Hendry (2003), combined with a detailed taxonomy of forecast errors developed in Clements and Hendry (1998b, 1999), and extended to a non-parametric representation in Clements and Hendry (2004). The former delineates the many steps from an entity being predictable to a forecast thereof; for each such step, the latter shows which of the resulting possible sources of forecast error are likely to induce forecast failure, and which are not.

Let:

$$\mathbf{y}_t = \mathbf{f}_t\left(\mathcal{I}_{t-1}\right) + \boldsymbol{\nu}_t \tag{1}$$

where $\mathbf{y}_t$ is the set of n variables of interest, generated by the DGP $\mathsf{D}_{\mathbf{y}_t}\left(\mathbf{y}_t|\mathcal{I}_{t-1}\right)$ where $\mathcal{I}_{t-1} \subseteq \mathcal{I}_t$ denotes the (cumulative) DGP information set, and $\boldsymbol{\nu}_t$ is an unpredictable non-degenerate vector random variable, defined by the property that over the period $\mathcal{T} = \{1, \ldots, T\}$, the conditional distribution $\mathsf{D}_{\boldsymbol{\nu}_t}\left(\boldsymbol{\nu}_t|\mathcal{I}_{t-1}\right)$ equals the unconditional $\mathsf{D}_{\boldsymbol{\nu}_t}\left(\boldsymbol{\nu}_t\right)$:

$$\mathsf{D}_{\boldsymbol{\nu}_t}\left(\boldsymbol{\nu}_t \mid \mathcal{I}_{t-1}\right) = \mathsf{D}_{\boldsymbol{\nu}_t}\left(\boldsymbol{\nu}_t\right) \quad \forall t \in \mathcal{T}. \tag{2}$$

To forecast $\mathbf{y}_{T+1}$, the in-sample model $\psi(\widehat{\mathcal{J}}_T, \widehat{\boldsymbol{\theta}}_T)$ is developed for some specification of the ℓ parameters $\boldsymbol{\theta} \in \mathbb{R}^\ell$ estimated as $\widehat{\boldsymbol{\theta}}_T$ from the full-sample information $\widehat{\mathcal{J}}_T$ where $\mathcal{J}_{t-1} \subseteq \mathcal{I}_{t-1}$ is the available information set at each point in time, measured by $\widehat{\mathcal{J}}_{t-1}$ such that:

$$\widehat{\mathbf{y}}_{T+1|T} = \boldsymbol{\psi}_1\left(\widehat{\mathcal{J}}_T, \widehat{\boldsymbol{\theta}}_T\right). \tag{3}$$

There are many ways to formulate the function $\boldsymbol{\psi}_1(\cdot)$ in (3) for a dynamic model $\boldsymbol{\psi}(\cdot)$, including ‘powering up’ and multi-step estimation, but only the former is considered below (on the latter, see Bhansali, 1996, 1999, 2002, Clements and Hendry, 1996b, and Chevillon and Hendry, 2005, *inter alia*). For simplicity of exposition, we focus on the first two moments for 1-step ahead forecasts, rather than the complete forecast distribution.

The key factors that determine the forecast error:

$$\widehat{\mathbf{u}}_{T+1|T} = \mathbf{y}_{T+1} - \widehat{\mathbf{y}}_{T+1|T} = \mathbf{f}_{T+1}\left(\mathcal{I}_T\right) + \boldsymbol{\nu}_{T+1} - \boldsymbol{\psi}_1(\widehat{\mathcal{J}}_T, \widehat{\boldsymbol{\theta}}_T),$$

are: the composition of the DGP information sets $\mathcal{I}_{t-1}$; how each $\mathcal{I}_{t-1}$ enters the DGP $\mathsf{D}_{\mathbf{y}_t}\left(\mathbf{y}_t|\mathcal{I}_{t-1}\right)$; how $\mathsf{D}_{\mathbf{y}_t}\left(\mathbf{y}_t|\mathcal{I}_{t-1}\right)$ changes over time in-sample; the limited information set $\mathcal{J}_{t-1} \subseteq \mathcal{I}_{t-1}$; the mapping of $\mathsf{D}_{\mathbf{y}_t}\left(\mathbf{y}_t|\mathcal{I}_{t-1}\right)$ into $\mathsf{D}_{\mathbf{y}_t}\left(\mathbf{y}_t|\mathcal{J}_{t-1}\right)$, called the local DGP (see Bontemps and Mizon, 2003), which induces:

$$\mathbf{g}_t\left(\mathcal{J}_{t-1}\right) = \mathsf{E}_t\left[\mathbf{f}_t\left(\mathcal{I}_{t-1}\right) \mid \mathcal{J}_{t-1}\right],$$

such that $\boldsymbol{\epsilon}_t = \mathbf{y}_t - \mathbf{g}_t\left(\mathcal{J}_{t-1}\right)$ is unpredictable relative to $\mathcal{J}_{t-1}$; how $\mathcal{J}_T$ will enter $\mathsf{D}_{\mathbf{y}_{T+1}}\left(\cdot|\mathcal{J}_T\right)$ for a forecast origin at T; the approximation of $\mathbf{g}_t\left(\mathcal{J}_{t-1}\right)$ by the model $\boldsymbol{\psi}\left(\mathcal{J}_{t-1}, \boldsymbol{\theta}\right)$; the specification of $\boldsymbol{\theta}$; measurement errors in each $\widehat{\mathcal{J}}_{t-1}$ for $\mathcal{J}_{t-1}$ (which may themselves change over time); and the estimation of $\boldsymbol{\theta}$ by $\widehat{\boldsymbol{\theta}}_T$, which together determine the properties of $\boldsymbol{\psi}_1(\cdot)$. The first six are aspects of predictability in the DGP; the second four of the formulation of forecasting models like $\boldsymbol{\psi}_1(\cdot)$ which seek to capture that predictability.

Given such a formulation, $\widehat{\mathbf{u}}_{T+1|T}$ can be decomposed into errors which derive from each of the main reduction or transformation steps, namely:

$$\begin{aligned}\widehat{\mathbf{u}}_{T+1|T} =\ & \boldsymbol{\nu}_{T+1} + \left[\mathbf{f}_{T+1}(\mathcal{I}_T) - \mathbf{g}_{T+1}(\mathcal{J}_T)\right] + \left[\mathbf{g}_{T+1}(\mathcal{J}_T) - \mathbf{g}_{T+1|T}(\mathcal{J}_T)\right] \\ & + \left[\mathbf{g}_{T+1|T}(\mathcal{J}_T) - \boldsymbol{\psi}_1(\mathcal{J}_T, \boldsymbol{\theta})\right] + \left[\boldsymbol{\psi}_1(\mathcal{J}_T, \boldsymbol{\theta}) - \boldsymbol{\psi}_1(\widehat{\mathcal{J}}_T, \boldsymbol{\theta})\right] + \left[\boldsymbol{\psi}_1(\widehat{\mathcal{J}}_T, \boldsymbol{\theta}) - \boldsymbol{\psi}_1(\widehat{\mathcal{J}}_T, \widehat{\boldsymbol{\theta}}_T)\right]\end{aligned} \tag{4}$$

where $\mathbf{g}_{T+1|T}(\mathcal{J}_T)$ is the 'extrapolated' value of $\mathbf{g}_{T+1}(\mathcal{J}_T)$ based on the forecast-origin form of $\mathbf{g}(\cdot)$. While decompositions such as (4) are not unique, they help pinpoint the potential sources of forecast failure, and which components are less likely to have a pernicious effect on forecast accuracy.

Taking the six right-hand side terms in (4) in turn, the first three are unknowable (in the absence of a crystal ball), being dependent on the future innovation $\boldsymbol{\nu}_{T+1}$, the reduction to the limited information set, and post-forecast-origin changes in the induced process: all 3 are, therefore, unpredictable, will affect the forecast-error variance, and may influence its mean. The first and second terms have expected values of zero for proper information sets $\mathcal{I}$ and $\mathcal{J}$, so will not affect $\mathsf{E}_{T+1}\left[\widehat{\mathbf{u}}_{T+1|T}|\mathcal{J}_T\right]$. Consequently, a lack of knowledge of the complete information set $\mathcal{I}$ is not an explanation for forecast failure, a general result of importance below, although using more (relevant) information will reduce the variance component deriving from $\mathbf{f}_{T+1}(\mathcal{I}_T) - \mathbf{g}_{T+1}(\mathcal{J}_T)$. However, the third term is a potential source of forecast failure when $\mathbf{g}_{T+1}(\mathcal{J}_T) \neq \mathbf{g}_{T+1|T}(\mathcal{J}_T)$. That requires an induced location shift to be non-zero on average, rather than just structural change in general. Conversely, this third term would be zero under constant parameters.

The next three terms depend on the goodness of the model for the local DGP $\mathsf{D}_{\mathbf{y}_T}(\mathbf{y}_T|\mathcal{J}_T)$ and on data accuracy, both in-sample and at the forecast origin, as well as the choice of estimator. Specifically, the fourth is a function of the adequacy of the model, the fifth of the data accuracy at T, and the last on the properties of the estimator $\widehat{\boldsymbol{\theta}}_T$ for $\boldsymbol{\theta}$ when the observed data are used. Thus, the fourth term would be zero for a correctly specified model, the fifth zero for accurate data, but the sixth only zero in an infinite sample. The focus in many derivations of forecast-error uncertainties on the impacts of parameter estimation and innovation error variances reflects assumptions of a well-specified model of a constant DGP based on accurate data.

The three main general lessons of value from this summary for the analysis below are:

that forecast failure is primarily due to induced location changes: this will account for the failure of the VEqCM after such shifts;

that limited information does not explain forecast failure: this will account for why a DDD, using almost no information, can dominate the VEqCM;

that more information helps reduce forecast error variances: this will account for the dominance of the DVEqCM over the simpler DDD.

These results will be established in the special case of a cointegrated DGP, which is now considered in detail.

3 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 local DGP is:

$$\mathbf{x}_t = \boldsymbol{\tau} + \boldsymbol{\Gamma}\mathbf{x}_{t-1} + \boldsymbol{\epsilon}_t \text{ where } \boldsymbol{\epsilon}_t \sim \mathsf{IN}_n\left[\mathbf{0}, \boldsymbol{\Omega}_\epsilon\right]. \tag{5}$$

$\boldsymbol{\Gamma}$ is an $n \times n$ matrix of coefficients and $\boldsymbol{\tau}$ is an n dimensional vector of intercepts. The specification in (5) is assumed constant in-sample, and the system is taken to be $\mathsf{I}(1)$, satisfying the $r < n$ cointegration relations:

$$\boldsymbol{\Gamma} = \mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}'. \tag{6}$$

In (6), $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $n \times r$ full-rank matrices, no roots of $|\mathbf{I} - \boldsymbol{\Gamma}L| = 0$ lie inside unit circle (where $L^s\mathbf{x}_t = \mathbf{x}_{t-s}$), and $\boldsymbol{\alpha}'_\perp\boldsymbol{\Gamma}\boldsymbol{\beta}_\perp$ 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 (5) is reparametrized as the VEqCM:

$$\Delta\mathbf{x}_t = \boldsymbol{\tau} + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \boldsymbol{\epsilon}_t. \tag{7}$$

Both $\Delta\mathbf{x}_t$ and $\boldsymbol{\beta}'\mathbf{x}_t$ are $\mathsf{I}(0)$, but may have non-zero means. However, $\mathsf{E}\left[\Delta\boldsymbol{\beta}'\mathbf{x}_t\right] = \mathsf{E}\left[\boldsymbol{\beta}'\Delta\mathbf{x}_t\right] = \mathbf{0}$ being the average growth of an $\mathsf{I}(0)$ variable. Let:

$$\boldsymbol{\tau} = \boldsymbol{\gamma} - \boldsymbol{\alpha}\boldsymbol{\mu} \tag{8}$$

and write (7) as:

$$\left(\Delta\mathbf{x}_t - \boldsymbol{\gamma}\right) = \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\mu}\right) + \boldsymbol{\epsilon}_t, \tag{9}$$

then pre-multiplying by $\boldsymbol{\beta}'$, and taking expectations:

$$\mathsf{E}\left[\boldsymbol{\beta}'\Delta\mathbf{x}_t\right] = \boldsymbol{\beta}'\boldsymbol{\gamma} + \boldsymbol{\beta}'\boldsymbol{\alpha}\mathsf{E}\left[\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\mu}\right] + \boldsymbol{\beta}'\mathsf{E}\left[\boldsymbol{\epsilon}_t\right] = \mathbf{0}.$$

When $\boldsymbol{\beta}'\boldsymbol{\alpha}$ is non-singular:

$$\mathsf{E}\left[\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\mu}\right] = -\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)^{-1}\boldsymbol{\beta}'\boldsymbol{\gamma}, \tag{10}$$

so when $\boldsymbol{\tau}$ lies in the cointegration space, $\boldsymbol{\gamma} = \mathbf{0}$, and $\mathsf{E}\left[\boldsymbol{\beta}'\mathbf{x}_{t-1}\right] = \boldsymbol{\mu}$, which matches the condition in Johansen and Juselius (1990). Taking expectations in (9) using (10):

$$\mathsf{E}\left[\Delta\mathbf{x}_t\right] = \boldsymbol{\gamma} + \boldsymbol{\alpha}\mathsf{E}\left[\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\mu}\right] + \mathsf{E}\left[\boldsymbol{\epsilon}_t\right] = \boldsymbol{\gamma} - \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)^{-1}\boldsymbol{\beta}'\boldsymbol{\gamma} = \left(\mathbf{I}_n - \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)^{-1}\boldsymbol{\beta}'\right)\boldsymbol{\gamma} = \mathbf{K}\boldsymbol{\gamma},$$

where $\mathbf{K}$ is non-symmetric idempotent with $\boldsymbol{\beta}'\mathbf{K} = \mathbf{0}'$ and $\mathbf{K}\boldsymbol{\alpha} = \mathbf{0}$ so $\boldsymbol{\Gamma}\mathbf{K} = \mathbf{K} = \mathbf{K}\boldsymbol{\Gamma}$ which implies that $\mathbf{K}\boldsymbol{\gamma} = \mathbf{K}\boldsymbol{\tau}$. Imposing $\mathsf{E}\left[\boldsymbol{\beta}'\mathbf{x}_{t-1}\right] = \boldsymbol{\mu}$ generally, so the long-run equilibrium mean is the constant:

$$\mathsf{E}\left[\boldsymbol{\beta}'\mathbf{x}_t\right] = \boldsymbol{\mu}, \tag{11}$$

then $\boldsymbol{\beta}'\boldsymbol{\gamma} = \mathbf{0}$ with $\boldsymbol{\Gamma}\boldsymbol{\gamma} = \boldsymbol{\gamma}$, and $\mathbf{K}\boldsymbol{\gamma} = \boldsymbol{\gamma}$ and:

$$\mathsf{E}\left[\Delta\mathbf{x}_t\right] = \boldsymbol{\gamma} + \boldsymbol{\alpha}\mathsf{E}\left[\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\mu}\right] + \mathsf{E}\left[\boldsymbol{\epsilon}_t\right] = \boldsymbol{\gamma}. \tag{12}$$

Thus, in (9), 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 from $\boldsymbol{\beta}'\boldsymbol{\gamma} = \mathbf{0}$, and $\boldsymbol{\mu}$ is $r \times 1$, leaving n unrestricted intercepts in total in (9). Consequently, $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\mu}$ are assumed to be variation free, although in principle, $\boldsymbol{\mu}$ could depend on $\boldsymbol{\gamma}$ as in (10): see Hendry and von Ungern-Sternberg (1981). Then $(\boldsymbol{\tau}, \boldsymbol{\Gamma})$ 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 (1996a).

3.1 Forecasting properties for a constant DGP

When the parameters of (9) 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 (9) coincide with the conditional expectation $\mathsf{E}_T\left[\Delta\mathbf{x}_{T+1}|\mathbf{x}_T\right]$, and are given by:

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

(this is an example of $\mathbf{g}_{T+1|T}(\cdot)$ above). The h-step ahead forecast errors for the growth rate are $\widehat{\boldsymbol{\epsilon}}_{T+h|T} = \Delta\mathbf{x}_{T+h} - \Delta\widehat{\mathbf{x}}_{T+h|T} = \boldsymbol{\epsilon}_{T+1}$.

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

$$\widehat{\mathbf{x}}_{T+1|T} = \mathbf{x}_T + \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}\right) = \boldsymbol{\tau} + \boldsymbol{\Gamma}\mathbf{x}_T, \tag{14}$$

so $\widetilde{\boldsymbol{\epsilon}}_{T+1|T} = \widehat{\boldsymbol{\epsilon}}_{T+1|T}$. The h-step ahead forecast errors from (14) are then generated recursively by:

$$\widehat{\mathbf{x}}_{T+h|T} = \boldsymbol{\tau} + \boldsymbol{\Gamma}\widehat{\mathbf{x}}_{T+h-1|T} = \sum_{i=0}^{h-1}\boldsymbol{\Gamma}^i\boldsymbol{\tau} + \boldsymbol{\Gamma}^h\mathbf{x}_T. \tag{15}$$

As:

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

for known parameters, we have:

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

with:

$$\mathsf{E}\left[\widetilde{\boldsymbol{\epsilon}}_{T+h|T}\right] = \mathbf{0} \text{ and } \mathsf{V}\left[\widetilde{\boldsymbol{\epsilon}}_{T+h|T}\right] = \sum_{i=0}^{h-1} \boldsymbol{\Gamma}^i \boldsymbol{\Omega}_\epsilon \boldsymbol{\Gamma}^{i\prime} \tag{16}$$

where $\mathsf{V}[\cdot]$ denotes the variance, which is $\mathsf{O}(h)$ in (16) because $\boldsymbol{\Gamma}^i$ increases in i, but eventually converges to $\mathbf{K}$ as we now show. We use the well-known results that (see e.g., Clements and Hendry, 1995):

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

and:

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

In a cointegrated system, $\boldsymbol{\Psi}$ corresponds to the eigenvalues of $\boldsymbol{\Gamma}$ which are strictly less than unity in absolute value, so $\boldsymbol{\Psi}^h \to \mathbf{0}$ as $h \to \infty$. Since $\mathbf{I}_r - \boldsymbol{\Psi} = \boldsymbol{\beta}'\boldsymbol{\alpha}$, then:

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

so that $\boldsymbol{\Gamma}^h \to \mathbf{K}$ as $h \to \infty$ with:

$$\mathbf{x}_{T+h} = \mathbf{x}_T + h\boldsymbol{\gamma} - \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)^{-1}\left(\mathbf{I}_r - \boldsymbol{\Psi}^h\right)\left(\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}\right) + \sum_{i=0}^{h-1} \boldsymbol{\Gamma}^i \boldsymbol{\epsilon}_{T+h-i}. \tag{17}$$

Thus, any disequilibrium at the forecast origin has an increasing impact over time on the level of the series as $\boldsymbol{\Psi}^h \to \mathbf{0}$, albeit possibly 'hidden' in practice by the increased noise from the cumulative error term.

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

$$\begin{aligned}
\Delta\mathbf{x}_{T+h} &= \boldsymbol{\Gamma}^{h-1}\boldsymbol{\tau} + \boldsymbol{\Gamma}^{h-1}\left(\boldsymbol{\Gamma} - \mathbf{I}_n\right)\mathbf{x}_T + \boldsymbol{\epsilon}_{T+h} + \left(\boldsymbol{\Gamma} - \mathbf{I}_n\right)\sum_{i=0}^{h-2} \boldsymbol{\Gamma}^i \boldsymbol{\epsilon}_{T+h-i-1} \\
&= \boldsymbol{\gamma} + \boldsymbol{\alpha}\boldsymbol{\Psi}^{h-1}\left(\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}\right) + \boldsymbol{\epsilon}_{T+h} - \boldsymbol{\alpha}\sum_{i=0}^{h-2} \boldsymbol{\Psi}^i \boldsymbol{\beta}' \boldsymbol{\epsilon}_{T+h-1-i}, \qquad (18)
\end{aligned}$$

so:

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

The impact of the initial disequilibrium fades in (18) in contrast to (17), a result of importance in explaining the differing behaviour of levels and growth-rate forecasts when there has been an unmodelled equilibrium shift. Since $\widehat{\boldsymbol{\epsilon}}_{T+h|T} = \Delta \mathbf{x}_{T+h} - \widehat{\Delta \mathbf{x}}_{T+h|T}$:

$$\mathsf{E}\left[\widehat{\boldsymbol{\epsilon}}_{T+h|T}\right] = \mathbf{0} \text{ and } \mathsf{V}\left[\widehat{\boldsymbol{\epsilon}}_{T+h|T}\right] = \boldsymbol{\Omega}_{\epsilon} + \sum_{i=0}^{h-2} \boldsymbol{\alpha}\boldsymbol{\Psi}^{i}\boldsymbol{\beta}'\boldsymbol{\Omega}_{\epsilon}\boldsymbol{\beta}\boldsymbol{\Psi}^{i\prime}\boldsymbol{\alpha}', \tag{19}$$

where $\mathsf{V}\left[\widehat{\boldsymbol{\epsilon}}_{T+h|T}\right]$ is $\mathsf{O}(1)$ in h in (19).

Parameter estimation adds terms of $\mathsf{O}_p(T^{-1})$ to $\mathsf{V}\left[\widehat{\boldsymbol{\epsilon}}_{T+h|T}\right]$ and $\mathsf{V}\left[\widetilde{\boldsymbol{\epsilon}}_{T+h|T}\right]$ for a sample of size T. Given this background, we now introduce location shifts into the DGP.

4 The impact of location shifts

The main 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 as well, reasonable changes to these rarely entail the same magnitude of forecast failure: see Hendry (2000). In a scalar setting for log-linear models, $\boldsymbol{\Omega}_{\epsilon} = \sigma_{\epsilon}^2$ where $\sigma_{\epsilon} = 0.015$ (1.5%) is a representative value for an error standard deviation. Being an unconditional growth rate, the sizes of changes to $\boldsymbol{\gamma}$ are limited for real variables (e.g.), $\boldsymbol{\gamma} = 0.006$ in quarterly data corresponds to 2.5% pa growth, so even a change of 0.006–which would double real growth to 5% pa–is only $0.4\sigma_{\epsilon}$. However, $\boldsymbol{\mu}$ need not have any 'natural units' (e.g., as in money demand), and even in cases where it does (as with consumption-income equations where values in the range 0.05–0.15 would be a feasible), changes could be very large relative to σ_{ϵ}, namely $3\sigma_{\epsilon}$ to $10\sigma_{\epsilon}$. In any case, shifts in $\boldsymbol{\gamma}$ are easily incorporated in the 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:

$$\Delta \mathbf{x}_{T+1} = \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}^*\right) + \boldsymbol{\epsilon}_{T+1} \tag{20}$$

so adding and subtracting $\boldsymbol{\alpha}\boldsymbol{\mu}$ in (20):

$$\Delta \mathbf{x}_{T+1} = \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}\right) + \boldsymbol{\epsilon}_{T+1} - \boldsymbol{\alpha}\nabla\boldsymbol{\mu}^* = \widehat{\Delta \mathbf{x}}_{T+1|T} - \boldsymbol{\alpha}\nabla\boldsymbol{\mu}^* + \boldsymbol{\epsilon}_{T+1}. \tag{21}$$

The term $\widehat{\Delta \mathbf{x}}_{T+1|T}$ in (21) is the constant-parameter forecast of $\Delta \mathbf{x}_{T+1}$ given by (13) (i.e., $\mathbf{g}_{T+1|T}(\cdot)$), whereas:

$$\mathsf{E}\left[\Delta \mathbf{x}_{T+1} - \widehat{\Delta \mathbf{x}}_{T+1|T}\right] = -\boldsymbol{\alpha}\nabla\boldsymbol{\mu}^*. \tag{22}$$

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

For h-steps ahead:

$$\mathsf{E}\left[\Delta \mathbf{x}_{T+h} - \widehat{\Delta \mathbf{x}}_{T+h|T}\right] = -\boldsymbol{\alpha}\boldsymbol{\Psi}^{h-1}\nabla\boldsymbol{\mu}^* \tag{23}$$

which tends to zero as h increases as $\boldsymbol{\Psi}^h \to \mathbf{0}$. Thus, following an equilibrium shift in an I(1) system, further ahead growth rates are forecast more accurately in mean than 1-step. This occurs because adjustment to 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. Such an outcome is very different from that obtaining in a constant-parameter process, where predictability cannot increase for a given future outcome the further ahead it is. However, the increased variance of multi-period forecasts will entail reduced precision.

Recommencing the h-steps ahead forecast sequence at $T+j$ using an unchanged model does not alter these results: (22) and (23) continue to hold with (e.g.) $\mathsf{E}[\Delta \mathbf{x}_{T+h+j} - \widehat{\Delta \mathbf{x}}_{T+h+j|T+j}] = -\boldsymbol{\alpha}\boldsymbol{\Psi}^{h-1}\nabla\boldsymbol{\mu}^*$.

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:

$$\mathsf{E}\left[\mathbf{x}_{T+h} - \widehat{\mathbf{x}}_{T+h|T}\right] = -\boldsymbol{\alpha}\sum_{i=0}^{h-1}\boldsymbol{\Psi}^i\nabla\boldsymbol{\mu}^* = \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)^{-1}\left(\mathbf{I}_r - \boldsymbol{\Psi}^h\right)\nabla\boldsymbol{\mu}^* \tag{24}$$

which increases to $\boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)^{-1}\nabla\boldsymbol{\mu}^*$ over the forecast horizon as $\boldsymbol{\Psi}^h \to \mathbf{0}$. As with (23), (24) persists even from a forecast origin of $T+j$ well after the break. 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 7 helps highlight some possible magnitudes. Considering inverse velocity adjusted for the foregone interest cost of holding money, we have approximately, $\alpha = -0.1$, and $\beta = 1$ with $\nabla\mu^* = 0.5$ (which is interpretable as 50% of the money stock) and $\sigma_\epsilon = 0.015$ (1.5%). Then (22) is initially $0.05 > 3\sigma_\epsilon$ but tends to zero, whereas (24) also starts at 0.05 but increases to 0.5 (since $(\beta\alpha)^{-1} = 10$), a huge forecast failure of $33\sigma_\epsilon$.

Sections 5 and 6 examine two closely related approaches to avoiding forecast failure by improving robustness to location shifts:

- forecasting from a double-differenced device (denoted DDD) which adjusts quickly to breaks;
- differencing the VEqCM in (9) to eliminate the equilibrium mean and growth intercept.

We take these two transformations in turn. It must be stressed that VEqCMs and DDDs 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 at any point after a break has already occurred, in which case the

VEqCM continues to perform just as badly (as shown above), but the DDD becomes relatively immune to the earlier break. As we will show below, differencing the VEqCM achieves a similar objective (for shifts in $\boldsymbol{\mu}$). Such breaks are more common in lower frequency data for a given occurrence rate, so the 'insurance costs' of differencing are probably less worth paying the higher the data frequency. 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.

5 Forecasting by a DDD

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

$$\mathsf{E}\left[\Delta^2\mathbf{x}_t\right] = \mathbf{0}, \tag{25}$$

and suggesting the forecasting rule:

$$\widetilde{\Delta\mathbf{x}}_{T+1|T} = \Delta\mathbf{x}_T. \tag{26}$$

This will deliver unconditionally unbiased, but noisy, forecasts when the DGP has the form (9), 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 (26) 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.

Figure 1 here

For example, from (20) 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 $\Delta\mathbf{x}_{T+2} - \widetilde{\Delta\mathbf{x}}_{T+2|T+1} = \widetilde{\mathbf{u}}_{T+2|T+1}$:

$$\mathsf{E}\left[\widetilde{\mathbf{u}}_{T+2|T+1}\right] = \mathsf{E}\left[\boldsymbol{\alpha}\boldsymbol{\beta}'\Delta\mathbf{x}_{T+1} + \Delta\boldsymbol{\epsilon}_{T+2}\right] = \mathsf{E}\left[\boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}^*\right)\right] = -\boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)\nabla\boldsymbol{\mu}^*. \tag{27}$$

Compared to (21), which will remain the 1-step error of the VEqCM from a forecast origin of $T+1$, (27) must be smaller. This pattern persists for 1-step errors h-periods after the shift:

$$\mathsf{E}\left[\Delta\mathbf{x}_{T+h} - \widetilde{\Delta\mathbf{x}}_{T+h|T+h-1}\right] = -\boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)\boldsymbol{\Psi}^{h-2}\nabla\boldsymbol{\mu}^* \tag{28}$$

whereas $\mathsf{E}[\Delta\mathbf{x}_{T+h} - \widehat{\Delta\mathbf{x}}_{T+h|T+h-1}] = -\boldsymbol{\alpha}\nabla\boldsymbol{\mu}^*$. For the scalar numerical example above, (27) delivers $-0.1 \times 0.1 \times 0.5$ which is a bias of $-0.005 = -0.3\sigma_\epsilon$, so is negligible.

In addition to the properties just noted, there is a deeper reason why a forecast of the form (26) may generally perform well. Consider an extended in-sample DGP which reflects the analysis in section 2 that the process may depend on a larger information set than the model:

$$\Delta \mathbf{x}_t = \boldsymbol{\gamma}_0 + \boldsymbol{\alpha}_0 \left(\boldsymbol{\beta}_0' \mathbf{x}_{t-1} - \boldsymbol{\mu}_0 \right) + \boldsymbol{\Upsilon}_0 \mathbf{z}_t + \boldsymbol{\nu}_t, \tag{29}$$

where $\boldsymbol{\nu}_t \sim \mathsf{IN}_n \left[\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\nu}} \right]$ independently of all the included variables and their history, with population parameter values denoted by the subscript 0. Thus, $\boldsymbol{\nu}_t$ is the DGP innovation such that $\boldsymbol{\epsilon}_t = \boldsymbol{\Upsilon}_0 \mathbf{z}_t + \boldsymbol{\nu}_t$. In (29), $\{\mathbf{z}_t\}$ denotes potentially many omitted effects, possibly all lagged, but which for consistency with only $\mathbf{x}_t$ being $\mathsf{I}(1)$, are $\mathsf{I}(0)$ perhaps because of ‘internal’ cointegration, differencing, or intrinsic stationarity. We assume $\mathbf{z}_t$ is generated by the k-dimensional mean-zero VAR:

$$\mathbf{z}_t = \boldsymbol{\Phi} \mathbf{z}_{t-1} + \boldsymbol{\eta}_t \text{ where } \boldsymbol{\eta}_t \sim \mathsf{IN}_k \left[\mathbf{0}, \boldsymbol{\Omega}_\eta \right], \tag{30}$$

where

$$\mathsf{V}\left[\mathbf{z}_t \right] = \mathbf{V}_z = \boldsymbol{\Phi} \mathbf{V}_z \boldsymbol{\Phi}' + \boldsymbol{\Omega}_\eta. \tag{31}$$

Although it is unrealistic, we take $\mathbf{z}_t$ to be orthogonal to $\boldsymbol{\beta}_0' \mathbf{x}_{t-1}$, with $\boldsymbol{\beta}_0' \boldsymbol{\Upsilon}_0 = \mathbf{0}$ so 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.[1] Now the VEqCM (13) 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.

5.1 Changed parameters

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

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

If $\Delta \mathbf{x}_{T+i} - \Delta \widehat{\mathbf{x}}_{T+i|T+i-1} = \mathbf{w}_{T+i|T+i-1}$ 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:

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

[1] It is feasible, though tedious, to relax this requirement, but the additional inconsistency of the estimated cointegration relations would serve to strengthen the argument.

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 $\mathsf{E}[\widehat{\gamma}] = \boldsymbol{\gamma}_p$ etc., for in-sample average values):

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

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 $(\mathbf{x}_{T+i-1}, \mathbf{z}_{T+i-1})$, $\mathbf{w}_{T+i|T+i-1}$ has an approximate mean forecast error relative to the relevant post-break distribution at $T+i$ of:

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

In general, ignoring chance cancellations, this will be considerably worse than (22)). Also, neglecting parameter estimation variance uncertainty as $\mathsf{O}_p(T^{-1})$, $\mathbf{w}_{T+i|T+i-1}$ has an approximate conditional forecast-error variance matrix:

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

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 (26):

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

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

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

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

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

All terms in the last line must be $\mathsf{I}(-1)$, so will be ‘noisy’, but systematic failure should not result.

There are two drawbacks to using (38) which partially offset its advantages: the unwanted presence of $\boldsymbol{\nu}_{T+i-1}$ 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:

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

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:

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

which is the MSFE matrix when $\mathsf{E}\left[\mathbf{u}_{T+i|T+i-1}\right] = \mathbf{0}$. Conventional analysis argues for the doubling of $\boldsymbol{\Omega}_{\boldsymbol{\nu}}$ in (41) relative to (37). However, only the innovation error variance component is doubled, so the variance component could even be smaller (as in section 5.2), clearly guaranteeing that the combined MSFE would be smaller than from the VEqCM.

5.1.1 Scalar illustration 1

Considering only a change in μ for illustrative purposes, with all other parameters constant, and no omitted variables, the VEqCM error is:

$$w_{T+i|T+i-1} = -\alpha_0\left(\mu_0^* - \mu_0\right) + \nu_{T+i} \tag{42}$$

with unconditional outcome:

$$\mathsf{E}\left[w_{T+i|T+i-1}\right] = -\alpha_0\left(\mu_0^* - \mu_0\right) \text{ and } \mathsf{V}\left[w_{T+i|T+i-1}\right] = \sigma_\nu^2 \tag{43}$$

so the 1-step sequence of MSFEs is approximately:

$$\mathsf{M}\left[w_{T+i|T+i-1}\right] = \alpha_0^2\left(\mu_0^* - \mu_0\right)^2 + \sigma_\nu^2. \tag{44}$$

In comparison that of the DDD in (41) is:

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

Using the earlier values $\alpha_0 = -0.1$ with $\nabla\mu_0^* = 0.5$ and $\sigma_\nu = 0.015$ 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. If $\mu_0^* = \mu_0$ with all other parameters constant, and no omitted variables, then the VEqCM must prevail. Surprisingly, however, once model mis-specification is allowed, (38) can outperform (33) even for constant parameters, as we now show.

5.2 Constant-parameter case

In the constant-parameter DGP (29), both VEqCM and DDD are mis-specified, but in different ways, so a contrast of their forecasts for $\mathbf{x}_t$ is useful. The 1-step forecast error from the VEqCM is $\boldsymbol{\Upsilon}_0\mathbf{z}_{T+1} + \boldsymbol{\nu}_{T+1}$ where from (30):

$$\mathsf{E}\left[\boldsymbol{\Upsilon}_0\mathbf{z}_{T+1} + \boldsymbol{\nu}_{T+1}\right] = \mathbf{0} \tag{46}$$

and:

$$\mathsf{V}\left[\boldsymbol{\Upsilon}_0\mathbf{z}_{T+1} + \boldsymbol{\nu}_{T+1}\right] = \boldsymbol{\Upsilon}_0\mathsf{V}\left[\mathbf{z}_t\right]\boldsymbol{\Upsilon}_0' + \boldsymbol{\Omega}_{\boldsymbol{\nu}}. \tag{47}$$

The DDD 1-step forecast error is $\Delta\mathbf{x}_{T+1} - \Delta\mathbf{x}_T = \mathbf{u}_{T+1|T}$ which has mean zero and variance:

$$\mathsf{V}\left[\mathbf{u}_{T+1}\right] = \boldsymbol{\alpha}_0\boldsymbol{\beta}_0'\mathsf{V}\left[\Delta\mathbf{x}_T\right]\boldsymbol{\beta}_0\boldsymbol{\alpha}_0' + \boldsymbol{\Upsilon}_0\mathsf{V}\left[\Delta\mathbf{z}_{T+1}\right]\boldsymbol{\Upsilon}_0' + 2\boldsymbol{\Omega}_{\boldsymbol{\nu}} \tag{48}$$

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:

$$\mathsf{V}\left[\Delta\mathbf{z}_t\right] = \left(\boldsymbol{\Phi} - \mathbf{I}_k\right)\mathsf{V}\left[\mathbf{z}_t\right]\left(\boldsymbol{\Phi} - \mathbf{I}_k\right)' + \boldsymbol{\Omega}_\eta. \tag{49}$$

Using (31) and (49), the difference between (47) and (48) is:

$$\boldsymbol{\Upsilon}_0\left(\boldsymbol{\Phi}\mathsf{V}\left[\mathbf{z}_t\right] + \mathsf{V}\left[\mathbf{z}_t\right]\boldsymbol{\Phi}' - \mathsf{V}\left[\mathbf{z}_t\right]\right)\boldsymbol{\Upsilon}_0' - \boldsymbol{\alpha}_0\boldsymbol{\beta}_0'\mathsf{V}\left[\Delta\mathbf{x}_T\right]\boldsymbol{\beta}_0\boldsymbol{\alpha}_0' - \boldsymbol{\Omega}_{\boldsymbol{\nu}}. \tag{50}$$

When $\boldsymbol{\Phi} = \mathbf{0}$ (or, of course, $\boldsymbol{\Upsilon}_0 = \mathbf{0}$), then, in the absence of parameter estimation uncertainty, the VEqCM forecast-error variance dominates that of the DDD, since (50) is negative semi-definite. However, if $\boldsymbol{\Phi} \simeq \mathbf{I}_k$, so $\mathbf{z}_t$ is near $\mathsf{I}(1)$ and the omitted variables are important in explaining $\mathbf{x}_t$, then the difference is:

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

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 unless these act like innovations as well.

5.2.1 Scalar illustration 2

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

$$\Delta x_{T+1} = \gamma_0 + \alpha_0\left(x_T - \mu\right) + \lambda_0 z_{T+1} + \nu_{T+1}.$$

Then (47) becomes:

$$\sigma_{\eta}^{2}\frac{\lambda_{0}^{2}}{1-\phi^{2}}+\sigma_{\nu}^{2} \tag{51}$$

since $\sigma_{z}^{2}=\sigma_{\eta}^{2}/\left(1-\phi^{2}\right)$; and (48) becomes:

$$2\alpha_{0}^{2}\frac{\left(\sigma_{\eta}^{2}\lambda_{0}^{2}+\sigma_{\nu}^{2}\right)}{2+\alpha_{0}}+2\sigma_{\eta}^{2}\lambda_{0}^{2}\frac{1}{1+\phi}+2\sigma_{\nu}^{2} \tag{52}$$

so the difference between (47) and (48) is:

$$\sigma_{\eta}^{2}\lambda_{0}^{2}\left(\frac{2\phi-1}{1-\phi^{2}}-\frac{2\alpha_{0}^{2}}{2+\alpha_{0}}\right)-\sigma_{\nu}^{2}\left(1+\frac{2\alpha_{0}^{2}}{2+\alpha_{0}}\right),$$

which will be positive only if $\phi>0.5$, but can certainly be positive (e.g., $\alpha_0=-0.1$, $\lambda_0=1$, $\sigma_{\eta}^{2}=\sigma_{\nu}^{2}$, $\phi>0.75$ would suffice). Thus, even in a constant parameter world, the 'naive' predictor $\widetilde{\Delta x}_{T+1|T}$ could outperform a (mis-specified) VEqCM.

5.3 Longer-period differences

A potential drawback of a DDD is its noisiness, so instead of (38), one might consider an average of recent growth rates, denoted ADD$_m$:

$$\Delta\breve{\mathbf{x}}_{T+1|T,m}=\frac{1}{(m+1)}\sum_{j=0}^{m}\Delta\mathbf{x}_{T-j}=\frac{1}{(m+1)}\Delta_{(m+1)}\mathbf{x}_{T}. \tag{53}$$

Notice that $\Delta\breve{\mathbf{x}}_{T+1|T,T-1}$ would use all the in-sample data when $m=T-1$, switching to (26) for $m=0$. A special case is the latest annual change:

$$\Delta\breve{\mathbf{x}}_{T+1|T,3}=\frac{1}{4}\sum_{j=0}^{3}\Delta\mathbf{x}_{T-j}=\frac{1}{4}\Delta_{4}\mathbf{x}_{T}. \tag{54}$$

While *ad hoc*, $\Delta\breve{\mathbf{x}}_{T+1|T,3}$ is an adaptive estimator of $\boldsymbol{\gamma}$ which is slower to reflect breaks than $\Delta\mathbf{x}_T$ but is smoother, so its empirical behaviour is noted below.[2]

6 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:

$$\Delta\mathbf{x}_{t}=\boldsymbol{\gamma}+\boldsymbol{\alpha}\Delta\left(\boldsymbol{\beta}'\mathbf{x}_{t-1}-\boldsymbol{\mu}\right)+\boldsymbol{\xi}_{t}=\boldsymbol{\gamma}+\boldsymbol{\alpha}\boldsymbol{\beta}'\Delta\mathbf{x}_{t-1}+\boldsymbol{\xi}_{t}. \tag{55}$$

[2] A referee noted that this explanation may also show why some *ad hoc* forecasting devices in the finance literature, such as a buy-and-hold strategy, moving-average trading (similar to ADD$_m$), or a momentum strategy (similar to (26)) are often found to out-perform more 'sophisticated' forecasts based on structural models.

In this simple setting, the effect in (55) 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 $\boldsymbol{\alpha\beta}'$. Since shifts in $\boldsymbol{\mu}$ are the most pernicious for forecasting, (55) might be more robust to such breaks than the original VEqCM (9). On the other hand, there will be a loss of information during periods where no breaks occur.

To examine the behaviour of (55) forecasting $\Delta\mathbf{x}_{T+2}$ from $T+1$ after a break in $\boldsymbol{\mu}$ at time T, let:[3]

$$\overline{\Delta\mathbf{x}}_{T+2|T+1} = \boldsymbol{\gamma} + \boldsymbol{\alpha\beta}'\Delta\mathbf{x}_{T+1} \tag{56}$$

so the forecast error is:

$$\Delta\mathbf{x}_{T+2} - \overline{\Delta\mathbf{x}}_{T+2|T+1} = \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{T+1} - \boldsymbol{\mu}^*\right) + \boldsymbol{\epsilon}_{T+2} - \boldsymbol{\gamma} - \boldsymbol{\alpha\beta}'\Delta\mathbf{x}_{T+1}. \tag{57}$$

Since:

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

then:

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

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 (57), 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:

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

so the mean forecast error will gradually decline. Although (56) will induce a smaller increase in the error variance than (38), namely $\boldsymbol{\Omega}_\epsilon + \boldsymbol{\alpha\beta}'\boldsymbol{\Omega}_\epsilon\boldsymbol{\beta\alpha}'$ rather than $2\boldsymbol{\Omega}_\epsilon$, merely eliminating the equilibrium mean by differencing does not seem advantageous. Moreover, (56) remains vulnerable to shifts in $\boldsymbol{\gamma}$.

6.1 Differencing the VEqCM

Since shifts in $\boldsymbol{\gamma}$ are the next most pernicious for forecasting, we consider forecasting not from (9) itself, but from a variant thereof which is the difference of the estimated congruent representation, namely:

$$\Delta\mathbf{x}_t = \Delta\mathbf{x}_{t-1} + \boldsymbol{\alpha\beta}'\Delta\mathbf{x}_{t-1} + \Delta\boldsymbol{\epsilon}_t = \left(\mathbf{I}_n + \boldsymbol{\alpha\beta}'\right)\Delta\mathbf{x}_{t-1} + \boldsymbol{\zeta}_t = \boldsymbol{\Gamma}\Delta\mathbf{x}_{t-1} + \boldsymbol{\zeta}_t \tag{58}$$

where (58) is just the first difference of the original VAR, denoted DVAR, since $\left(\mathbf{I}_n + \boldsymbol{\alpha\beta}'\right) = \boldsymbol{\Gamma}$, but with the rank restriction from cointegration imposed. Alternatively, $\Delta\mathbf{x}_{t-1}$ can be interpreted as

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

a highly adaptive estimator of γ in (9), where $\boldsymbol{\mu}$ is then also approximated by the previous value of the cointegrating combination, $\boldsymbol{\beta}'\mathbf{x}_{t-2}$, so both parameters are replaced by instantaneous unbiased estimators, leading to:[4]

$$\Delta\mathbf{x}_t = \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\mu}\right) + \boldsymbol{\epsilon}_t \simeq \Delta\mathbf{x}_{t-1} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\beta}'\mathbf{x}_{t-2}\right) + \boldsymbol{\epsilon}_t^* = \Delta\mathbf{x}_{t-1} + \boldsymbol{\alpha}\boldsymbol{\beta}'\Delta\mathbf{x}_{t-1} + \boldsymbol{\epsilon}_t^*. \tag{59}$$

In (59):

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

so the errors' composition in $\mathsf{I}(0)$ (rather than $\mathsf{I}(-1)$) terms is revealed on this interpretation, albeit still being $\Delta\boldsymbol{\epsilon}_t$. This form has the added advantage of clarifying the problem at the point at which breaks actually occur in $\boldsymbol{\gamma}$ and $\boldsymbol{\mu}$, namely:

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

so the forecast bias is $\left(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}\right) + \boldsymbol{\alpha}\left(\boldsymbol{\mu} - \boldsymbol{\mu}^*\right)$ at that point, but rapidly vanishes as the values of $\Delta\mathbf{x}$ and $\boldsymbol{\beta}'\mathbf{x}$ converge on their new means.

A second representation from (58) is:

$$\Delta^2\mathbf{x}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\Delta\mathbf{x}_{t-1} + \boldsymbol{\zeta}_t. \tag{60}$$

where (60) can be interpreted as augmenting the DDD forecast by $\boldsymbol{\alpha}\boldsymbol{\beta}'\Delta\mathbf{x}_{t-1}$, 'adding back' to the DDD the main observable component omitted by using just the lagged first difference, shown in (40). Since zero-mean shifts in parameters do not have a major effect on forecast accuracy, such a strategy is likely to have some benefits even if $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$ change. The remaining terms are all unknown, so absent knowledge of some of the $\mathbf{z}_t$, no further improvements are feasible down this route. Thus, a DDD is not only the difference of a first-order DVAR, but is also obtained by dropping the mean-zero term $\boldsymbol{\alpha}\boldsymbol{\beta}'\Delta\mathbf{x}_{t-1}$ from the simplest DVEqCM.

A third representation of (58) is:

$$\Delta\mathbf{x}_t = \Delta\mathbf{x}_{t-1} + \boldsymbol{\alpha}\boldsymbol{\beta}'\Delta\mathbf{x}_{t-1} + \Delta\boldsymbol{\epsilon}_t = \boldsymbol{\gamma} + \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\mu}\right) + \left[\Delta\mathbf{x}_{t-1} - \boldsymbol{\gamma} - \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{t-2} - \boldsymbol{\mu}\right)\right] + \Delta\boldsymbol{\epsilon}_t \tag{61}$$

The term in brackets, $\Delta\mathbf{x}_{t-1} - \boldsymbol{\gamma} - \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\mathbf{x}_{t-2} - \boldsymbol{\mu}\right) = \boldsymbol{\epsilon}_{t-1}$, so acts like an intercept correction (IC), adding back the previous realised error (see e.g., Hendry and Clements, 1994). However, the effect is not the same as an IC given the different error term which results from adding and subtracting $\boldsymbol{\epsilon}_{t-1}$.

Like a DDD, the DVEqCM in (60) has no deterministic terms (e.g., constants or linear trends, t), and hence does not equilibrium correct, thereby reducing the risks attached to VEqCMs. However, it will produce noisy forecasts, although smoothed variants are easily formulated like ADD_m. When there

[4] I am indebted to an anonymous referee for this suggestion.

are no locations shifts, the 'insurance' of differencing must worsen forecast accuracy and precision, so is a clear cost of using that strategy. Conversely, if location shifts do occur, the policy will pay. We consider both cases, beginning with location shifts. From section 5, we can recover results for mis-specified models of an extended DGP by substituting $\boldsymbol{\epsilon}_t = \boldsymbol{\Upsilon}_0 \mathbf{z}_t + \boldsymbol{\nu}_t$: that section also established most of the algebraic results needed below.

To trace the behaviour of (58) after a break in $\boldsymbol{\mu}$, let:

$$\widetilde{\widetilde{\Delta \mathbf{x}}}_{T+1|T} = \left(\mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}'\right) \Delta \mathbf{x}_T \tag{62}$$

where from (21):

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

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:

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

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:

$$\mathsf{E}\left[\Delta \mathbf{x}_{T+2} - \widetilde{\widetilde{\Delta \mathbf{x}}}_{T+2|T+1}\right] = \boldsymbol{\gamma} - \boldsymbol{\alpha}\boldsymbol{\Psi}\nabla\boldsymbol{\mu}^* - \left(\boldsymbol{\gamma} - \boldsymbol{\alpha}\nabla\boldsymbol{\mu}^*\right) + \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{\alpha}\nabla\boldsymbol{\mu}^* = \mathbf{0}. \tag{63}$$

Thus, the differenced VEqCM 'misses' only for the forecast origin, 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 $\boldsymbol{\mu}$, (60) should outperform, especially if $\boldsymbol{\gamma}$ also alters. Moreover, (63) demonstrates that $\widetilde{\widetilde{\Delta \mathbf{x}}}$ dominates $\widetilde{\Delta \mathbf{x}}$ in (28) in mean.

If all parameters are constant, (62) remains unbiased but inefficient, even at the forecast origin:

$$\mathsf{E}\left[\Delta \mathbf{x}_{T+1} - \widetilde{\widetilde{\Delta \mathbf{x}}}_{T+1|T}\right] = \mathbf{0}.$$

The next sub-section considers the impact of unnecessary differencing on forecast-error variances, in the context of 1-step ahead forecasts.

6.1.1 Forecast-error variances

Let $\widetilde{\widetilde{\mathbf{e}}}_{T+h|T+h-1} = \Delta \mathbf{x}_{T+h} - \widetilde{\widetilde{\Delta \mathbf{x}}}_{T+h|T+h-1}$ be the sequence of 1-step forecast errors from updating (62), then, ignoring parameter estimation variances as $\mathsf{O}_p\left(T^{-1}\right)$:

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

whereas:

$$\widetilde{\widetilde{\mathbf{e}}}_{T+2|T+1} = \Delta\boldsymbol{\epsilon}_{T+2}.$$

Relative to a DDD, therefore, there is a gain from the DVEqCM, since the former also 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 and any omitted variables errors. Thus, both central tendency and variability should be better for the DVEqCM than a DDD, at least in the absence of parameter estimation uncertainty. When $\boldsymbol{\epsilon}_t = \boldsymbol{\Upsilon}_0\mathbf{z}_t + \boldsymbol{\nu}_t$ we obtain similar results to those for a DDD depending on the properties of $\{\mathbf{z}_t\}$.

These findings provide a clear ranking of forecast performance after a location shift in $\boldsymbol{\mu}$: the DVEqCM should dominate DDD in both mean and variance and hence in MSFE; DDD should dominate VEqCM in mean, but not necessarily in variance, so MSFE comparisons depend on the relative magnitudes of shifts to error variances; and hence DVEqCM should dominate VEqCM in mean, but not necessarily in variance or MSFE depending on the magnitude of any shift. Additional unknown omitted variables strengthen such an ordering. We turn to an empirical illustration with a known location shift to evaluate these predictions.

7 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, zero before 1984(2),
R_{net}	$R_{la} - R_o$.

All computations are based on PcGive: see Doornik and Hendry (2001).

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.[5]

Figure 2 here

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 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}.[6] 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 $R_n^c = R_{la} - 0.072 \times 1_{\{t>1985(2)\}}$ 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 used 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.

7.1 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 $\widehat{\boldsymbol{\beta}}'\mathbf{x}_t$, denoted EqCM[$R_{la}$]: this model

[5] 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.

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

uses current-dated values of R_{la} 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, denoted DEqCM[R_{la}]: there is a very substantial improvement, with no systematic under-forecasting, suggesting that the adaptation proposed in section 6.1 can be effective in the face of equilibrium-mean shifts. All 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 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 here

Figure 3c shows the good performance on 20 1-step forecasts of the 'correct' model EqCM[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, denoted DEqCM[R_{net}]. This produces similar forecasts to EqCM[R_{net}], 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 DDD from section 5, are shown in figure 4 panel a. The DDD 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.[7] Moreover, the benefits from using either DEqCM are marked relative to the DDD, consistent with the earlier theory that including $\boldsymbol{\alpha}\boldsymbol{\beta}'\Delta\mathbf{x}_{t-1}$ would improve performance. Finally, that the MSFE has essentially doubled relative to EqCM[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.

Figure 4 here

The ADD_4 forecasts shown in figure 4c are distinctly better than the DDD, having a mean forecast

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

error of -0.07% and an RMSFE of 1.8%. Hence some degree of smoothing seems to pay. This is also true of the ADD_4 and DDD forecasts for R_{net} shown in figure 4 panels b and d (ADD_4 has an RMSFE of 1.5% as against DDD of 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}] forecasts.

7.2 System behaviour

In a system context, there are three major aspects to consider for most of the methods, although the DDD and ADD_4 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[R_{net}] forecast errors for $\Delta(m-p)$ for comparison with the conditional single-equation forecast errors. It also shows the corresponding DVEqCM[R_{net}] forecast errors to highlight the small loss from the additional differencing of the correct specification. The forecasts from VEqCM[R_{la}] are as poor as the single equation ones for $\Delta(m-p)$, but differencing that VEqCM again corrects the main forecast error bias, delivering errors similar to those of DEqCM[R_{la}].

Figure 5 here

Secondly, multi-step forecasts can now be calculated. These serve to confirm the above results, and while more representative 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 up to the break, so are their multi-step forecasts for any horizon h: for later forecast origins, however, they will behave distinctly differently, with that based on R_{la} continuing to perform poorly whereas that based on R_{net} will perform well once the location shift is past. Conversely, the DDD 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 VEqCM[R_{la}] becomes a shift in the R_{net} equation in VEqCM[R_{net}], 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 VEqCM[R_{net}] does noticeably better for $(m-p)$ but is still unable to forecast R_{net} very well.

Figure 6 here

Other aspects of adaptive forecasting could be incorporated with any of the above devices, including intercept corrections, recursive updating of parameter estimates, and re-selecting the relevant variables (see e.g., Phillips, 1994). The first of these should 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 DDD than a VEqCM, as the additional differencing eliminates some of the adverse effects of the shift. The third accelerates the tendency just noted.

8 Conclusions

Using a cointegrated linear dynamic system with breaks over the forecast horizon as the illustrative DGP, two adaptations were considered to improve the robustness of forecasts from equilibrium-correction systems. The first was using second differences to forecast (DDD); the second was forecasting from the differenced VEqCM (denoted DVEqCM). Both proposals were shown to help robustify forecasts against unanticipated location shifts, particularly shifts in the equilibrium means.

A new theoretical explanation for the relative success of the DDD was proposed, as capturing all changes and mis-specifications in the DGP, albeit with a lag. Despite the DDD being mis-specified for the local DGP, it could outperform a VEqCM even in constant-parameter processes, when the latter used incomplete information, as seems inevitable in practice. Then the DVEqCM was related to the DDD as also retaining one of the key observable components that affected the latter's forecast errors, namely the change in the equilibrium correction.

These results provided a ranking across the three methods when forecasting after a location shift: the DVEqCM should dominate DDD in both mean and variance and hence in MSFE; DDD should dominate VEqCM in mean, but not necessarily in variance, so MSFE comparisons depended on the magnitudes of shifts relative to error variances; and hence DVEqCM should dominate VEqCM in mean, but not necessarily in variance or MSFE. Additional unknown omitted variables would strengthen such an ordering.

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. The VEqCM confronting the location shift had a RMSFE of 4.9% as against 2.25% for the DDD and 1.8% for its DVEqCM (compared to an equation standard error of 1.3%). The corresponding mean forecast errors were 4.4%, 0.001% and 0.4% respectively. These demonstrate the potentially large 'insurance payoff' from differencing in the face of a location shift. There seemed to be some benefit from using a smoothed DDD, in that the four-period moving average, denoted ADD_4, had a mean forecast error of -0.07% and an RMSFE of 1.8%, which is competitive with both DVEqCMs, and suggests investigating some smoothing for those.

For the post-break correctly-specified EqCM, where there was no location shift, its RMSFE was 1.14% as against 1.79% for its DEqCM, which is essentially the same as from differencing the incorrect model. That measures the 'insurance cost' when needlessly differencing a constant model, and is close to a factor of $\sqrt{2}$. However, when the post-break behaviour of the new opportunity-cost measure also had to be forecast in a VEqCM, much of the benefit of the appropriate specification was lost, making either DVEqCM a competitive forecasting alternative when location shifts occur.

All the approaches behaved as anticipated from the theory, and demonstrated the difficulty of out-performing 'naive extrapolative devices' when these 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.

Acknowledgements

Financial support from the ESRC under grant RES 051 270035, and helpful comments from Jennifer Castle, Mike Clements, Neil Ericsson, Katarina Juselius, Kirstin Hubrich, Grayham Mizon, and two anonymous referees are gratefully acknowledged.

References

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

Allen, P. G., and Fildes, R. A. (2004). Levels, differences and ECMs–principles for improved econometric forecasting. Discussion paper, Department of Management Science, University of Lancaster.

Bhansali, R. J. (1996). Asymptotically efficient autoregressive model selection for multistep prediction. *Annals of Institute of Statistical Mathematics*, **48**, 94–134.

Bhansali, R. J. (1999). Parameter estimation and model selection for multistep prediction of time series: a review. In Gosh, S. (ed.), *Asymptotics, Nonparametrics and Time Series*, pp. 201–225. New York, NY: Marcel Dekker.

Bhansali, R. J. (2002). Multi-step forecasting. in Clements, and Hendry (2002), pp. 206–221.

Bontemps, C., and Mizon, G. E. (2003). Congruence and encompassing. In Stigum, B. P. (ed.), *Econometrics and the Philosophy of Economics*, pp. 354–378. Princeton: Princeton University Press.

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

Chevillon, G., and Hendry, D. F. (2005). Non-parametric direct multi-step estimation for forecasting economic processes. *International Journal of Forecasting*, **21**, 201–218.

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. (1996a). 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. (1996b). Multi-step estimation for forecasting. *Oxford Bulletin of Economics and Statistics*, **58**, 657–684.

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.

Clements, M. P., and Hendry, D. F. (eds.)(2002). *A Companion to Economic Forecasting*. Oxford: Blackwells.

Clements, M. P., and Hendry, D. F. (2004). Forecasting with breaks in data processes. Chapter prepared for *Handbook of Econometrics on Forecasting*, Granger, C.W.J., Timmermann, A. and Elliott, G. (eds), Elsevier.

Doornik, J. A., and Hendry, D. F. (2001). *Modelling Dynamic Systems using PcGive 10: Volume II*. London: Timberlake Consultants Press.

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, and Hendry (2002), pp. 322–253.

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; and in J. Campos, N.R. Ericsson and D.F. Hendry (eds.), *General to Specific Modelling*. Edward Elgar, 2005.

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, R. (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 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 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. Reprinted in J. Campos, N.R. Ericsson and D.F. Hendry (eds.), *General to Specific Modelling*. Edward Elgar, 2005.

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.

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.

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.

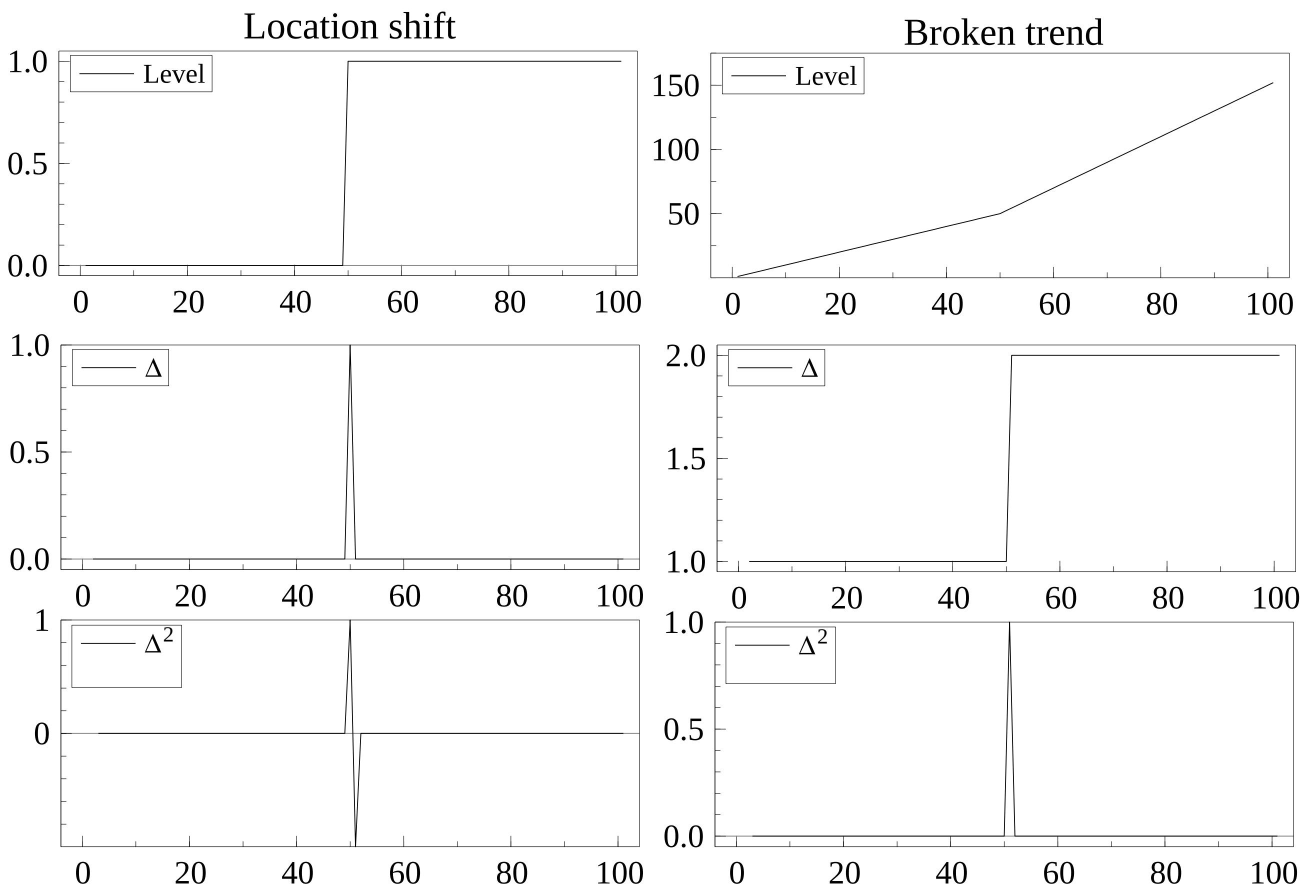

Figure 1 Location shifts and broken trends.

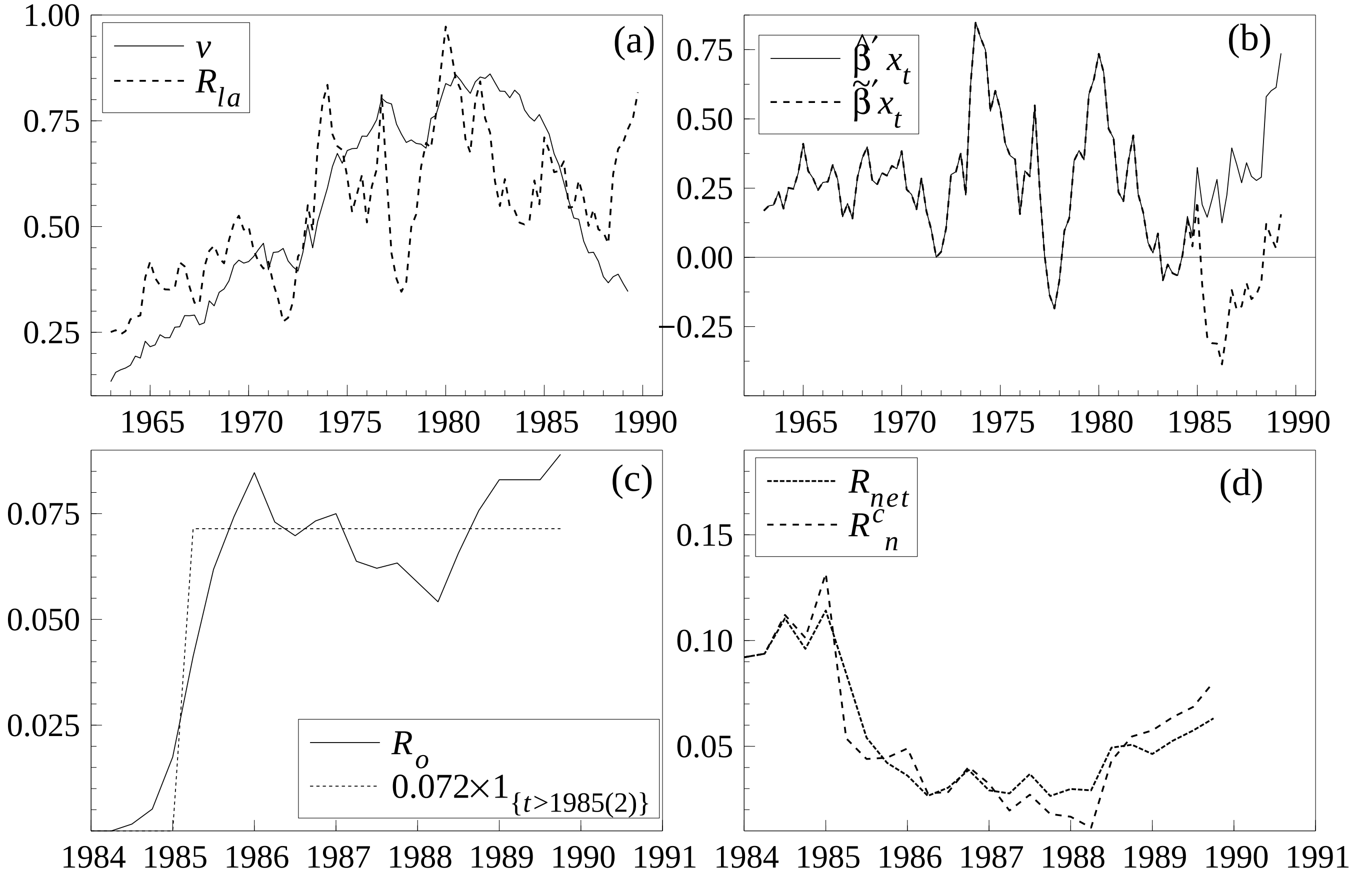

Figure 2 Effects of the 1984 Banking Act on UK M1.

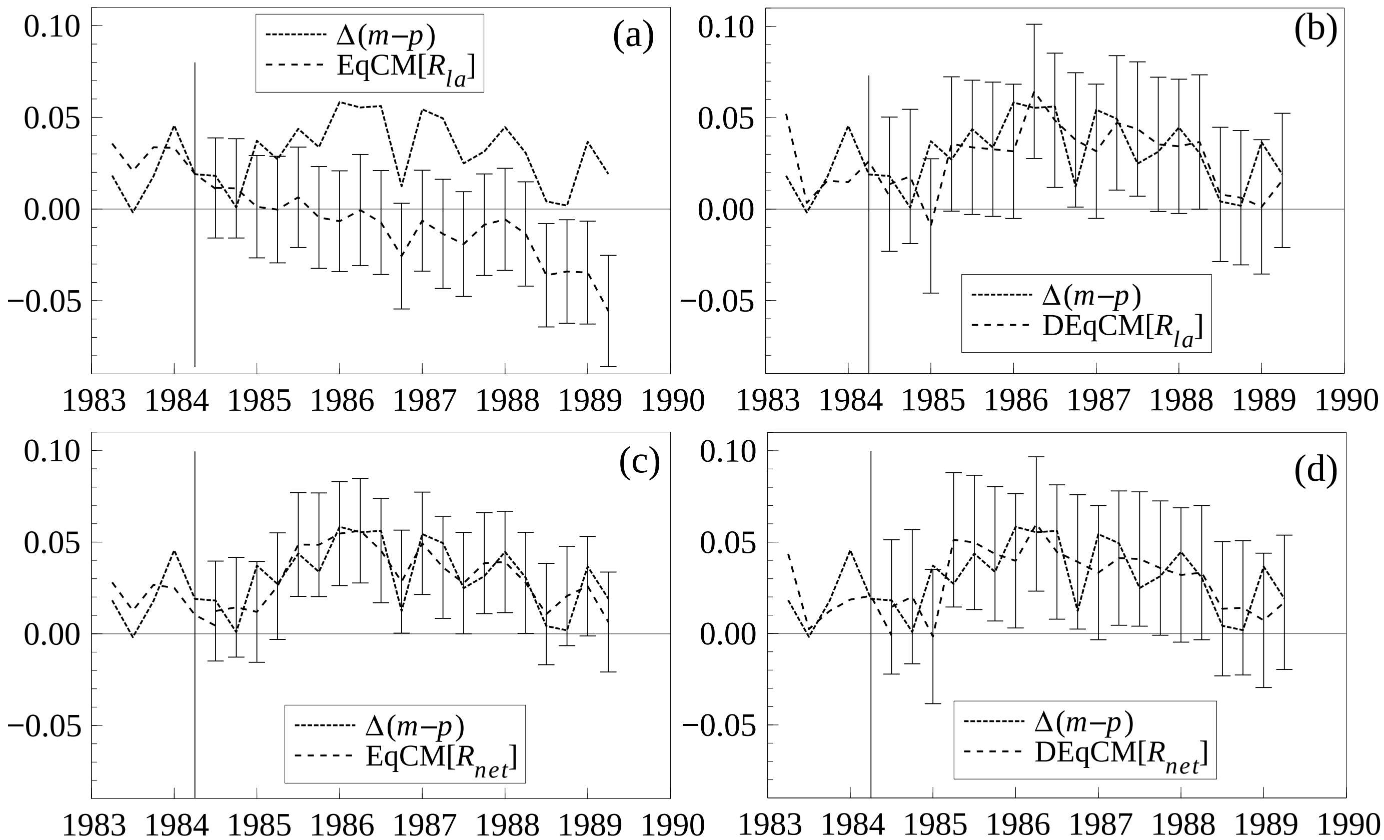

Figure 3 1-step forecasts of UK M1 from conditional models.

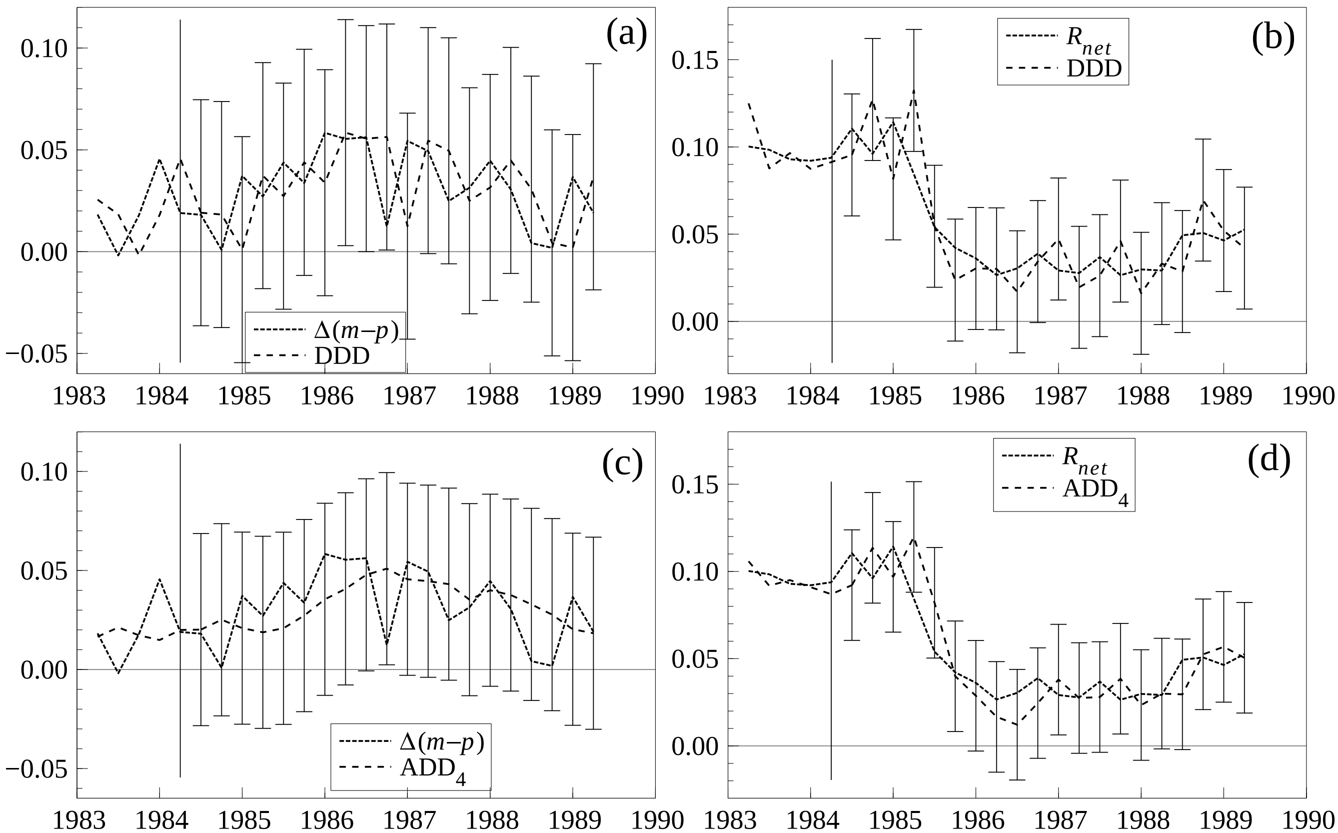

Figure 4 DDD and ADD_4 1-step forecasts of UK M1.

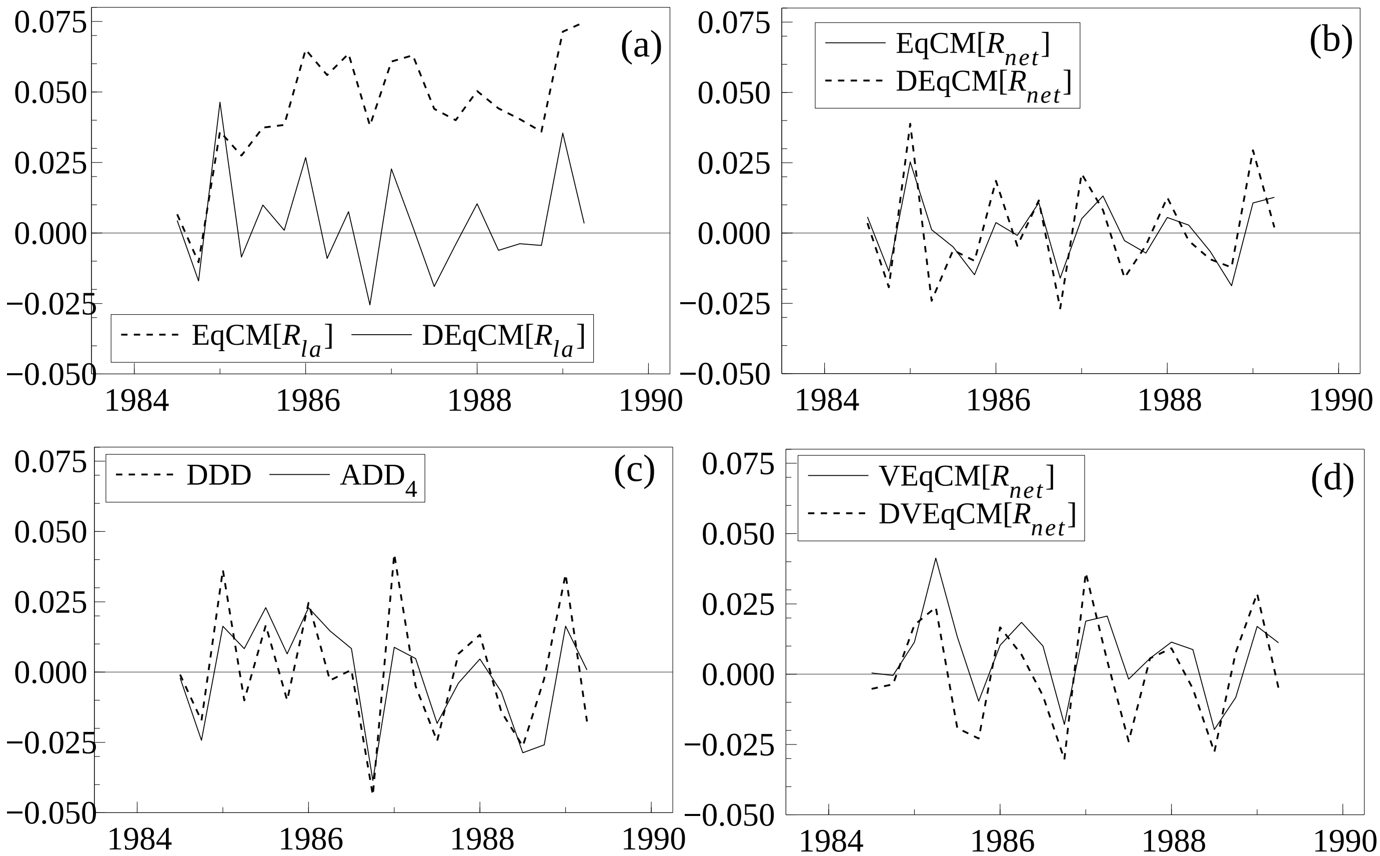

Figure 5 1-step forecast errors for different models of UK M1.

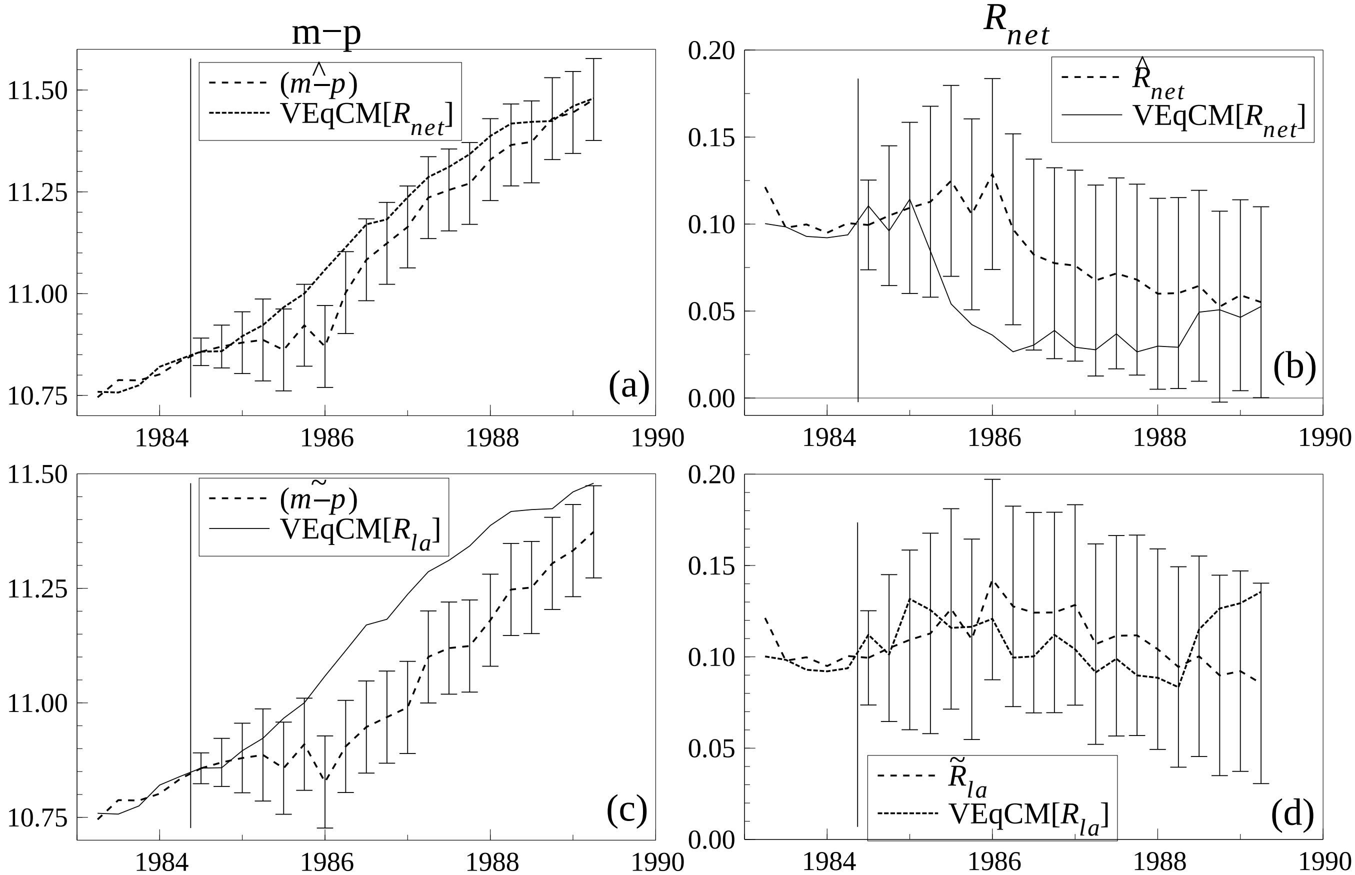

Figure 6 System 4-step forecasts from two VEqCMs of UK M1.