

Intercept Corrections and Structural Change

Michael P. Clements
Department of Economics,
University of Warwick

and

David F. Hendry*
Institute of Economics and Statistics, and
Nuffield College, Oxford.

March 31, 2008

Abstract

Analyses of forecasting that assume a constant, time-invariant data generating process (DGP), and so implicitly rule out structural change or regime shifts in the economy, ignore an aspect of the real world responsible for some of the more dramatic historical episodes of predictive failure. Some models may offer greater protection against unforeseen structural breaks than others, and various tricks may be employed to robustify forecasts to change. We show that in certain states of nature, vector autoregressions in the differences of the variables (in the spirit of Box-Jenkins time-series modelling), can outperform vector ‘equilibrium-correction’ mechanisms. However, appropriate intercept corrections can enhance the performance of the latter, albeit that reductions in forecast bias may only be achieved at the cost of inflated forecast error variances.

First draft February 1995. Revised May 1996.

1 Introduction

In recent work (Clements and Hendry, 1994, 1995b), we have sought to establish a theory of economic forecasting that captures three aspects of the real world in which the forecasting venture is to be undertaken. First, that the data generation process (DGP) is non-stationary due to unit roots; secondly, that it is susceptible to structural breaks; and thirdly, that the forecasting model typically differs from the (unknown) DGP. These features provide a rationale for the commonplace practice of making adjustments or ‘intercept corrections’ to purely model-based forecasts (see Hendry and Clements, 1994a, 1994b).¹ A Monte Carlo study showed that forecasts generated from vector autoregressions in differences (DVARs) may be more robust than models in levels to certain forms of structural change, but that intercept corrections may help vector ‘equilibrium-correction’ mechanisms (VECMs) to match the performance of DVARs.²

An interesting example of the benefit of ignoring long-run information for forecasting would appear to be Mizon (1995), who shows that only a DVAR has a satisfactory forecasting performance in the context of modelling UK wages and prices over the period 1966–1993. Models such as VECMs, which include long-run information, tend to fail badly. The models are estimated on data up to Mrs. Thatcher

*Financial support from the U.K. Economic and Social Research Council under grant L116251015 is gratefully acknowledged. The computations were performed using PcFiml version 8 (Doornik and Hendry, 1994) and the Gauss programming language, Aptech Systems, Inc., Washington. Helpful comments were received from the editors, Frank Diebold and Mark Watson, and numerous seminar participants.

¹Recognition of the potential for such adjustments has a long lineage: see, for example, Theil (1961) and Klein (1971).

²Following Davidson, Hendry, Srba and Yeo (1978), these terms have been known as ‘error-corrections’. However, they may play the opposite role when the equilibrium changes by ‘correcting’ to an inappropriate equilibrium. Hence the change in terminology, although the acronym is unchanged.

coming to power (1979:2), and then used to forecast the behaviour of wages, prices, and unemployment during the 1980s. There is anecdotal evidence of structural change in the 1980s in response to the dislocating effect of the 1979–81 recession, and the change in government economic policies may have altered the long-run relationship between these three macro aggregates. Thus, models which include long-run information tend to ‘error-correct’ on the basis of an out-dated structure, and manifest significant forecast errors, while models that eschew such information perform reasonably well.

The purpose of this paper is two-fold. First, we deduce analytically the classes of structural breaks for which, conditional on the break having occurred at the time of forecasting, time-series models in differences should outperform econometric models. We then look at the usefulness of intercept corrections for the class of breaks that affect the deterministic variables (constant and linear trend), and derive expressions for the forecast bias and error variances, allowing an assessment of intercept-correcting strategies in terms of squared-error loss. Secondly, we apply some of these correction strategies to econometric models based on Mizon (1995), where wider classes of breaks than those we analyse analytically may be important, to investigate their usefulness in practice. The three-variable system of wages, prices, and unemployment is simpler than would be countenanced by the large-scale macro-econometric modelling groups, but has the virtue of allowing a ‘statistical analysis’ of intercept corrections. Previous studies, such as those carried out by the ESRC Macroeconomic Modelling Bureau at Warwick, have assessed the impact of intercept corrections on actual forecasts of the main modelling teams (see, in particular, Wallis and Whitley, 1991, and Turner, 1990), and have been confined to only a small number of forecasts (typically less than three). Our setup allows an assessment of the impact of intercept corrections on forecast uncertainty using empirical mean-square forecast errors, taking account of their dependence on the forecast horizon and on the transformation of the data for which forecast accuracy is assessed. We distinguish between 1-step and h -steps ahead forecasting performance, given the lack of invariance of mean-square forecast errors (MSFEs) to evaluating forecasts of levels versus changes (say), and the likely poor discriminatory performance of evaluation in differences (see Clements and Hendry, 1993, 1995a).

Sections 2 and 3 derive analytical results on the forecast performance of the VECM and DVAR models when there are structural breaks, and on the impact of intercept correcting the VECM. Section 4 explains the relevance of some of our recent work on forecast evaluation and multi-step forecasting for the present analysis. Section 5 introduces the empirical example of Mizon (1995), and section 6 contains the results of the empirical study of the efficacy of intercept corrections. Section 7 provides some concluding remarks.

2 Forecasting under structural breaks

2.1 The data generation process

For illustrative purposes, we assume a linear, closed system so that all non-deterministic variables are forecast within the system. The vector of all n variables is denoted by \mathbf{w}_t and the system is represented by a first-order VAR which includes a constant and a linear deterministic trend:

$$\mathbf{w}_t = \boldsymbol{\tau}_0 + \boldsymbol{\tau}_1 t + \boldsymbol{\Upsilon} \mathbf{w}_{t-1} + \boldsymbol{\nu}_t \quad (1)$$

where $\boldsymbol{\nu}_t \sim \text{IN}_n(\mathbf{0}, \boldsymbol{\Omega})$. The system is assumed to be integrated, and to satisfy $r < n$ cointegration relations such that (see, for example, Johansen, 1988):

$$\boldsymbol{\Upsilon} = \mathbf{I}_n + \boldsymbol{\alpha} \boldsymbol{\beta}',$$

where α and β are $n \times r$ matrices of rank r . Then (1) can be reparameterized as a VECM:

$$\Delta \mathbf{w}_t = \tau_0 + \tau_1 t + \alpha \beta' \mathbf{w}_{t-1} + \nu_t. \quad (2)$$

The impact of the deterministic components on the series depends on the relationship between α and τ_0, τ_1 . Following Johansen (1994), decompose the $2n$ parameters in $\tau_0 + \tau_1 t$ as:

$$\tau_0 + \tau_1 t = \alpha_{\perp} \zeta_0 - \alpha \lambda_0 - \alpha \lambda_1 t + \alpha_{\perp} \zeta_1 t \quad (3)$$

where $\lambda_i = -(\alpha' \alpha)^{-1} \alpha' \tau_i$ ($2r$ free parameters) and $\zeta_i = (\alpha'_{\perp} \alpha_{\perp})^{-1} \alpha'_{\perp} \tau_i$ ($2(n-r)$ free parameters) when $\alpha' \alpha_{\perp} = \mathbf{0}$. Then $\alpha \lambda_i$ and $\alpha_{\perp} \zeta_i$ are orthogonal by construction. When the \mathbf{w}_t process does not contain a quadratic trend, $\alpha_{\perp} \zeta_1 = \mathbf{0}$, and λ_0, ζ_0 and λ_1 can all be varied freely. The \mathbf{w}_t process may still contain linear trends, which will also be a feature of the cointegrating vectors, as seems to be the case in the empirical example. Then $\alpha \lambda_0 + \alpha \lambda_1 t$ lies in the cointegration space, and (2) can be written as:

$$\Delta \mathbf{w}_t = \alpha_{\perp} \zeta_0 + \alpha (\beta' \mathbf{w}_{t-1} - \lambda_0 - \lambda_1 t) + \nu_t. \quad (4)$$

When the system grows at the (vector) rate:

$$\mathbb{E} [\Delta \mathbf{w}_t] = \gamma \quad (5)$$

from (4), we obtain:

$$\alpha \mathbb{E} [\beta' \mathbf{w}_{t-1}] = \gamma - \alpha_{\perp} \zeta_0 + \alpha (\lambda_0 + \lambda_1 t). \quad (6)$$

When $\beta' \alpha$ is non-singular, as assumed throughout:

$$\mathbb{E} [\beta' \mathbf{w}_{t-1}] = (\beta' \alpha)^{-1} \beta' (\gamma - \alpha_{\perp} \zeta_0) + \lambda_0 + \lambda_1 t = \psi + \lambda_0 + \lambda_1 t. \quad (7)$$

Hence, in (4):

$$\Delta \mathbf{w}_t = \alpha_{\perp} \zeta_0 + \alpha \psi + \alpha (\beta' \mathbf{w}_{t-1} - \psi - \lambda_0 - \lambda_1 t) + \nu_t.$$

It is useful for subsequent calculations to introduce the idempotent matrix $\mathbf{K} = (\mathbf{I}_n - \alpha (\beta' \alpha)^{-1} \beta')$ such that $\mathbf{K} \alpha = \mathbf{0}$, $\beta' \mathbf{K} = \mathbf{0}$, $\mathbf{K}^2 = \mathbf{K}$ and $\Upsilon \mathbf{K} = \mathbf{K}$, implying that $\mathbf{K} \tau_0 = \mathbf{K} \gamma = \mathbf{K} \alpha_{\perp} \zeta_0$ and $\mathbf{K} \tau_1 = \mathbf{0}$. The cost of orthogonality in (4) is that the cointegrating vectors are no longer deviations about their means, so the ‘intercept’ $\alpha_{\perp} \zeta_0$ is not the growth rate γ which can be expressed as:³

$$\gamma = \alpha_{\perp} \zeta_0 + \alpha \psi = \mathbf{K} \alpha_{\perp} \zeta_0 + \alpha (\beta' \alpha)^{-1} \lambda_1 \quad (8)$$

since from (7) and (5):

$$\Delta \mathbb{E} [\beta' \mathbf{w}_t] = \beta' \mathbb{E} [\Delta \mathbf{w}_t] = \lambda_1 \text{ so that } \beta' \gamma = \lambda_1. \quad (9)$$

Consequently, we can rewrite the VECM as:

$$\Delta \mathbf{w}_t = \gamma + \alpha (\beta' \mathbf{w}_{t-1} - \mu_0 - \mu_1 t) + \nu_t \quad (10)$$

where $\mu_0 = \psi + \lambda_0$ and $\mu_1 = \lambda_1$ with:

$$\psi = (\beta' \alpha)^{-1} \beta' (\gamma - \alpha_{\perp} \zeta_0) = (\beta' \alpha)^{-1} (\lambda_1 - \beta' \alpha_{\perp} \zeta_0)$$

³In previous work (eg. Clements and Hendry, 1995b, p.1005) with $\tau_1 = \mathbf{0}$, we have used a simpler, non-orthogonal decomposition of τ_0 . We are grateful to Bent Nielsen for bringing to our attention the problems with such an approach in the presence of the linear trend term.

Finally, a VAR in differences (DVAR) may be used, which within sample is mis-specified relative to the VECM unless $r = 0$. The simplest is:

$$\Delta \mathbf{w}_t = \gamma + \boldsymbol{\eta}_t, \quad (11)$$

so when $\alpha = \mathbf{0}$, the VECM and DVAR coincide. In practice, lagged $\Delta \mathbf{w}_t$ may be used to approximate the omitted cointegrating vectors, but we do not consider such a model here as its behaviour under structural breaks is rather complicated: Appendix I notes its derivation.

2.2 VECM forecast errors

We now consider dynamic forecasts and their errors when parameters are subject to change in the forecast period. We draw on the analyses in Clements and Hendry (1994) and Hendry and Clements (1994b), focusing on the bias and variance components. We also consider the implications of the deterministic terms lying in the cointegrating space. For simplicity, we abstract from many of the potential sources of forecast error discussed in those two papers. Thus, the forecast commences from correct initial conditions (equal to the true value of the process, \mathbf{w}_T), we assume that the model matches the DGP in-sample, and we abstract from parameter estimation uncertainty, so that the forecast functions are based on the true (but sample period) values of the process. The algebra represents a Monte Carlo where the same in-sample value is used, but for different forecast values; to vary both and get the average outcome, we must take expectations over \mathbf{w}_T . It is crucial how this is done relative to the structural change: (a) change occurs at $T + 1$; (b) change occurred at T .

Under these assumptions, the j -step ahead forecasts for the levels of the process are given by $\hat{\mathbf{w}}_{T+j} = \mathbf{E}[\mathbf{w}_{T+j} \mid \mathbf{w}_T]$:

$$\hat{\mathbf{w}}_{T+j} = \boldsymbol{\tau}_0 + \boldsymbol{\tau}_1(T+j) + \boldsymbol{\Upsilon} \hat{\mathbf{w}}_{T+j-1} = \sum_{i=0}^{j-1} \boldsymbol{\Upsilon}^i \boldsymbol{\tau}(i) + \boldsymbol{\Upsilon}^j \mathbf{w}_T \text{ for } j = 1, \dots, h \quad (12)$$

where we let $\boldsymbol{\tau}_0 + \boldsymbol{\tau}_1(T+j-i) = \boldsymbol{\tau}(i)$ for notational convenience. The associated forecast errors are:

$$\hat{\boldsymbol{\nu}}_{T+j} = \mathbf{w}_{T+j} - \hat{\mathbf{w}}_{T+j}.$$

We consider the situation where the system experiences a step change between the estimation and forecast periods, such that $(\boldsymbol{\tau}_0 : \boldsymbol{\tau}_1 : \boldsymbol{\Upsilon})$ changes to $(\boldsymbol{\tau}_0^* : \boldsymbol{\tau}_1^* : \boldsymbol{\Upsilon}^*)$ over $j = 1, \dots, h$, but the variance, autocorrelation, and distribution of the disturbance term remain unaltered. Thus, the data generated by the process for the next h periods is given by:

$$\begin{aligned} \mathbf{w}_{T+j} &= \boldsymbol{\tau}_0^* + \boldsymbol{\tau}_1^*(T+j) + \boldsymbol{\Upsilon}^* \mathbf{w}_{T+j-1} + \boldsymbol{\nu}_{T+j} \\ &= \sum_{i=0}^{j-1} (\boldsymbol{\Upsilon}^*)^i \boldsymbol{\tau}^*(i) + \sum_{i=0}^{j-1} (\boldsymbol{\Upsilon}^*)^i \boldsymbol{\nu}_{T+j-i} + (\boldsymbol{\Upsilon}^*)^j \mathbf{w}_T. \end{aligned} \quad (13)$$

Then, the j -step ahead forecast error can be written as:

$$\begin{aligned} \hat{\boldsymbol{\nu}}_{T+j} &= \sum_{i=0}^{j-1} (\boldsymbol{\Upsilon}^*)^i \boldsymbol{\tau}^*(i) + \sum_{i=0}^{j-1} (\boldsymbol{\Upsilon}^*)^i \boldsymbol{\nu}_{T+j-i} + (\boldsymbol{\Upsilon}^*)^j \mathbf{w}_T - \sum_{i=0}^{j-1} \boldsymbol{\Upsilon}^i \boldsymbol{\tau}(i) - \boldsymbol{\Upsilon}^j \mathbf{w}_T \\ &= \left(\sum_{i=0}^{j-1} (\boldsymbol{\Upsilon}^*)^i \boldsymbol{\tau}^*(i) - \sum_{i=0}^{j-1} \boldsymbol{\Upsilon}^i \boldsymbol{\tau}(i) \right) + \sum_{i=0}^{j-1} (\boldsymbol{\Upsilon}^*)^i \boldsymbol{\nu}_{T+j-i} + \left((\boldsymbol{\Upsilon}^*)^j - \boldsymbol{\Upsilon}^j \right) \mathbf{w}_T. \end{aligned} \quad (14)$$

The three components of forecast error are due to the changed intercepts and slope parameters; error accumulation; and an interaction term occasioned by the change in the slope parameter which includes the initial condition.

The expectation of the j -step forecast error conditional on \mathbf{w}_T is:

$$\mathbb{E}[\hat{\nu}_{T+j} \mid \mathbf{w}_T] = \left(\sum_{i=0}^{j-1} (\mathbf{\Upsilon}^*)^i \boldsymbol{\tau}^*(i) - \sum_{i=0}^{j-1} \mathbf{\Upsilon}^i \boldsymbol{\tau}(i) \right) + \left((\mathbf{\Upsilon}^*)^j - \mathbf{\Upsilon}^j \right) \mathbf{w}_T \quad (15)$$

so that the conditional forecast error variance is:

$$\mathbb{V}[\hat{\nu}_{T+j} \mid \mathbf{w}_T] = \sum_{i=0}^{j-1} (\mathbf{\Upsilon}^*)^i \boldsymbol{\Omega} (\mathbf{\Upsilon}^*)^{i'}.$$

We now consider a number of special cases where only the impact of the deterministic components changes. With the assumption that $\mathbf{\Upsilon}^* = \mathbf{\Upsilon}$, we obtain (noting the dependence of $\boldsymbol{\tau}^*(i)$ on $T+j$):

$$\begin{aligned} \mathbb{E}[\hat{\nu}_{T+j} \mid \mathbf{w}_T] &= \sum_{i=0}^{j-1} \mathbf{\Upsilon}^i ([\boldsymbol{\tau}_0^* + \boldsymbol{\tau}_1^*(T+j-i)] - [\boldsymbol{\tau}_0 + \boldsymbol{\tau}_1(T+j-i)]) \\ &= \sum_{i=0}^{j-1} \mathbf{\Upsilon}^i [(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}) + \boldsymbol{\alpha}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_0^*) + \boldsymbol{\alpha}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_1^*)(T+j-i)]. \end{aligned} \quad (16)$$

The bias is increasing in j due to the first term in square brackets. The impacts of the second and third terms eventually level off because:

$$\lim_{i \rightarrow \infty} \mathbf{\Upsilon}^i = \mathbf{I}_n - \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{\alpha})^{-1} \boldsymbol{\beta}' = \mathbf{K}$$

and $\mathbf{K}\boldsymbol{\alpha} = \mathbf{0}$. When the linear trend is absent and the constant term can be restricted to the cointegrating space (ie. $\boldsymbol{\tau}_1 = \mathbf{0}$ and $\boldsymbol{\zeta}_0 = \mathbf{0}$, which implies $\boldsymbol{\lambda}_1 = \mathbf{0}$ and therefore $\boldsymbol{\mu}_1 = \boldsymbol{\gamma} = \mathbf{0}$) then only the second term appears, and the bias is $\mathcal{O}(1)$ in j . The formulation in (16) assumes that $\mathbf{\Upsilon}$, and therefore the cointegrating space, remains unaltered. Moreover, the coefficient on the linear trend alters but still lies in the cointegrating space. Otherwise, after the structural break, \mathbf{w}_t would be propelled by quadratic trends.

The conditional forecast error variance is:

$$\mathbb{V}[\hat{\nu}_{T+j} \mid \mathbf{w}_T] = \sum_{i=0}^{j-1} \mathbf{\Upsilon}^i \boldsymbol{\Omega} \mathbf{\Upsilon}^{i'} \quad (17)$$

which is $\mathcal{O}(j)$.

2.3 DVAR forecast errors

Now, consider forecasts from a simplified DVAR. Forecasts from the DVAR for $\Delta \mathbf{w}_t$ are defined by setting $\Delta \mathbf{w}_{T+j}$ equal to the population growth rate $\boldsymbol{\gamma}$:

$$\Delta \tilde{\mathbf{w}}_{T+j} = \boldsymbol{\gamma} \quad (18)$$

so that j -step ahead forecasts of the level of the process are obtained by integrating (18) from the initial condition \mathbf{w}_T :

$$\tilde{\mathbf{w}}_{T+j} = \tilde{\mathbf{w}}_{T+j-1} + \boldsymbol{\gamma} = \mathbf{w}_T + j\boldsymbol{\gamma} \text{ for } j = 1, \dots, h. \quad (19)$$

When Υ is unchanged over the forecast period, the expected value of the conditional j -step ahead forecast error is:

$$\mathbb{E} [\tilde{\nu}_{T+j} \mid \mathbf{w}_T] = \sum_{i=0}^{j-1} \Upsilon^i [\tau_0^* + \tau_1^* (T+j-i)] - j\gamma + (\Upsilon^j - \mathbf{I}_n) \mathbf{w}_T. \quad (20)$$

The occurrence of \mathbf{w}_T in (20) is awkward for comparisons with the VECM in (16). Thus, we average over \mathbf{w}_T to give the unconditional bias $\mathbb{E}_{\mathbf{w}_T} [\tilde{\nu}_{T+j}]$. Since $\Upsilon = \mathbf{I}_n + \alpha\beta'$, for $j > 0$:

$$\Upsilon^j = (\mathbf{I}_n + \alpha\beta')^j = \Upsilon^{j-1} (\mathbf{I}_n + \alpha\beta') = \Upsilon^{j-1} + \Upsilon^{j-1} \alpha\beta' = \dots = \mathbf{I}_n + \sum_{i=0}^{j-1} \Upsilon^i \alpha\beta', \quad (21)$$

so from (7) using:

$$(\Upsilon^j - \mathbf{I}_n) = \sum_{i=0}^{j-1} \Upsilon^i \alpha\beta' \doteq \mathbf{A}_j \alpha\beta' \quad (22)$$

we obtain:

$$\mathbb{E}_{\mathbf{w}_T} [(\Upsilon^j - \mathbf{I}_n) \mathbf{w}_T] = \mathbf{A}_j \alpha \mathbb{E}_{\mathbf{w}_T} [\beta' \mathbf{w}_T] = \mathbf{A}_j \alpha \mathbf{f}_T \quad (23)$$

where $\mathbf{f}_T = \mathbb{E}_{\mathbf{w}_T} [\beta' \mathbf{w}_T] = \mu_0^a + \beta' \gamma^a (T+1)$, say, where the values of μ_0^a and γ^a depend on the regime. Substituting from (23) into (20):

$$\mathbb{E}_{\mathbf{w}_T} [\tilde{\nu}_{T+j}] = \sum_{i=0}^{j-1} \Upsilon^i [\gamma^* - \alpha\mu_0^* - \alpha\mu_1^* (T+j-i)] - j\gamma + \mathbf{A}_j \alpha \mathbf{f}_T. \quad (24)$$

From (21), as $\Upsilon^i = \mathbf{I}_n + \mathbf{A}_i \alpha\beta'$:

$$\mathbf{A}_j = \sum_{k=0}^{j-1} \Upsilon^k = \sum_{k=0}^{j-1} (\mathbf{I}_n + \mathbf{A}_k \alpha\beta') = j\mathbf{I}_n + \left(\sum_{k=0}^{j-1} \mathbf{A}_k \right) \alpha\beta' = j\mathbf{I}_n + \mathbf{B}_j \alpha\beta'. \quad (25)$$

Thus from (24), since $\beta' \gamma = \mu_1$ and $\beta' \gamma^* = \mu_1^*$:

$$\begin{aligned} \mathbb{E} [\tilde{\nu}_{T+j}] &= \mathbf{A}_j \gamma^* - \mathbf{A}_j \alpha \mu_0^* - \mathbf{A}_j \alpha \beta' \gamma^* (T+j) + \sum_{i=1}^{j-1} i \Upsilon^i \alpha \beta' \gamma^* - j\gamma + \mathbf{A}_j \alpha \mathbf{f}_T \\ &= j(\gamma^* - \gamma) + \mathbf{A}_j \alpha \mathbf{f}_T - \mu_0^* - \beta' \gamma^* T + \left(\sum_{i=1}^{j-1} i \Upsilon^i - j\mathbf{A}_j + \mathbf{B}_j \right) \alpha \beta' \gamma^* \\ &= j(\gamma^* - \gamma) + \mathbf{A}_j \alpha (\mu_0^a - \mu_0^* - \beta' [\gamma^* - \gamma^a] (T+1)) + \mathbf{C}_j \alpha \beta' \gamma^* \end{aligned} \quad (26)$$

where $\mathbf{C}_j = (\mathbf{D}_j + \mathbf{B}_j - (j-1)\mathbf{A}_j)$ when $\mathbf{D}_j = \sum_{i=1}^{j-1} i \Upsilon^i$. However, $\mathbf{C}_j \alpha \beta' = \mathbf{0}$ as follows. Since $\Upsilon^j = \mathbf{I}_n + \mathbf{A}_j \alpha \beta'$ from (22), then:

$$j\mathbf{A}_j \alpha \beta' = j\Upsilon^j - j\mathbf{I}_n,$$

and so eliminating $j\mathbf{I}_n$ using (25):

$$(\mathbf{B}_j - j\mathbf{A}_j) \alpha \beta' = \mathbf{A}_j - j\Upsilon^j.$$

Also:

$$\mathbf{D}_j = \sum_{i=1}^j i \Upsilon^i - j\Upsilon^j = \sum_{i=1}^j \Upsilon^i - j\Upsilon^j + \left(\sum_{i=1}^{j-1} i \Upsilon^i \right) \Upsilon = \mathbf{A}_j \Upsilon - j\Upsilon^j + \mathbf{D}_j \Upsilon.$$

Since $\Upsilon = \mathbf{I}_n + \alpha\beta'$:

$$\mathbf{D}_j \alpha \beta' = j \Upsilon^j - \mathbf{A}_j - \mathbf{A}_j \alpha \beta'.$$

Combining these results:

$$\mathbf{C}_j \alpha \beta' = (\mathbf{D}_j + \mathbf{B}_j - (j-1) \mathbf{A}_j) \alpha \beta' = j \Upsilon^j - \mathbf{A}_j - \mathbf{A}_j \alpha \beta' + \mathbf{A}_j - j \Upsilon^j + \mathbf{A}_j \alpha \beta' = \mathbf{0}. \quad (27)$$

Thus:

$$\mathbb{E} [\tilde{\nu}_{T+j}] = j (\gamma^* - \gamma) + \mathbf{A}_j \alpha ([\mu_0^a - \mu_0^*] - \beta' [\gamma^* - \gamma^a] (T+1)). \quad (28)$$

In the same notation, the VECM results from (16) are:

$$\mathbb{E} [\hat{\nu}_{T+j}] = j (\gamma^* - \gamma) + \mathbf{A}_j \alpha ([\mu_0 - \mu_0^*] - \beta' [\gamma^* - \gamma] (T+1)). \quad (29)$$

Thus, (29) and (28) coincide when $\mu_0^a = \mu_0$, and $\gamma^a = \gamma$ as will occur if either there is no regime shift, or the shift occurs after the start of the forecast period.

2.4 Forecast biases under deterministic shifts

We now consider a number of interesting special cases of (28) and (29) which highlight the different behaviour of the DVAR and VECM under regime changes. Note that when $\gamma^* = \gamma$, then $\beta' \gamma^* = \beta' \gamma$, whereas $\gamma^* \neq \gamma$ does not necessarily entail that $\beta' \gamma^* \neq \beta' \gamma$. If we view (τ_0, τ_1) as the primary parameters, then it is informative to map changes in these parameters, via the orthogonal decomposition into $(\zeta_0, \lambda_0, \lambda_1)$, to the parameterization in terms of (γ, μ_0, μ_1) that underpins (28) and (29). From section 2.1 we can summarise the interdependencies as: $\gamma(\zeta_0, \lambda_1)$, $\mu_0(\zeta_0, \lambda_0, \lambda_1)$, $\mu_1(\lambda_1)$.

Case I $\tau_0^* = \tau_0$, $\tau_1^* = \tau_1$. Here, there is no structural change, $\mu_0^a = \mu_0$ and $\gamma^a = \gamma$ and so:

$$\mathbb{E} [\hat{\nu}_{T+j}] = \mathbb{E} [\tilde{\nu}_{T+j}] = \mathbf{0}. \quad (30)$$

Thus, the forecast error biases in the DVAR and VECM coincide when there is no regime change, even when the DVAR omits an ECM which includes a non-zero trend.

Case II $\tau_0^* \neq \tau_0$, $\tau_1^* = \tau_1$, but $\zeta_0^* = \zeta_0$. Then $\gamma^* = \gamma$; $\mu_0^* \neq \mu_0$.

$$\mathbb{E} [\hat{\nu}_{T+j}] = \mathbf{A}_j \alpha (\mu_0 - \mu_0^*) \quad (31)$$

$$\mathbb{E} [\tilde{\nu}_{T+j}] = \mathbf{A}_j \alpha (\mu_0^a - \mu_0^*). \quad (32)$$

The biases are equal if $\mu_0^a = \mu_0$; i.e. the shock is after the initial condition. However, $\mathbb{E} [\tilde{\nu}_{T+j}] = \mathbf{0}$ when $\mu_0^a = \mu_0^*$, and hence the DVAR wins uniformly if the shock has occurred prior to the commencement of forecasting. In this example the component of the constant term orthogonal to α is unchanged so that the growth rate is unaffected.

Case III $\tau_0^* \neq \tau_0$, $\tau_1^* = \tau_1$ (as in Case II), but now $\lambda_0^* = \lambda_0$ which implies $\zeta_0^* \neq \zeta_0$ and therefore $\mu_0^* \neq \mu_0$ and $\gamma^* \neq \gamma$. However, $\beta' \gamma^* = \beta' \gamma$ holds so that:

$$\mathbb{E} [\hat{\nu}_{T+j}] = j (\gamma^* - \gamma) + \mathbf{A}_j \alpha (\mu_0 - \mu_0^*) \quad (33)$$

$$\mathbb{E} [\tilde{\nu}_{T+j}] = j (\gamma^* - \gamma) + \mathbf{A}_j \alpha (\mu_0^a - \mu_0^*). \quad (34)$$

since μ_1 depends only on τ_1 . Consequently, the errors coincide when $\mu_0^a = \mu_0$, but differ when $\mu_0^a = \mu_0^*$, though it is unclear whether the terms augment or attenuate each other.

Case IV $\tau_0^* = \tau_0$, $\tau_1^* \neq \tau_1$. All of μ_0 , μ_1 and γ change. If $\beta' \gamma^* \neq \beta' \gamma$ then we have (28) and (29), and otherwise the biases of Case III.

Alternatively, the difference in the forecast bias between the DVAR and the VECM is given by:

$$\mathbf{d}_j = \mathbb{E}[\tilde{\nu}_{T+j}] - \mathbb{E}[\hat{\nu}_{T+j}] = \mathbf{A}_j \boldsymbol{\alpha} ([\boldsymbol{\mu}_0 - \boldsymbol{\mu}_0^a] + \boldsymbol{\beta}' [\boldsymbol{\gamma} - \boldsymbol{\gamma}^a] (T+1)). \quad (35)$$

This is zero when $\boldsymbol{\mu}_0^a = \boldsymbol{\mu}_0$ and $\boldsymbol{\gamma}^a = \boldsymbol{\gamma}$, but otherwise enhances or attenuates the existing biases, generally inducing a smaller forecast error bias in the DVAR when there are unmodelled regime shifts.

An alternative representation is in terms of the changes between successive forecast errors. For the VECM, from (28) and that equation lagged:

$$\mathbb{E}[\tilde{\nu}_{T+j}] - \mathbb{E}[\tilde{\nu}_{T+j-1}] = (\boldsymbol{\gamma}^* - \boldsymbol{\gamma}) + \boldsymbol{\Upsilon}^{j-1} \boldsymbol{\alpha} ([\boldsymbol{\mu}_0 - \boldsymbol{\mu}_0^*] + \boldsymbol{\beta}' [\boldsymbol{\gamma} - \boldsymbol{\gamma}^*] (T+1)) \quad (36)$$

since $\mathbf{A}_j - \mathbf{A}_{j-1} = \boldsymbol{\Upsilon}^{j-1}$. The first term is constant, and hence can be removed by an intercept correction once the regime shift has occurred; the second tends to zero as j increases since:

$$\boldsymbol{\Upsilon}^{j-1} \boldsymbol{\alpha} \rightarrow \mathbf{K} \boldsymbol{\alpha} = \mathbf{0}.$$

The results for the DVAR are similar:

$$\mathbb{E}[\tilde{\nu}_{T+j}] - \mathbb{E}[\tilde{\nu}_{T+j-1}] = (\boldsymbol{\gamma}^* - \boldsymbol{\gamma}) + \boldsymbol{\Upsilon}^{j-1} \boldsymbol{\alpha} ([\boldsymbol{\mu}_0^a - \boldsymbol{\mu}_0^*] + \boldsymbol{\beta}' [\boldsymbol{\gamma}^a - \boldsymbol{\gamma}^*] (T+1)). \quad (37)$$

When $\boldsymbol{\mu}_0^a = \boldsymbol{\mu}_0^*$ and $\boldsymbol{\beta}' \boldsymbol{\gamma}^a = \boldsymbol{\beta}' \boldsymbol{\gamma}^*$, then only a constant error ensues; when $\boldsymbol{\mu}_0^a = \boldsymbol{\mu}_0$ and $\boldsymbol{\beta}' \boldsymbol{\gamma}^a = \boldsymbol{\beta}' \boldsymbol{\gamma}$, the outcome coincides with the VECM; and otherwise, it lies in between.

3 Intercept corrections

We can show that if there is a one-off change in the value of the deterministic parameters $\boldsymbol{\tau}_0$ in the DGP, which has occurred prior to the period on which the forecasts are conditioned, then the optimal (in the sense of yielding unbiased forecasts) intercept correction for the VECM is simply to add in the period T residual each step ahead. Below we denote forecasts generated by this method by $\dot{\mathbf{w}}_{T+h}$. We also consider some other general strategies for intercept correcting, which are applied in the empirical work. In practice of course, the circumstances under which the $\dot{\mathbf{w}}_{T+h}$ strategy is optimal are unlikely to hold exactly: the $\boldsymbol{\tau}_1$ parameters may shift as well; $\boldsymbol{\tau}_0$ may change a number of times over the forecast period; the slope parameters may also alter, etc. Thus it may be of interest empirically to compare the performance of other adjustment schemes, and in this section we consider what can be deduced analytically about their properties.

We may also wish to base the adjustment on some average of recent errors rather than simply the observed model error at the forecast origin.

Below, we assume that the period T residual embodies the change in the process, and the discussion is in terms of the VECM even though there may also be scope for adjustments to the DVAR. Assuming that $\boldsymbol{\Upsilon}$ does not change, the period T residual is given by:

$$\hat{\nu}_T = \mathbf{w}_T - \hat{\mathbf{w}}_T = (\boldsymbol{\tau}_0^* - \boldsymbol{\tau}_0) + (\boldsymbol{\tau}_1^* - \boldsymbol{\tau}_1) T + \boldsymbol{\nu}_T. \quad (38)$$

We now consider a number of options for forecasting h -steps ahead. The intercept correction (IC) can be held constant over the forecast period, so that the period T error is added in at each step ahead. This is perhaps the most commonly-used form of intercept correction, where the adjustment over the future is held constant at an average of the most recent errors (in our example, just the period T error). This amounts to solving:

$$\dot{\mathbf{w}}_{T+h} = \boldsymbol{\tau}_0 + \boldsymbol{\tau}_1 (T+h) + \boldsymbol{\Upsilon} \dot{\mathbf{w}}_{T+h-1} + \hat{\nu}_T \quad (39)$$

where $\dot{\mathbf{w}}_T = \mathbf{w}_T$, so that:

$$\dot{\mathbf{w}}_{T+h} = \hat{\mathbf{w}}_{T+h} + \sum_{i=0}^{h-1} \Upsilon^i \hat{\nu}_T = \hat{\mathbf{w}}_{T+h} + \mathbf{A}_h \hat{\nu}_T. \quad (40)$$

Secondly, only adjust the 1-step forecast:

$$\begin{aligned} \overrightarrow{\mathbf{w}}_{T+1} &= \dot{\mathbf{w}}_{T+1} \\ \vdots &\quad \quad \quad \vdots \\ \overrightarrow{\mathbf{w}}_{T+h} &= \boldsymbol{\tau}_0 + \boldsymbol{\tau}_1 (T+h) + \Upsilon \overrightarrow{\mathbf{w}}_{T+h-1} \end{aligned} \quad (41)$$

which implies that:

$$\overrightarrow{\mathbf{w}}_{T+h} = \hat{\mathbf{w}}_{T+h} + \Upsilon^{h-1} \hat{\nu}_T. \quad (42)$$

Thirdly, one may adjust the h -step forecast by the full amount of the period T error:

$$\overleftrightarrow{\mathbf{w}}_{T+h} = \hat{\mathbf{w}}_{T+h} + \hat{\nu}_T. \quad (43)$$

Finally, tailing-off the adjustment induces:

$$\overline{\mathbf{w}}_{T+h} = \boldsymbol{\tau} + \Upsilon \overline{\mathbf{w}}_{T+h-1} + \mathbf{H}^{h-1} \hat{\nu}_T \quad (44)$$

so that:

$$\overline{\mathbf{w}}_{T+h} = \hat{\mathbf{w}}_{T+h} + \sum_{i=0}^{h-1} \Upsilon^i \mathbf{H}^{h-1-i} \hat{\nu}_T$$

where \mathbf{H} may be diagonal, say, with typical element $|h_{ii}| < 1$.

3.1 Biases

It is straightforward to derive expressions for the biases associated with the above forecasts, since in each case the forecasts are written as the conditional expectation ($\hat{\mathbf{w}}_{T+h}$) plus another term, and the conditional expectation forecast biases are given by (16) or (29). For the first strategy, from (38):

$$\mathbb{E} [\hat{\nu}_T] = (\boldsymbol{\tau}_0^* - \boldsymbol{\tau}_0) + (\boldsymbol{\tau}_1^* - \boldsymbol{\tau}_1) T,$$

and noting that the conditional and unconditional (over \mathbf{w}_T) expectations coincide here, then in the $(\boldsymbol{\tau}_0, \boldsymbol{\tau}_1)$ notation:

$$\mathbb{E} [\hat{\nu}_{T+h} \mid \mathbf{w}_T] = \mathbb{E} [\hat{\nu}_{T+h} - \mathbf{A}_h \hat{\nu}_T] = [h\mathbf{A}_h - \mathbf{D}_h] (\boldsymbol{\tau}_1^* - \boldsymbol{\tau}_1). \quad (45)$$

As claimed above, the constant-adjustment strategy yields unbiased forecasts when $\boldsymbol{\tau}_1^* = \boldsymbol{\tau}_1$.

In terms of the notation of section 2.2:

$$\mathbf{A}_h \hat{\nu}_T = \mathbf{A}_h (\gamma^* - \gamma) - \mathbf{A}_h \boldsymbol{\alpha} (\mu_0^* - \mu_0) - \mathbf{A}_h \boldsymbol{\alpha} \beta' (\gamma^* - \gamma) T,$$

and taking $\mathbb{E} [\hat{\nu}_{T+h}]$ from (29) we obtain:

$$\mathbb{E} [\hat{\nu}_{T+h} \mid \mathbf{w}_T] = (h - \mathbf{A}_h - \mathbf{A}_h \boldsymbol{\alpha} \beta') (\gamma^* - \gamma) = -(\mathbf{A}_h + \mathbf{B}_h) \boldsymbol{\alpha} \beta' (\gamma^* - \gamma),$$

which is zero when the time trend is absent since $\beta' (\gamma^* - \gamma) = \mu_1^* - \mu_1 = 0$.

Expressions for the biases resulting from the other adjustment schemes are:

$$\mathbb{E}[\vec{\nu}_{T+h} | \mathbf{w}_T] = \mathbb{E}[\hat{\nu}_{T+h} - (\mathbf{A}_h - \mathbf{A}_{h-1}) \hat{\nu}_T]$$

so:

$$\mathbb{E}[\vec{\nu}_{T+h} | \mathbf{w}_T] = h(\gamma^* - \gamma) - \mathbf{A}_{h-1} [\alpha(\mu_0^* - \mu_0) + \alpha\beta'(\gamma^* - \gamma)(T+1)] - (\mathbf{A}_h - \mathbf{A}_{h-1}) [\mathbf{I}_n + \alpha\beta'](\gamma^* - \gamma) \quad (46)$$

and:

$$\mathbb{E}[\overleftarrow{\nu}_{T+h} | \mathbf{w}_T] = h(\gamma^* - \gamma) - (\mathbf{A}_h - \mathbf{I}_n) [\alpha(\mu_0^* - \mu_0) + \alpha\beta'(\gamma^* - \gamma)(T+1)] - [\mathbf{I}_n + \alpha\beta'](\gamma^* - \gamma). \quad (47)$$

Finally:

$$\begin{aligned} \mathbb{E}[\bar{\nu}_{T+h} | \mathbf{w}_T] &= h(\gamma^* - \gamma) - \left(\sum_{i=0}^{h-1} \Upsilon^i (\mathbf{I}_n - \mathbf{H}^{h-1-i}) \right) [\alpha(\mu_0^* - \mu_0) + \alpha\beta'(\gamma^* - \gamma)(T+1)] \\ &\quad - \sum_{i=0}^{h-1} \Upsilon^i \mathbf{H}^{h-1-i} [\mathbf{I}_n + \alpha\beta'](\gamma^* - \gamma). \end{aligned} \quad (48)$$

When the time trend is absent, the biases with ICs become:

$$\begin{aligned} \mathbb{E}[\vec{\nu}_{T+h} | \mathbf{w}_T] &= h(\gamma^* - \gamma) - \mathbf{A}_{h-1} [\alpha(\mu_0^* - \mu_0)] - (\mathbf{A}_h - \mathbf{A}_{h-1})(\gamma^* - \gamma) \\ &= (h - \mathbf{A}_h + \mathbf{A}_{h-1})(\gamma^* - \gamma) - \mathbf{A}_{h-1} [\alpha(\mu_0^* - \mu_0)]. \end{aligned} \quad (49)$$

$$\begin{aligned} \mathbb{E}[\overleftarrow{\nu}_{T+h} | \mathbf{w}_T] &= h(\gamma^* - \gamma) - (\gamma^* - \gamma) - (\mathbf{A}_h - \mathbf{I}_n) [\alpha(\mu_0^* - \mu_0)] \\ &= (h-1)(\gamma^* - \gamma) - \alpha \sum_{i=1}^{h-1} [\mathbf{I} + \beta' \alpha]^i (\mu_0^* - \mu_0) \end{aligned} \quad (50)$$

$$\begin{aligned} \mathbb{E}[\bar{\nu}_{T+h} | \mathbf{w}_T] &= h(\gamma^* - \gamma) - \left(\sum_{i=0}^{h-1} \Upsilon^i (\mathbf{I}_n - \mathbf{H}^{h-1-i}) \right) [\alpha(\mu_0^* - \mu_0)] \\ &\quad - \sum_{i=0}^{h-1} \Upsilon^i \mathbf{H}^{h-1-i} (\gamma^* - \gamma) \end{aligned} \quad (51)$$

Thus, when $\gamma^* = \gamma$ but $\mu_0^* \neq \mu_0$, just adjusting the h -step forecast (that is, $\overleftarrow{\mathbf{w}}_{T+h}$, in (50)) will yield a larger bias than adjusting only the 1-step forecast (see (49)). Even if the process remains unchanged there is no penalty in terms of bias from intercept correcting.

3.2 Variances

The penalty to intercept correcting when the process is unchanged is in terms of increased uncertainty. The conditional forecast error variances for strategies (40) to (44) are given by (52) to (55) below, and all exceed the conditional expectation forecast error variance $\mathbb{V}[\hat{\nu}_{T+h} | \mathbf{w}_T]$ in (17) by a positive semi-definite matrix:

$$\mathbb{V}[\dot{\nu}_{T+h} | \mathbf{w}_T] = 2\mathbb{V}[\hat{\nu}_{T+h} | \mathbf{w}_T] + \sum_{j=0}^{h-1} \sum_{i=0}^{h-1} \Upsilon^j \Omega \Upsilon^{i'} \quad j \neq i \quad (52)$$

$$\mathbb{V}[\vec{\nu}_{T+h} | \mathbf{w}_T] = \mathbb{V}[\hat{\nu}_{T+h} | \mathbf{w}_T] + \Upsilon^{h-1} \Omega \Upsilon^{h-1'} \quad (53)$$

$$V \left[\overleftrightarrow{\nu}_{T+h} \mid \mathbf{w}_T \right] = V \left[\hat{\nu}_{T+h} \mid \mathbf{w}_T \right] + \Omega \quad (54)$$

$$V \left[\bar{\nu}_{T+h} \mid \mathbf{w}_T \right] = V \left[\hat{\nu}_{T+h} \mid \mathbf{w}_T \right] + \sum_{j=0}^{h-1} \sum_{i=0}^{h-1} \mathbf{r}^i \left(\mathbf{I} - \mathbf{H}^{h-1-i} \right) \Omega \left(\mathbf{I} - \mathbf{H}^{h-1-i'} \right) \mathbf{r}^{i'}. \quad (55)$$

For example, the error variance more than doubles for the constant-adjustment strategy. The problem is apparent from (38), since the intercept correction comprises terms reflecting the change in the intercept and trend parameter *plus* the full value of the period T disturbance, which has an (unconditional) variance of Ω . A more precise estimate of the change-in-parameter component could be obtained by averaging a number of recent errors, provided the break occurred sufficiently far back.

Nevertheless, summing the (squared) bias and variance components, for a sufficiently large change in τ_0 , holding the adjustment constant over the forecast period will result in the smallest MSFE (since the bias components of the other adjustment schemes can be made arbitrarily large).

4 Forecast evaluation and multi-step forecasts

Clements and Hendry (1993) show that forecast comparisons based on MSFE may depend upon which isomorphic representation of the system is selected for the assessment. For example, one method or model may appear to predict the levels of the series more accurately, another the changes in the variables. Potential ranking reversals suggest caution in interpreting the results of MSFE comparisons. In our empirical work we check that our findings are not simply an artefact of one particular chosen transformation of the data by reporting results for both levels and differences.

Alternatively, we could employ invariant criteria, such as the generalized forecast error second moment (GFESM) or its determinant. This is formed by stacking the (vectors of) forecast errors from all previous step ahead forecasts. The 1-step forecast errors determine the complete ranking for the GFESM when there is no parameter uncertainty and the model is correctly specified. However, the choice between mis-specified models may depend on the forecast horizon, so that, independent of the measure of forecast accuracy, one model may dominate at certain horizons, and another at other horizons. Thus 1-step forecast performance may not be a reliable guide to the outcome of comparisons for h -steps ahead (see, for example, Baillie, 1993; Fama and French, 1988): for this reason we report the results of comparisons for multi-step forecasts.

Clements and Hendry (1995a) show that, in the absence of structural change or regime shifts, MSFE evaluations of the ability of models to predict differences of the variables may have low power in identifying models which incorrectly impose too many unit roots, i.e. VARs in differences when there is cointegration. We shall check whether all the models have a similar ability to predict the differences of the data in the presence of structural breaks.

5 Modelling wages and prices in the UK

Mizon (1995) analyzes the relationships between the following three variables for the UK, over the period 1965:1 to 1993:1; e_t , the natural logarithm of earnings per man-hour (loosely referred to as ‘wages’), the log of the retail price index, r_t , and the log of the unemployment rate u_t . Precise definitions and sources are given in Mizon (1995) or Clements and Mizon (1991). The latter used a dataset that also included average hours worked and productivity per man-hour to illustrate the ‘encompassing the VAR’ approach of Hendry and Mizon (1993), while Mizon (1995) uses an extended sample on the three variables.

Briefly, Mizon obtained a simultaneous model of the determination of wages, prices and employment, in which there is a single cointegrating vector that enters only the wage and price equations (see his Table 10). If the full-sample estimates of the cointegrating vector are retained, with the model otherwise being estimated up to 1979:2, and used to forecast (the 55 observations) 1979:3–1993:1, the model's 1-step ahead forecast performance is satisfactory, as is that of a DVAR estimated, and used to forecast, over the same period. However, if the cointegrating vector is also determined from the sub-sample estimation period (as would necessarily be the case in an *ex ante* forecasting exercise), the model fails to provide reliable forecasts.

This finding is not peculiar to the model specification found on the full-sample. A 'general-to-simple' model selection strategy applied to the sub-sample led to a model which differed somewhat from the full-sample specification (see Table 1 for the parameter estimates), but nevertheless possessed the feature that the 'error correction' terms caused the 1-step forecasts to go awry.⁴

Table 2 gives the first two moments of the 1-step forecast errors and two forecast test statistics (see Doornik and Hendry, 1994, for details) for a (third-order) DVAR, the model given in Mizon (1995) with the cointegrating vector estimated over the full sample, (GM_{FS}), the same specification with the cointegrating vector estimated on data up to 1979:2 (GM_{SS}), and a model specified and estimated on the sub-sample alone, VECM. Finally, VECM* is VECM but with the equilibrium-correction terms omitted.

From the first part of the table, it is apparent that the 1-step forecasts from the GM_{SS} and VECM models for the quarter on quarter growth in earnings are on average 0.8% points too high. This is approximately twice the size of the bias from the other three models (which either neglect the long-run information or base it on the full sample). By way of comparison, the average growth in earnings over the forecast period (1980:1–1993:1) was 2% a quarter. GM_{SS} and VECM also record larger biases in forecasting Δr . The standard deviations of the forecast errors are appreciably higher for both Δe and Δr for the GM_{SS} and VECM models. The forecast tests in the third part of the table depict the resounding rejection of the GM_{SS} and VECM models alluded to above. A comparison of the results for VECM with those for VECM* highlights the problems caused by the equilibrium-correction terms.

The tests have F -distributions with degrees of freedom as specified in the table, the quantities in '[]' following the test statistic values are p -values. The first test headed η_1 is an index of numerical parameter constancy, and ignores parameter uncertainty and serial correlation between the forecast errors, the second (headed η_2) allows for parameter uncertainty (see Doornik and Hendry, 1994, p. 197, for full details).

Figure 1 depicts time series plots of the 'equilibrium errors' for the full sample where the coefficients of the cointegrating vector have been estimated both from the full sample and from the sub-sample. The sub-sample error has a sharp downward trend in the 1980s, no longer appears to be an $I(0)$ process, and by the 1990s is over 10% lower than at the beginning of the period.

6 An empirical study of the impact of intercept corrections

Of the models incorporating long-run information discussed in section 5, we will explore the behaviour of VECM when use is made of intercept corrections: GM_{FS} and GM_{SS} either result from, or are motivated by, knowledge of the full sample, which is ruled out for *ex ante* forecasting.

Three of the forms of intercept correction discussed in section 3 are implemented: a constant adjustment throughout the period, a one-off adjustment in the first period of the forecast, and an adjustment which is tailed off as we forecast further ahead (the matrix \mathbf{H} in (44) is a diagonal matrix with $\frac{1}{2}$ s on

⁴We use the sub-sample 1965:1–1979:4 rather than 1965:1–1979:2, which makes little difference to the results.

the leading diagonal). In fact the constant adjustment to unemployment led to a ballooning of MSFEs (the unemployment equation is solely in differences), so by ‘constant adjustment to all equations’ is to be understood constant adjustment to earnings and prices but tailed-off adjustment to unemployment.

Forecasts can be adjusted using the value of an equation’s error at the forecast origin, or an average of that and the previous $(p - 1)$ errors. More elaborate schemes might utilise information from other equations, for example, or conduct significance tests, but we shall consider adjustments based on the latest own-error [1] and the latest four errors [4]. We also consider the impact of only adjusting the equations with ‘error-correction’ mechanisms, that is, e and r .

The precise way in which the forecasts are obtained is as follows. The models are estimated only once on data up to 1979:4. Then, 1979:4 is taken as the first forecast origin (initial condition) from which we calculate forecasts for 1 through to 12 steps ahead. The exercise is repeated with 1980:1 as the forecast origin, and so on, subject to the constraint that we have data on the period being forecast (the sample ends in 1993:1). This gives fifty-three 1-step forecasts, where the last forecast is made in 1992:4 of 1993:1, fifty-two 2 step forecasts (a forecast made in 1979:4 of 1980:2 through to a forecast made in 1992:3 of 1993:1), and so on to forty-two 12 steps. The means and forecast errors for each length of forecast are calculated by averaging over the available number of forecasts, corresponding to averaging over the forecast origins.

The results are summarised in figures 2, 3, 4 and 5. In each figure, the x -axis denotes the forecast lead, and the y -axis the trace mean-square forecast error (TMSFE). The first two figures are TMSFEs for predicting the levels of the data, the second two for the differences. Consider figure 2. The figure conveys information for the VECM, for one-off adjustments to the VECM based on the latest residual (one-off [1]) and an average of the latest 4 (one-off [4]); and for constant adjustments to the VECM (const [1], and const [4]). In each case, the line is scaled by the TMSFE for the DVAR, to aid interpretation. Figure 2 is based on corrections to e and r only, and figure 3 on corrections to all equations (but recall the ‘constant’ adjustment to u is a ‘tailed’ adjustment). The fully tailed-off adjustment cases are not depicted.

The figure shows that the TMSFE of the VECM model almost always exceeds that of the DVAR, and is over 40% higher for 12-steps ahead.⁵ Constant adjustments appear to work best on average (across horizons) for longer horizons, and one-off adjustments result in a better performance at long horizons when applied to all equations (as in figure 3) rather than only the e and r equations (figure 2). Averaging residuals ([4] versus [1]) works best for constant adjustments, and for horizons up to 10 quarters ahead, the VECM model now outperforms the DVAR on the TMSFE measure (figure 3). For predicting the changes in the data, the one-off adjustments are less useful than the constant adjustments, underlining the danger of relying on a single transformation of the data (just levels, say) when forecast models/methods are compared in terms of MSFE (Clements and Hendry, 1993). The finding in Clements and Hendry (1995a) that evaluation in terms of differences might be expected to have low power to discriminate between models does not apply in the presence of structural breaks. The adjustments are not quite as successful in improving the forecast performance of the VECM model.

It is apparent that automatic adjustments bolster the forecast performance of the model incorporating long-run information. Indeed, at short horizons, the VECM model outperforms the DVAR by about 20% (due to its better fit combined with the intercept correction offsetting the bias), and only when the variance of the intercept correction becomes large does the DVAR win. This suggests the possible use of selection criteria or longer averages when correcting.

We now look at whether the improvements due to the intercept corrections result from reductions

⁵Perhaps surprisingly given the test statistic results reported in table 2, there is little to choose between the two at 1-step ahead.

in the bias or variance components of the MSFEs. Table 3 records the bias and forecast error standard deviations for the VECM model, the DVAR, and the constant adjustments (to e and r only), and the one-off adjustments (to all equations), for a selection of forecast lead times, for the levels of the variables.

The constant adjustments (const [1]) reduce the biases in forecasting e and r at all horizons, but at the cost of larger forecast error variances. Averaging four residuals to form the constant-adjustment intercept correction (const [4]) is generally less successful in reducing biases, but has a smaller inflationary effect on the variances, consistent with the analysis in section 3. The impact of the one-off adjustments on biases is largely short term, and there is a minimal impact on the variance component.

7 Conclusions

In the empirical illustration, based on modelling and forecasting wages, prices and unemployment, we find that vector autoregressions in the differences of the variables (which eschew long-run information) forecast well compared to a vector equilibrium-correction system (VECM) when the long-run relationships are subject to structural change. However, the VECM clearly provides a better description of the data within-sample, and the equilibrium-correction terms are significant at conventional significance levels.

In line with the analysis in section 3, in the empirical illustration we found that intercept corrections can improve the forecasts of econometric models when the long-run relationships appear to alter over the forecast period. In particular, the corrections resulted in significant reductions in forecast bias. The precise form of the adjustments affects the forecast bias and variance, which also depend on the forecast horizon, emphasising the need to consider multi-step forecasts.

References

- Baillie, R. T. (1993). Comment on ‘On the limitations of comparing mean squared forecast errors’, by M.P. Clements and D.F. Hendry. *Journal of Forecasting*, **12**, 639–641.
- Clements, M. P. and Hendry, D. F. (1993). On the limitations of comparing mean squared forecast errors. *Journal of Forecasting*, **12**, 617–637. With discussion.
- Clements, M. P. and Hendry, D. F. (1994). Towards a theory of economic forecasting. In Hargreaves, C. (ed.), *Non-stationary Time-series Analyses and Cointegration*, pp. 9–52. Oxford: Oxford University Press.
- Clements, M. P. and Hendry, D. F. (1995a). Forecasting in cointegrated systems. *Journal of Applied Econometrics*, **10**, 127–146.
- Clements, M. P. and Hendry, D. F. (1995b). Macro-economic forecasting and modelling. *Economic Journal*, **105**, 1001–1013.
- Clements, M. P. and Mizon, G. E. (1991). Empirical analysis of macroeconomic time series: VAR and structural models. *European Economic Review*, **35**, 887–932.
- Davidson, J. E. H., Hendry, D. F., Srba, F. and Yeo, S. (1978). Econometric modelling of the aggregate time-series relationship between consumers’ expenditure and income in the United Kingdom. *Economic Journal*, **88**, 661–692. Reprinted in Hendry D. F. (1993), *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.
- Doornik, J. A. and Hendry, D. F. (1994). *PcFiml 8: An Interactive Program for Modelling Econometric Systems*. London: International Thomson Publishing.

- Fama, E. F. and French, K. R. (1988). Permanent and temporary components of stock prices. *Journal of Political Economy*, **96**, 2, 246–273.
- Hendry, D. F. and Clements, M. P. (1994a). Can econometrics improve economic forecasting?. *Swiss Journal of Economics and Statistics*, **130**, 267–298.
- Hendry, D. F. and Clements, M. P. (1994b). 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 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.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, **12**, 231–254.
- Johansen, S. (1994). The role of the constant and linear terms in cointegration analysis of nonstationary variables. *Econometric Reviews*, **13**, 205–229.
- Klein, L. R. (1971). *An Essay on the Theory of Economic Prediction*. Chicago: Markham Publishing Company.
- Mizon, G. E. (1995). Progressive modelling of macroeconomic time series: the LSE methodology. In Hoover, K. D. (ed.), *Macroeconometrics: Developments, Tensions and Prospects*. Dordrecht: Kluwer Academic Press. Forthcoming.
- Theil, H. (1961). *Economic Forecasts and Policy* 2nd edn. Amsterdam: North-Holland Publishing Company.
- Turner, D. S. (1990). The role of judgement in macroeconomic forecasting. *Journal of Forecasting*, **9**, 315–345.
- Wallis, K. F. and Whitley, J. D. (1991). Sources of error in forecasts and expectations: U.K. economic models 1984–8. *Journal of Forecasting*, **10**, 231–253.

8 Appendix I: A dynamic DVAR

Here we derive the population parameters of a DVAR for $\Delta \mathbf{w}_t$ regressed on $\Delta \mathbf{w}_{t-1}$ and an intercept. First, from (10):

$$\alpha'_{\perp} \Delta \mathbf{w}_t = \alpha'_{\perp} \gamma + \alpha'_{\perp} \nu_t \quad (56)$$

is a white-noise innovation process, accounting for $n - r$ of the equations in the DVAR. Next:

$$\beta' \mathbf{w}_t - (\mu_0 + \mu_1 (t + 1)) = (\mathbf{I}_r + \beta' \alpha) [\beta' \mathbf{w}_{t-1} - (\mu_0 + \mu_1 t)] + \beta' \nu_t.$$

Also:

$$\Delta \beta' \mathbf{w}_t = \mu_1 + \beta' \alpha (\beta' \mathbf{w}_{t-1} - \mu_0 - \mu_1 t) + \beta' \nu_t$$

Let:

$$\beta' \mathbf{w}_t - \mu_0 - \mu_1 (t + 1) = \mathbf{x}_t$$

be the remaining r variables in the DVAR. Then:

$$\mathbf{x}_t = \mathbf{A} \mathbf{x}_{t-1} + \mathbf{u}_t$$

where:

$$\mathbf{A} = \mathbf{I}_r + \beta' \alpha$$

has all its eigenvalues inside the unit circle. So:

$$\begin{aligned}
 E [\Delta \mathbf{x}_t \Delta \mathbf{x}'_{t-1}] &= 2E [\mathbf{x}_t \mathbf{x}'_{t-1}] - E [\mathbf{x}_t \mathbf{x}'_{t-2}] - E [\mathbf{x}_{t-1} \mathbf{x}'_{t-1}] \\
 &= 2\mathbf{A}E [\mathbf{x}_{t-1} \mathbf{x}'_{t-1}] - \mathbf{A}^2 E [\mathbf{x}_{t-2} \mathbf{x}'_{t-2}] - E [\mathbf{x}_{t-1} \mathbf{x}'_{t-1}] \\
 &= (\mathbf{I}_r - 2\mathbf{A} + \mathbf{A}^2) \mathbf{M} = (\mathbf{I}_r - \mathbf{A})^2 \mathbf{M} = -(\beta' \alpha)^2 \mathbf{M}
 \end{aligned}$$

and:

$$\begin{aligned}
 E [\Delta \mathbf{x}_{t-1} \Delta \mathbf{x}'_{t-1}] &= 2E [\mathbf{x}_{t-1} \mathbf{x}'_{t-1}] - E [\mathbf{x}_{t-1} \mathbf{x}'_{t-2}] - E [\mathbf{x}_{t-2} \mathbf{x}'_{t-1}] \\
 &= (\mathbf{I}_r - \mathbf{A}) \mathbf{M} + \mathbf{M} (\mathbf{I}_r - \mathbf{A}') = -\beta' \alpha \mathbf{M} - \mathbf{M} \alpha' \beta.
 \end{aligned}$$

Finally:

$$E [\mathbf{x}_t \mathbf{x}'_t] = \mathbf{M} = \mathbf{A} \mathbf{M} \mathbf{A}' + \beta' \Omega \beta.$$

Noting that $\Delta \mathbf{x}_t = \beta' \Delta \mathbf{w}_t - \mu_1$ and letting:

$$\Delta \beta' \mathbf{w}_t = -\beta' \alpha \mu_1 + \mathbf{A} \Delta \beta' \mathbf{w}_{t-1} + \mathbf{e}_t \quad (57)$$

then:

$$\mathbf{A} = (\beta' \alpha)^2 \mathbf{M} (\beta' \alpha \mathbf{M} + \mathbf{M} \alpha' \beta)^{-1}.$$

Thus, we stack the $n - r$ equations from (56) with the r from (57) to yield:

$$\Delta \mathbf{w}_t = \begin{pmatrix} \alpha'_\perp \\ \beta' \end{pmatrix}^{-1} \left[\begin{pmatrix} \alpha'_\perp \gamma \\ -\beta' \alpha \mu_1 \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \mathbf{A} \beta' \end{pmatrix} \Delta \mathbf{w}_{t-1} + \begin{pmatrix} \alpha'_\perp \nu_t \\ \mathbf{e}_t \end{pmatrix} \right].$$

GM_{FS}: Mizon (1995), Table 10. Estimated by FIML. 1966:1--1993:1

$\Delta e =$	+0.3145 Δr (0.138)	+0.3291 Δe_2 (0.0742)	+0.3022 Δr_3 (0.0820)
	-0.03728 Δu_2 (0.01723)	-0.1269 ecm_1 (0.05549)	+0.0259 D793 (0.01212)
	+0.03758 D745 (0.004799)	-1.095 (0.4838)	
$\Delta r =$	+0.3737 Δr_1 (0.05695)	-0.03827 Δu_2 (0.0104)	+0.1815 ecm_1 (0.02358)
	+0.04739 D793 (0.00720)	+0.00675 D745 (0.00304)	+0.01119 Budget (0.00105)
	+1.585 (0.205)		
$\Delta u =$	+0.7554 Δe_1 (0.1652)	+0.7851 Δu_1 (0.0570)	+0.1218 Δu_2 (0.0620)
	-0.142 Δu_3 (0.0554)	+0.129 Expansion (0.0094)	-0.01232 (0.00535)
$\sigma_e = 0.0108$	$\sigma_r = 0.0071$	$\sigma_u = 0.0310$	

where $ecm = e - 0.88r - 0.017u - 0.008t$.

VECM model, estimated by FIML. 1966:1--1979:4

$\Delta e =$	+0.3538 Δe_2 (0.0862)	-0.1005 $ecm794_1$ (0.0498)	-0.8204 Constant (0.4161)
	+0.03685 D745 (0.00504)	+0.03616 D793 (0.01194)	
$\Delta r =$	+0.1906 Δe_1 (0.0600)	+0.2877 Δr_1 (0.0752)	-0.06187 Δu_2 (0.01602)
	+0.1994 $ecm794_1$ (0.0427)	+1.675 (0.357)	+0.04376 D793 (0.00860)
	+0.00872 Budget (0.00155)		
$\Delta u =$	+0.4669 Δe_1 (0.1492)	+0.7541 Δu_1 (0.0719)	+0.1264 Expansion (0.0122)
$\sigma_e = 0.0118$	$\sigma_r = 0.0086$	$\sigma_u = 0.0407$	

where $ecm794 = e - 0.76r + 0.076u - 0.013t$.

D793, D745, Budget, and Expansion are dummy variables, see Mizon (1995).

Table 1 Full-sample model and the sub-sample VECM model.

Mean of forecast errors: 'bias'. 1980:1 - 1993:1.

Model	Δe	Δr	Δu
DVAR	-0.0037	-0.0017	0.0010
GM_{FS}	-0.0040	-0.0005	0.0007
GM_{SS}	-0.0080	0.0039	0.0004
VECM	-0.0076	0.0026	-0.0050
VECM*	-0.0041	-0.0023	-0.0049
Average value of actuals	0.0212	0.0157	0.0186

Standard deviations of forecast errors. 1980:1 - 1993:1.

Model	Δe	Δr	Δu
DVAR	0.0120	0.0079	0.0281
GM_{FS}	0.0105	0.0069	0.0219
GM_{SS}	0.0147	0.0152	0.0219
VECM	0.0136	0.0139	0.0232
VECM*	0.0121	0.0077	0.0235
Average value of actuals	0.0137	0.0127	0.0581

Forecast Tests.

Model	$F[.,.]$	η_1	η_2
DVAR	159,44	0.82 [.81]	0.69 [.95]
GM_{FS}	159,49	0.69 [.96]	0.64 [.98]
GM_{SS}	159,49	2.62 [.0001]	1.96 [.004]
VECM	159,51	2.21 [.0007]	1.72 [.01]
VECM*	159,52	0.73 [.92]	0.70 [.95]

Table 2 1-step forecast performance of models .

1-step forecast errors

	Mean			s . dev			
Model	<i>e</i>	<i>r</i>	<i>u</i>	<i>e</i>	<i>r</i>	<i>u</i>	
DVAR	-0.004	-0.002	0.001	0.012	0.008	0.028	
VECM	-0.009	0.004	-0.005	0.014	0.012	0.023	
const [1]	-0.001	-	-0.005	0.012	0.010	0.023	
1-off [1]	-0.001	-	-	0.012	0.010	0.026	
const [4]	-0.002	0.001	-0.005	0.011	0.008	0.023	
1-off [4]	-0.002	0.001	0.001	0.011	0.008	0.025	

4-step forecast errors

	Mean			s . dev			
Model	<i>e</i>	<i>r</i>	<i>u</i>	<i>e</i>	<i>r</i>	<i>u</i>	
DVAR	-0.019	-0.013	-0.020	0.026	0.018	0.152	
VECM	-0.036	-	-0.065	0.038	0.034	0.141	
const [1]	-0.009	0.001	-0.048	0.044	0.043	0.133	
1-off [1]	-0.030	0.003	-0.043	0.033	0.033	0.129	
const [4]	-0.013	0.003	-0.051	0.029	0.027	0.136	
1-off [4]	-0.031	0.003	-0.041	0.032	0.032	0.141	

8-step forecast errors

	Mean			s . dev			
Model	<i>e</i>	<i>r</i>	<i>u</i>	<i>e</i>	<i>r</i>	<i>u</i>	
DVAR	-0.046	-0.034	-0.099	0.025	0.027	0.314	
VECM	-0.068	-0.024	-0.212	0.044	0.046	0.285	
const [1]	-0.025	0.002	-0.166	0.084	0.093	0.268	
1-off [1]	-0.065	-0.018	-0.182	0.042	0.047	0.268	
const [4]	-0.030	0.002	-0.172	0.049	0.064	0.268	
1-off [4]	-0.065	-0.019	-0.176	0.041	0.047	0.280	

12-step forecast errors

	Mean			s . dev			
Model	<i>e</i>	<i>r</i>	<i>u</i>	<i>e</i>	<i>r</i>	<i>u</i>	
DVAR	-0.070	-0.054	-0.236	0.028	0.033	0.409	
VECM	-0.089	-0.055	-0.414	0.037	0.050	0.381	
const [1]	-0.032	0.005	-0.341	0.122	0.152	0.346	
1-off [1]	-0.087	-0.049	-0.375	0.038	0.053	0.370	
const [4]	-0.041	0.001	-0.354	0.080	0.106	0.341	
1-off [4]	-0.088	-0.050	-0.375	0.036	0.053	0.381	

Table 3 Effect of intercept corrections on means and standard deviations of forecast errors. A ‘—’ denotes a number less than 10^{-3} in absolute value .

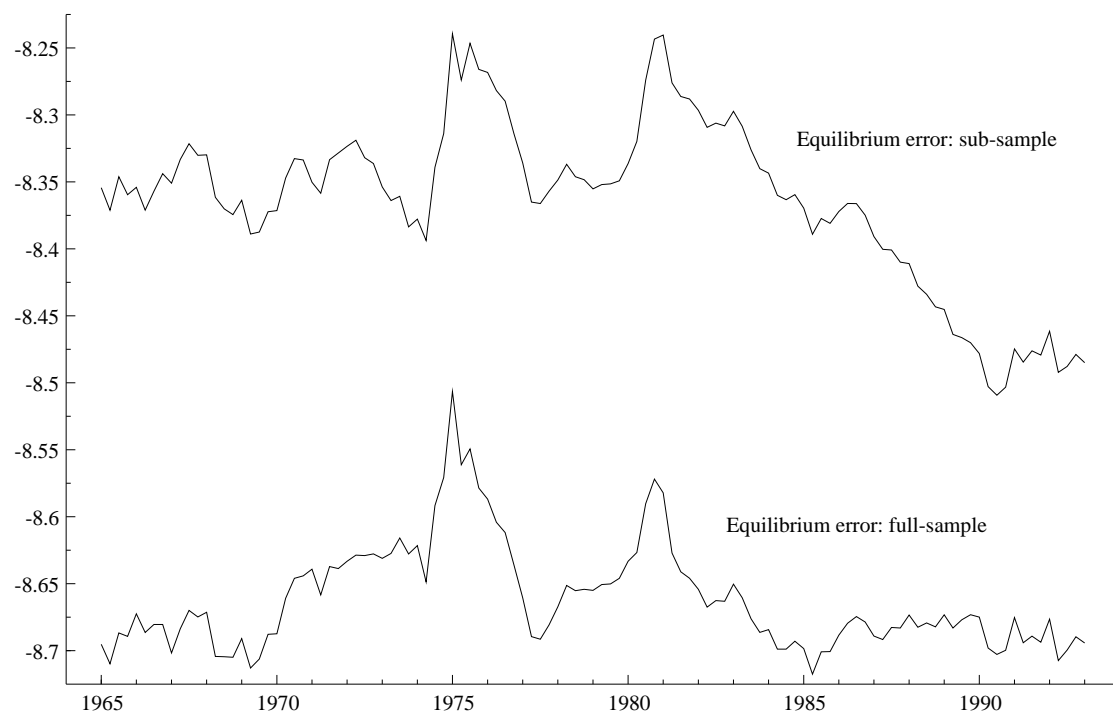


Figure 1 'Equilibrium errors': full and sub-sample estimates.

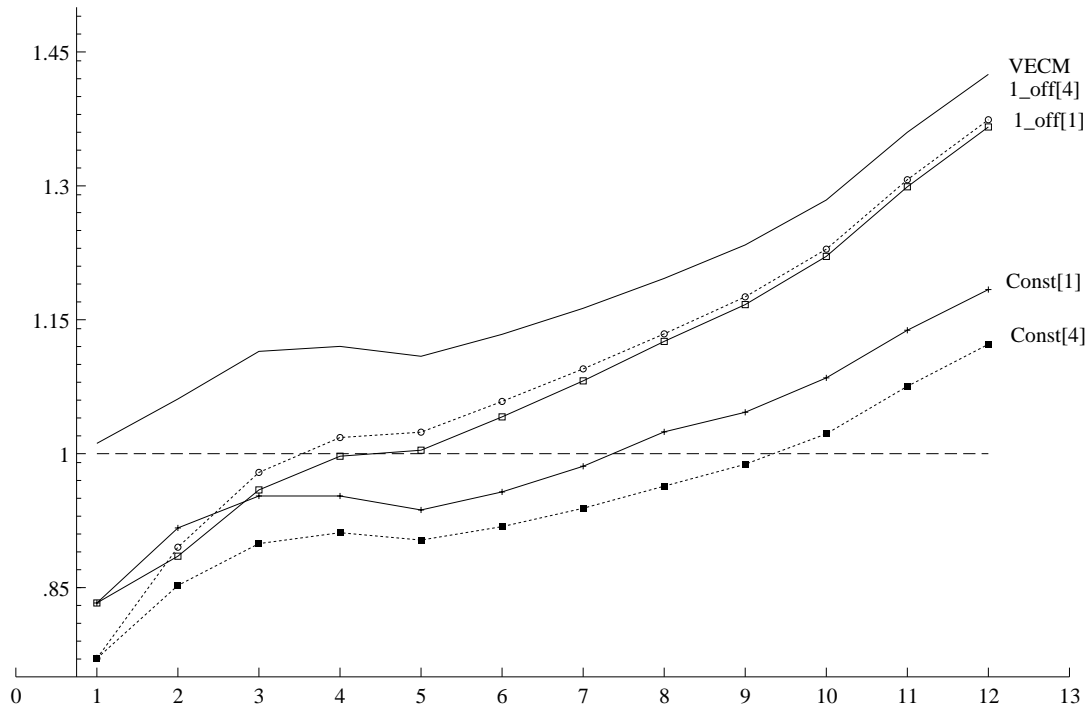


Figure 2 TMSFEs: levels of the variables. Adjustments to e and r equations only.

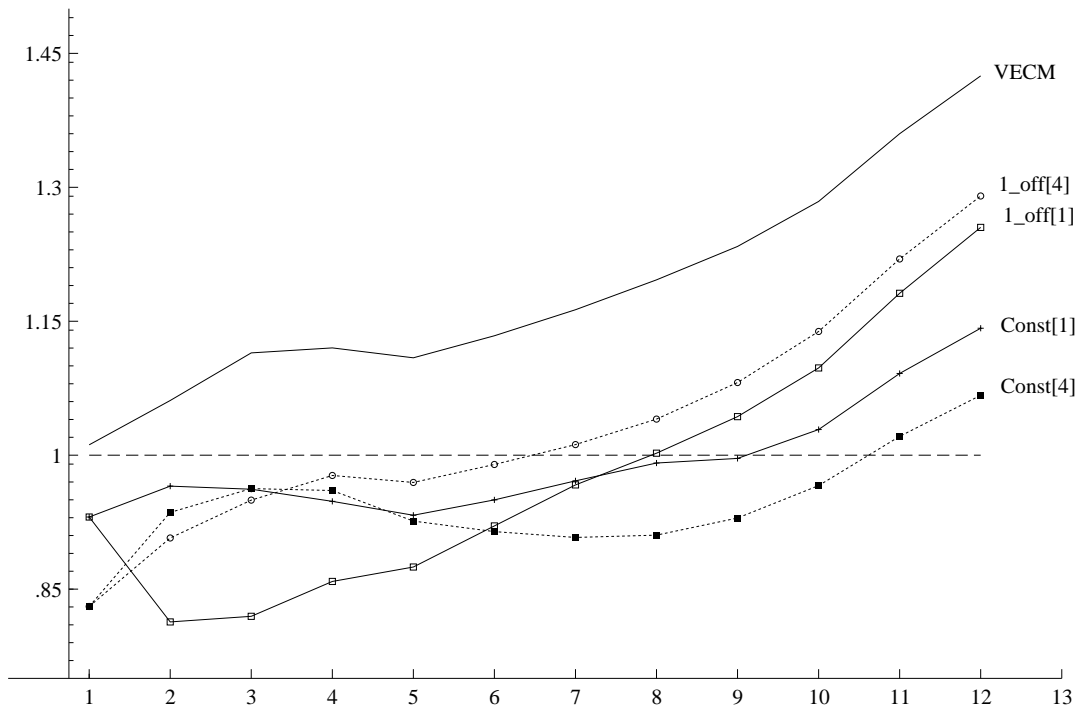


Figure 3 TMSFEs: levels of the variables. Adjustments to all equations.

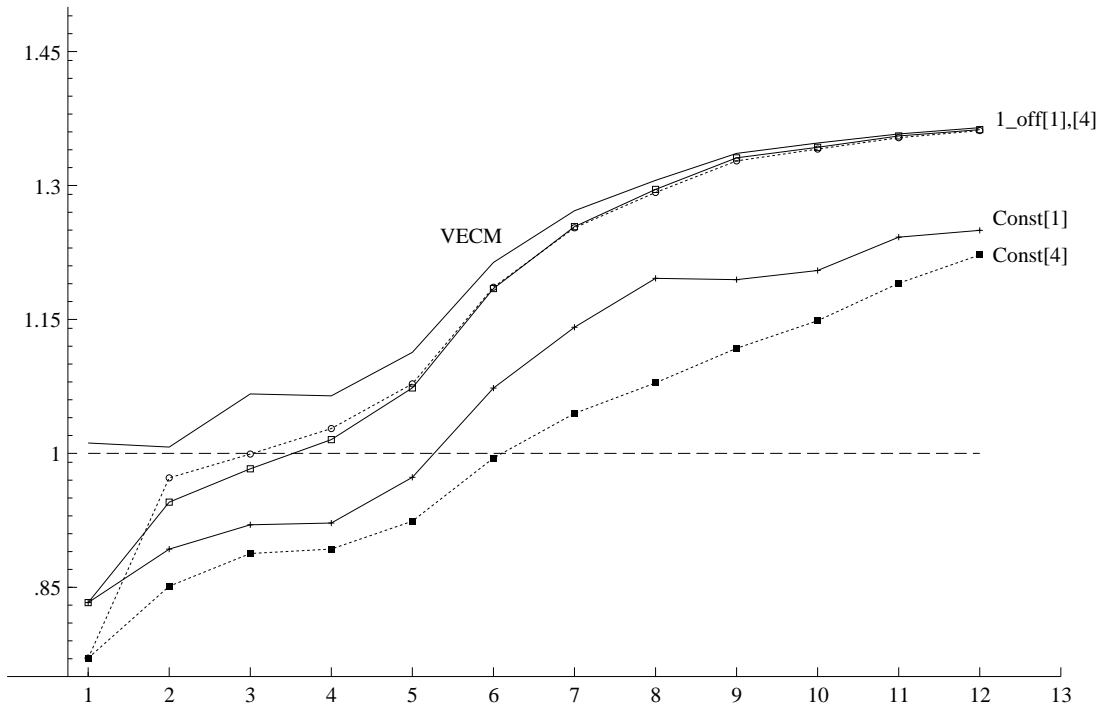


Figure 4 TMSFEs: differences of the variables. Adjustments to e and r equations only.

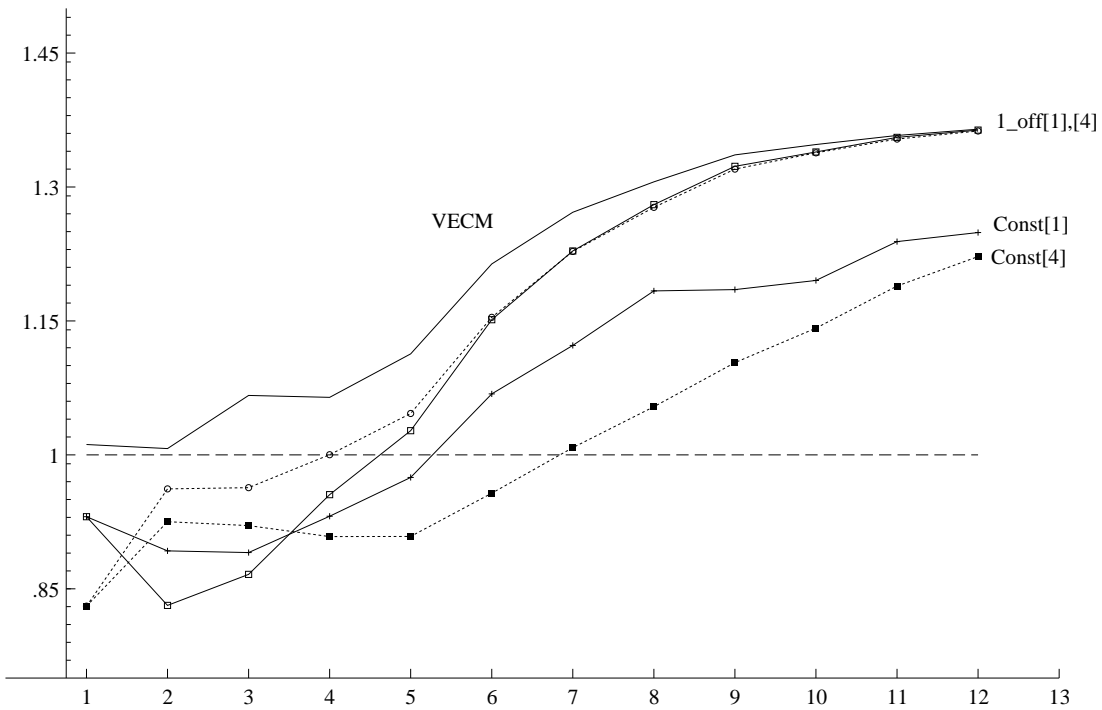


Figure 5 TMSFEs: differences of the variables. Adjustments to all equations.