ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

FORECASTING WITH EQUILIBRIUM-CORRECTION MODELS DURING STRUCTURAL BREAKS

Jennifer L. Castle, Nicholas W.P. Fawcett and David F. Hendry

Number 408
October 2008

Manor Road Building, Oxford OX1 3UQ

Forecasting with Equilibrium-correction Models during Structural Breaks

Jennifer L. Castle, Nicholas W. P. Fawcett, and David F. Hendry*
Economics Department, Oxford University

May 29, 2008

Abstract

When breaks occur, equilibrium-correction models (EqCMs) based on cointegration face forecasting problems. We investigate approaches to alleviate forecast failure following a location shift, including updating, intercept corrections, differencing, and estimating the future impact of an 'internal' break during its progress. Although updating can lead to a loss of cointegration when an EqCM suffers an equilibrium-mean shift, we show that updating can help when collinearites are changed by an 'external' break and the EqCM itself remains constant. Both mechanistic corrections help compared to just retaining a pre-break estimated model, but an estimated model of the break process could outperform. Throughout, we apply the approaches to the much-studied example of EqCMs for UK M1, and compare with updating a learning function as the break evolves.

JEL classifications: C1, C53.

1 Introduction

Building on the earlier tradition of 'error correction' in Phillips (1954), Sargan (1964), and Davidson, Hendry, Srba and Yeo (1978), the cointegration revolution sparked by Granger (1981) and Engle and Granger (1987) has greatly improved the econometric modeling of integrated time series: see e.g., Hendry (2004), Hendry and Juselius (2000, 2001) for recent surveys. By imposing convergence to equilibrium trajectories from long-run economic analysis, based on the isomorphism between cointegration and equilibrium-correction mechanisms (EqCMs), many new empirical economic insights have been gained.

Unfortunately, that very strength of cointegration when modeling has proved to be its Achilles heel in forecasting, namely a susceptibility to systematic forecast failure: see Clements and Hendry (1998, 1999). The forms of structural breaks that are pernicious for forecasting with cointegrated systems are location shifts, namely, breaks which change the underlying equilibrium means of the cointegration relationships. When such changes are not modeled, a cointegrated system converges back to its pre-break equilibrium–irrespective of where the post-break data are located–inducing systematic forecast failure: see Clements and Hendry (2002, 2006). Shifts in other parameters of a cointegrated system–when equilibrium means and growth rates are unchanged–are much less damaging to forecasts, and indeed may even be difficult to detect *ex post*: see Hendry (2000). *Ex post*, however, locations shifts are relatively easily detected and modeled.

*Financial support from the UK Economic and Social Research Council under grant RES-062-23-0061 and award PTA-030-2003-00904 (Fawcett) is gratefully acknowledged. We are grateful to Peter Boswijk, J. James Reade, Timo Teräsvirta, and an anonymous referee for helpful comments on an earlier draft.

On the other hand, for *ex ante* forecasting, problems abound: predicting the precise timing, form, and magnitude of location shifts seems beyond the scope of present methods. Thus, we consider whether forecast performance can be improved by forecasting during a break, which either alters the parameters of the model in use, or is induced by changes elsewhere in the economy. As shown in Hendry (2006), the class of EqCMs includes most econometric systems without imposed unit roots (regressions, VARs, DSGEs, GARCH and special cases thereof), and they all share a range of generic properties, of which the crucial feature here is their convergence back to a pre-shift equilibrium. Consequently for clarity, we establish the behavior of any proposed solution in the simplest EqCM exemplar, rather than in a general cointegrated system. This serves to highlight the key aspect of each proposal, and guides the empirical modeling illustration.

When a break is not predicted, possible solutions to avoiding continued forecast failure include the well-known techniques of intercept corrections and differencing, both being 'robust' solutions to past unmodeled location shifts. These are considered below as benchmarks against which to judge improvements in forecasting during breaks. For the former, Hendry and Santos (2005) consider 'setting the model on track' at the forecast origin, while Reade (2007) investigates smoothing recent corrections, in both cases for a step shift. Here, we examine their behavior during an evolving break. For the latter, Hendry (2006) proposes an explanation for the forecasting success of so called 'naive' devices, such as differencing, and also derives a hybrid that transforms a vector EqCM to robustify forecasts against past breaks, while retaining the causal information embodied in the cointegration relationships. That approach is investigated here for an evolving break, since it performs well even in the absence of knowledge of the causes of the break when applied to forecast UK M1 after 1984, which is the illustration we use throughout.

While both these 'corrections' can improve dramatically over simply maintaining the system estimated prior to the break, and retain the causal information embodied in the equilibrium-correction terms–of importance in any policy context–neither is optimal: both double the forecast-error variances as a trade-off against reductions in biases. Moreover, little is also known about updating as the time after a break increases. Thus, this paper also considers potential improvements in forecasting during location shifts by estimating the impacts of breaks during their progress. Even if the initial onset of a break is not forecast, since the final effect will generally differ from its impact in dynamic processes, methods for estimating the outcomes of breaks during their progress may still prove valuable. In particular, by modeling the break process, its future progress can be anticipated, which we show could improve forecasts. Doing so requires a modified forecasting model which embodies an 'auxiliary' device to forecast the progress of the break, such that the resulting forecasts are usefully accurate and their forecast-error uncertainty is accurately measured: the type of approach in (e.g.) Pesaran, Pettenuzzo and Timmermann (2006) offers possibilities for the latter, and is not considered here. These are demanding requirements because of the very changes inherently induced by location shifts in equilibria, which thereby automatically alter the pre-break collinearity structure. This can lead to a marked and unavoidable increase in uncertainty with an adverse impact on mean square forecast errors (MSFEs), despite the increased information which results from the break. Such an impact turns out to be unavoidable, so for example, deleting collinear variables does not help unless they are in fact essentially irrelevant: see Clements and Hendry (2005). Nevertheless, section 3 establishes that rapid information updates at the forecast origin can dramatically reduce the forecast uncertainty deriving from changed collinearities.

The structure of the paper is as follows. Section 2 reviews the forecast failure resulting from location shifts in a cointegrated or equilibrium-correction system using UK M1 as an example, following the Banking Act of 1984, which legalized interest-bearing sight deposits, and hence radically shifted the opportunity costs of holding money, altering demand relative to the prevailing information. Section 3 investigates forecast accuracy following an 'external break' that alters collinearity but leaves the forecasting model unchanged, and demonstrates the advantages of updating parameter estimates in that setting.

Section 4 analyzes the relative performance of various forecasting devices during an internal break, including intercept corrections, differencing and estimating a break-adjustment function in a simple setting relevant to the empirical illustration in section 5. Section 5 investigates forecasting during a break for a model of UK M1, and contrasts the 'agnostic' modeling of the break effects with the learning function approach in Baba, Hendry and Starr (1992) and Hendry and Ericsson (1991) which uses economic theory to constrain the effect through the shift in opportunity costs of holding money. Section 6 concludes.

2 Forecast failure in a cointegrated system

Location shifts – changes in equilibrium means – are the most pernicious form of breaks for cointegrated systems as they induce non-stationarity and systematic forecast failure. The theory behind this claim is established in Clements and Hendry (1998, 2006), and confirmed by their taxonomies of forecast errors. In practice, forecast failure remains common – and systematic – as figure 1 illustrates for a model of UK M1 estimated pre-1984: this failure will be the focus of the empirical illustration in section 5.

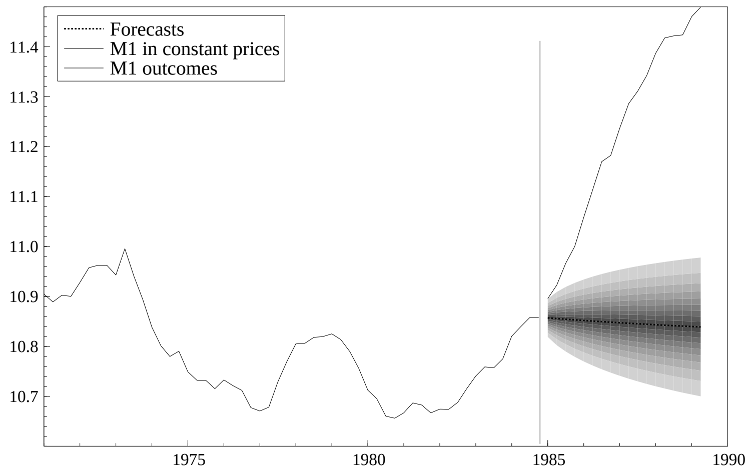

Figure 1: Outcomes for M1 with 95% forecast interval fans.

The sample of quarterly observations is over 1964(3) to 1989(2), for the following series:[1]

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_S	the interest rate on sight bank deposits

The main reason for the drastic failure seen in figure 1 is a shift in the equilibrium mean relative to its in-sample value. When that shift is not modeled, the estimated feedback coefficient converges on zero as the estimation sample grows, which allows better tracking of the changing data at the loss of cointegration: see e.g., Perron (1989) and Hendry and Neale (1991). Figure 2 reports recursive estimates of the associated feedback coefficient for the cointegrating relation $m - p = i - 8.7\Delta p + 6.6R_{LA}$ (lower case denotes logs). As can be seen, the intercept (which measures the growth rate of real money in a differenced relation), falls towards 0.007 (panel a), the estimated feedback coefficient ($\widehat{\alpha}_1$) converges on

[1]The data terminate in 1989(2) as the conversion to a bank by a major building society (the Abbey National) radically altered M1, when its retail sight deposits became part of M1.

zero (b), the 1-step ahead forecast errors lie outside the pre-existing 95% forecast interval (c), and the sequence of forecast Chow (1964) tests increasingly strongly rejects (d). The bottom row panels show the improvement in 'tracking' from full sample estimation (*ex post*, panel e), where $\widetilde{\alpha}_1 = -0.05\ (0.01)$ as against *ex ante*, where $\widehat{\alpha}_1 = -0.10\ (0.008)$ (panel f).

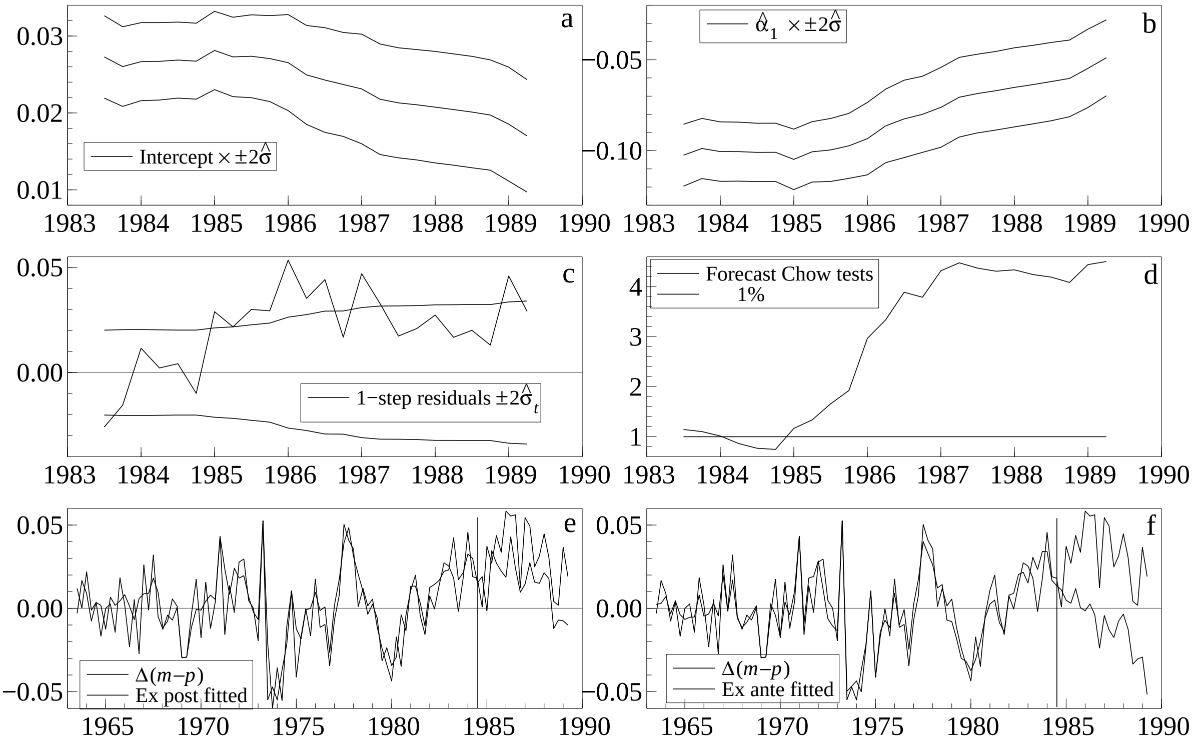

Figure 2: Recursive estimates of intercept, α_1, associated constancy tests, and fitted values.

Figure 3 shows how an unmodeled break affects the cointegrating vectors. Both panels show the coefficient on the interest rate R_{LA}, in the long-run money demand vector on the left, and the long-run output vector on the right, when estimated recursively after the break. If the model is left unadjusted, then the impact of the equilibrium-mean shift manifests in the interest rate coefficient in the money demand vector. This tries to adjust to compensate for a fall in the opportunity cost of holding money – which was hitherto measured by R_{LA} – even though, in reality, the slope parameters in the cointegrating relationship have not changed.

Since other (mean-zero) breaks are of less relevance for forecast failure, the crucial information needed to avoid systematic forecast failure in cointegrated systems is to forecast location shifts. As this is currently infeasible in general, one might attempt to forecast their ongoing effects after such shifts occur, as is the focus of this paper. At a minimum, adapting rapidly to a break can help considerably, so we also examine such strategies. Early detection of breaks using higher-frequency data could help in quick implementation of either adaptation, but a forecast-error taxonomy for time disaggregation shows that it does not *per se* reduce the impacts of location shifts on forecast errors (see Castle and Hendry, 2008), and measurement errors in high-frequency real-time data may attenuate that benefit. However, we first need to address the problem that, even though breaks increase data second moments and thereby lower estimation uncertainty, most breaks also induce collinearity changes, usually increasing it–at precisely the wrong moment from the forecaster's perspective.

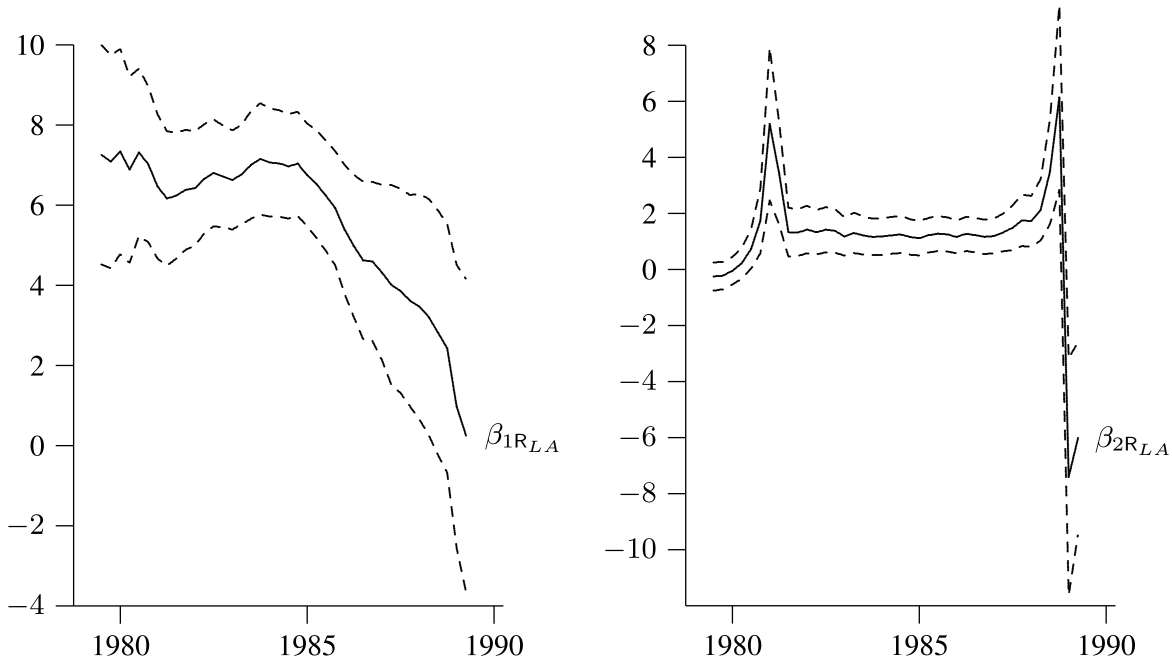

Figure 3: Recursive estimates of the coefficients on the interest rate R_{LA} in both cointegrating vectors. The left-hand panel shows the coefficient in the long-run money demand vector; the right shows that of the long-run trend output vector. Dotted lines show 95% confidence intervals.

3 Forecasting during an external break

Structural breaks generally alter the collinearities between variables, and so despite there being an increase in the information content of the data due to the shift, there is an offsetting adverse impact on MSFEs due to changes in collinearity. This is unavoidable, as deleting collinear variables does not mitigate the problem, unless the variables are irrelevant or nearly so. Hence, despite knowing the DGP, and even correctly predicting a location shift, parameter estimation updating is valuable in mitigating the impact of changing collinearity as we now show.

To illustrate the issues involved, we consider the simplest EqCM, namely a constant-parameter DGP correctly modeled by a conditional regression model:

$$y_t = \boldsymbol{\beta}'\mathbf{z}_t + \epsilon_t \text{ where } \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right] \tag{1}$$

with the marginal process:

$$\mathbf{z}_t \sim \mathsf{IN}_n\left[\boldsymbol{\mu}, \boldsymbol{\Sigma}\right] \tag{2}$$

distributed independently of $\{\epsilon_t\}$, where $\boldsymbol{\Omega} = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}' = \mathbf{H}'\boldsymbol{\Lambda}\mathbf{H}$ with $\mathbf{H}'\mathbf{H} = \mathbf{I}_n$ in-sample. Allowing for a post-sample change that:

$$\mathsf{E}\left[\mathbf{z}_{T+1}\mathbf{z}_{T+1}'\right] = \boldsymbol{\Omega}^* = \mathbf{H}'\boldsymbol{\Lambda}^*\mathbf{H},$$

after estimating (1) over the full sample as:

$$\widehat{\boldsymbol{\beta}}_{(1,T)} = \left(\sum_{t=1}^{T}\mathbf{z}_t\mathbf{z}_t'\right)^{-1}\sum_{t=1}^{T}\mathbf{z}_t y_t \simeq (T\Omega)^{-1}\sum_{t=1}^{T}\mathbf{z}_t y_t \tag{3}$$

with:

$$\mathsf{V}\left[\widehat{\boldsymbol{\beta}}_{(T)}\right] \simeq T^{-1}\sigma_\epsilon^2\Omega^{-1},$$

using

$$\widehat{y}_{T+h|T+h-1} = \widehat{\boldsymbol{\beta}}'_{(1,T)}\mathbf{z}_{T+h} \quad (4)$$

where $\widehat{\epsilon}_{T+1|T} = y_{T+1} - \widehat{y}_{T+1|T}$, the 1-step ahead MSFE for known regressors is:

$$\mathsf{E}\left[\widehat{\epsilon}^2_{T+1|T}\right] = \sigma^2_\epsilon \left(1 + T^{-1}\mathsf{E}\left[\mathbf{z}'_{T+1}\Omega^{-1}\mathbf{z}_{T+1}\right]\right) = \sigma^2_\epsilon \left(1 + \frac{1}{T}\sum_{i=1}^{n}\frac{\lambda^*_i}{\lambda_i}\right). \quad (5)$$

From (5), the increase in uncertainty will depend most on the ratio of the smallest eigenvalues, $\lambda^*_{(n)}/\lambda_{(n)}$ say, which could be large. If there is no shift, so $\lambda^*_i = \lambda_i \ \forall i$, then (5) reduces to the conventional $\sigma^2_\epsilon (1 + n/T)$, but (e.g.) for $T = 50$, $\lambda_{(n)} = 0.0001$ and $\lambda^*_{(n)} = 0.05$, then $\lambda^*_{(n)}/T\lambda_{(n)} = 10$, inducing a huge increase in MSFE. Other ratios may also matter, and (5) only rises when collinearity is reduced by a break, but that seems likely for a shift in $\boldsymbol{\mu}$.

A new result here is that one period later, after estimation updating:

$$\widetilde{\boldsymbol{\beta}}_{(1,T+1)} = \left(\frac{1}{T+1}\sum_{t=1}^{T+1}\mathbf{z}_t\mathbf{z}'_t\right)^{-1}\frac{1}{T+1}\sum_{t=1}^{T+1}\mathbf{z}_t y_t \simeq \left(T\boldsymbol{\Omega} + \boldsymbol{\Omega}^*\right)^{-1}\sum_{t=1}^{T+1}\mathbf{z}_t y_t, \quad (6)$$

with:

$$\mathsf{V}\left[\widetilde{\boldsymbol{\beta}}_{(1,T+1)}\right] \simeq \sigma^2_\epsilon \left(T\Omega + \Omega^*\right)^{-1} = \sigma^2_\epsilon \mathbf{H}'\left(T\boldsymbol{\Lambda} + \boldsymbol{\Lambda}^*\right)^{-1}\mathbf{H},$$

then:

$$\begin{aligned}\mathsf{E}\left[\widetilde{\epsilon}^2_{T+2|T+1}\right] &= \sigma^2_\epsilon \left(1 + tr\left\{\mathbf{H}'\left(T\boldsymbol{\Lambda} + \boldsymbol{\Lambda}^*\right)^{-1}\mathbf{H}\mathsf{E}\left[\mathbf{z}_{T+2}\mathbf{z}'_{T+2}\right]\right\}\right) \\ &= \sigma^2_\epsilon \left(1 + \sum_{i=1}^{n}\frac{\lambda^*_i}{T\lambda_i + \lambda^*_i}\right). \quad (7)\end{aligned}$$

The reduction in uncertainty depends most on the effect from the smallest eigenvalue ratio. If $\widehat{\boldsymbol{\beta}}_{(1,T)}$ is retained, then (5) would continue to hold as $\lambda^*_{(n)}/\left(\left[T+1\right]\lambda_{(n)}\right)$, as against $\lambda^*_{(n)}/\left(T\lambda_{(n)} + \lambda^*_{(n)n}\right)$ in (7). For $T = 50$, $\lambda_{(n)} = 0.0001$ and $\lambda^*_{(n)} = 0.05$, the former is $0.05/\left(51 \times 0.0001\right) = 9.8$, and the latter $0.05/\left(0.05 + 0.05\right) = 0.5$, leading to a dramatic reduction of the estimation uncertainty contribution.

We also obtain the impact on rolling-window estimators $\check{\boldsymbol{\beta}}_{(k,T)}$:

$$\mathsf{E}\left[\check{\epsilon}^2_{T+1|T}\right] = \sigma^2_\epsilon \left(1 + \sum_{i=1}^{n}\frac{\lambda^*_i}{\left(T-k+1\right)\lambda_i}\right) \quad (8)$$

which must be worse than (5). However at $T+1$ using $\widetilde{\boldsymbol{\beta}}_{(k+1,T+1)}$, so some of the break period is in-sample, then a substantive reduction ensues:

$$\mathsf{E}\left[\check{\epsilon}^2_{T+2|T+1}\right] = \sigma^2_\epsilon \left(1 + \sum_{i=1}^{n}\frac{\lambda^*_i}{\left(T-k+1\right)\lambda_i + \lambda^*_i}\right). \quad (9)$$

These impacts of a change in collinearity at the forecast origin were assessed in a Monte Carlo experiment based on (1) in which $\mathbf{z}_t$ consists of an intercept and two white-noise processes. The parameter values set to result in a population non-centrality of 2 using (14), with $T = 50$ and a forecast horizon $h = 10$. The intercept $\beta_0 = 0.28$ for a non-centrality of 2. We consider changes in Ω at time T from Ω to Ω^*. 1-step forecasts are computed for $T+1$ to $T+h$, for known regressors, and MSFEs are recorded for each forecast horizon. The forecasts considered include those from (4) based on in-sample parameter estimates (3); forecasts $\widetilde{y}_{T+h|T+h-1} = \widetilde{\boldsymbol{\beta}}'_{(1,T+h-1)}\mathbf{z}_{T+h}$ based on updating parameter estimates using the available sample:

$$\widetilde{\boldsymbol{\beta}}_{(1,T+h-1)} = \left(\sum_{t=1}^{T+h-1} \mathbf{z}_t \mathbf{z}_t'\right)^{-1} \sum_{t=1}^{T+h-1} \mathbf{z}_t y_t \tag{10}$$

and forecasts $\overline{y}_{T+h|T+h-1} = \overline{\boldsymbol{\beta}}'_{(k+h-1,T+h-1)}\mathbf{z}_{T+h}$ based on rolling parameter estimates:

$$\overline{\boldsymbol{\beta}}_{(k+h-1,T+h-1)} = \left(\sum_{t=k+h}^{T+h-1} \mathbf{z}_t \mathbf{z}_t'\right)^{-1} \sum_{t=k+h}^{T+h-1} \mathbf{z}_t y_t \tag{11}$$

where $k = 10, 25$.[2]

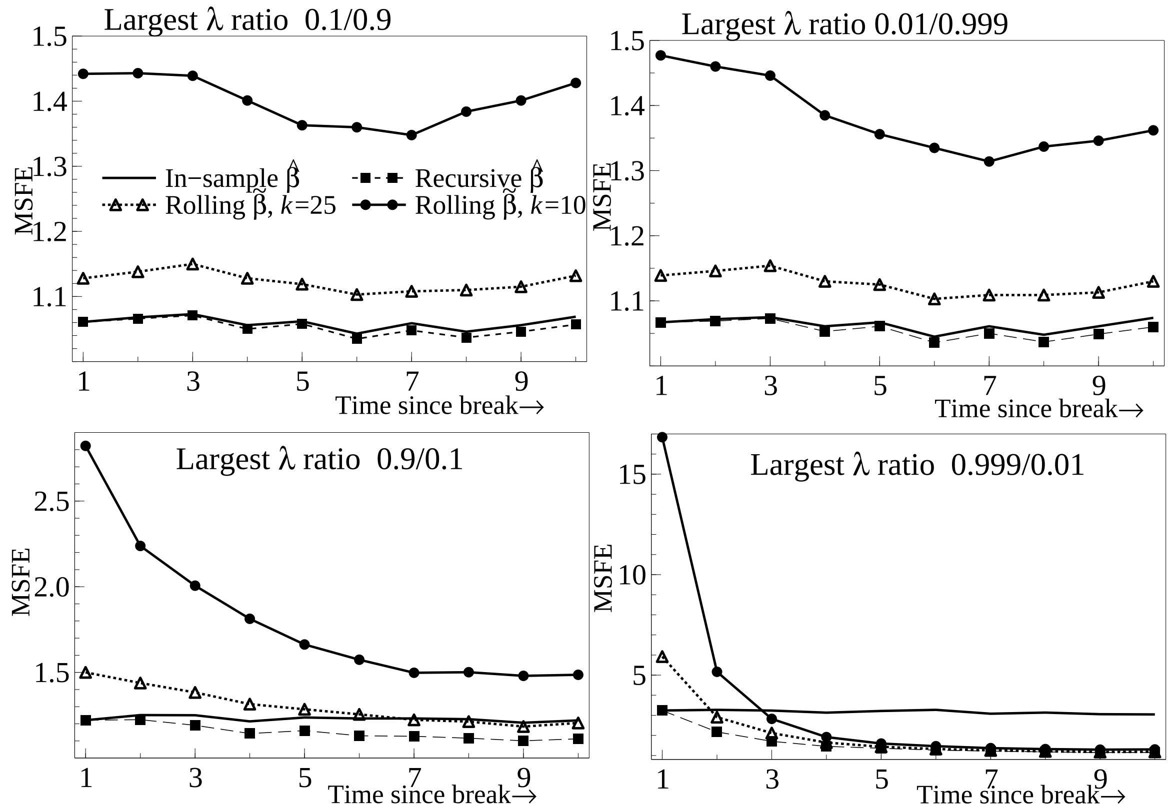

Figure 4: Breaks in collinearity

Figure 4 records the results for the various Ω and Ω^* reported in table 1. When the smallest eigenvalue ratio is small, there is little impact on the estimation contribution (see panels a and b), and short rolling windows are very detrimental. When the smallest eigenvalue ratio is larger, recursive estimation yields a greater improvement over the in-sample estimation, and rolling estimation does reasonably well some time after the break: see panel c. For a large eigenvalue ratio, panel d, the impact of changing collinearity is evident. After 2 steps, all updating methods are preferable to the forecasts based on in-sample estimates. Rapid updating is required to capture both the location shift and the change in collinearity between variables caused by the location shift. Observe that the impact of the shift depends on the ratio of eigenvalues and not the levels of correlation themselves. A shift from a correlation of 0.99 to 0.001 appears extreme, but a shift of the same magnitude results from $\lambda_{(n)} = 0.001$ to $\lambda^*_{(n)} = 0.1$.

[2] $M = 10,000$ replications were undertaken with $\sigma^2_\epsilon = 1$.

Panel	$\mathbf{\Sigma}$	$\mathbf{\Sigma}^*$	$\boldsymbol{\lambda}$	Smallest λ ratio
a	$\begin{pmatrix} 1 & 0.1 \\ 0.1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$	$0.9 \to 0.1$	0.11
b	$\begin{pmatrix} 1 & 0.001 \\ 0.001 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.99 \\ 0.99 & 1 \end{pmatrix}$	$0.999 \to 0.01$	0.01
c	$\begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.1 \\ 0.1 & 1 \end{pmatrix}$	$0.1 \to 0.9$	9
d	$\begin{pmatrix} 1 & 0.99 \\ 0.99 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.001 \\ 0.001 & 1 \end{pmatrix}$	$0.01 \to 0.999$	99.9

Table 1: Ω, Ω^* and smallest eigenvalue ratio for Monte Carlo experiments in figure 4.

Model mis-specification can exacerbate the changing collinearity problem. Partition $\mathbf{z}_t = \left(\mathbf{z}'_{1,t} : \mathbf{z}'_{2,t}\right)$, where $n_1 + n_2 = n$ with $\boldsymbol{\beta}' = \left(\boldsymbol{\beta}'_1 : \boldsymbol{\beta}'_2\right)$, and the $\mathbf{z}'_{i,t}$ are mutually orthogonal, as the 'best case'. Collinearity between $\mathbf{z}_{1,t}$ and $\mathbf{z}_{2,t}$ can be handled by orthogonalization, as happens anyway on omitting the latter. Omitting $\mathbf{z}_{2,t}$, the forecast is:

$$\overline{y}_{T+1|T} = \mathbf{z}'_{1,T+1}\overline{\boldsymbol{\beta}}_1, \tag{12}$$

with unconditional MSFE:

$$\mathsf{E}\left[\overline{\epsilon}^2_{T+1|T}\right] = \sigma^2_\epsilon \left(1 + \sum_{i=1}^{n_1} \frac{\lambda^*_i}{T\lambda_i}\right) + \boldsymbol{\beta}'_2\boldsymbol{\Lambda}^*_{22}\boldsymbol{\beta}_2. \tag{13}$$

This trade-off is key to forecast-model selection. Since $\boldsymbol{\beta}'_2\boldsymbol{\Lambda}^*_{22}\boldsymbol{\beta}_2 = \sum_{i=n_1+1}^{n} \beta_i^2\lambda_i$, defining the non-centrality of the squared t-test of H_0: $\beta_j = 0$ by:

$$\tau^2_{\beta_j} \simeq \frac{T\beta_j^2\lambda_j}{\sigma^2_\epsilon} \tag{14}$$

the unconditional MSFE is (see Clements and Hendry, 2005):

$$\mathsf{E}\left[\overline{\epsilon}^2_{T+1|T}\right] \simeq \mathsf{E}\left[\widehat{\epsilon}^2_{T+1|T}\right] + \frac{\sigma^2_\epsilon}{T}\sum_{j=n_1+1}^{n}\left(\tau^2_{\beta_j} - 1\right)\frac{\lambda^*_j}{\lambda_j}. \tag{15}$$

The impact of interactions between collinearity and mis-specification were assessed in a Monte Carlo experiment based on the previous design in which the mis-specification is omitting one of the regressors. Figure 5 illustrates the results for the largest shift in collinearity in the previous simulation with a smallest λ ratio of 99.9. For $\tau^2 > 1$ we set $\beta_0 = 0.2, \beta_1 = \beta_2 = 2$ such that $\tau^2 = 2$ in-sample for an eigenvalue of $\lambda_i = 0.01$, and for $\tau^2 < 1$, we set $\beta_0 = 0.1, \beta_1 = \beta_2 = 1$ such that $\tau^2 = 0.5$.

When $\tau^2 > 1$, the MSFEs of the mis-specified model are substantially worse for most estimation methods. Under constant collinearity, the mis-specified model's forecasts using rolling windows of 10 and 25 observations outperform the equivalent correctly-specified forecasts due to estimation uncertainty. For changing collinearity, all mis-specified models perform much worse, except for a lower MSFE in the period immediately after the break ($T + 1$) and using a rolling window with 10 observations. Note the much greater scale on the graphs for changing collinearity.

However, when $\tau^2 < 1$, so the omitted variable has a low non-centrality, the mis-specified model can do better than the estimated DGP under both constant and changing collinearity. Under constant collinearity, the forecasts from the mis-specified models all improve on the correctly-specified models,

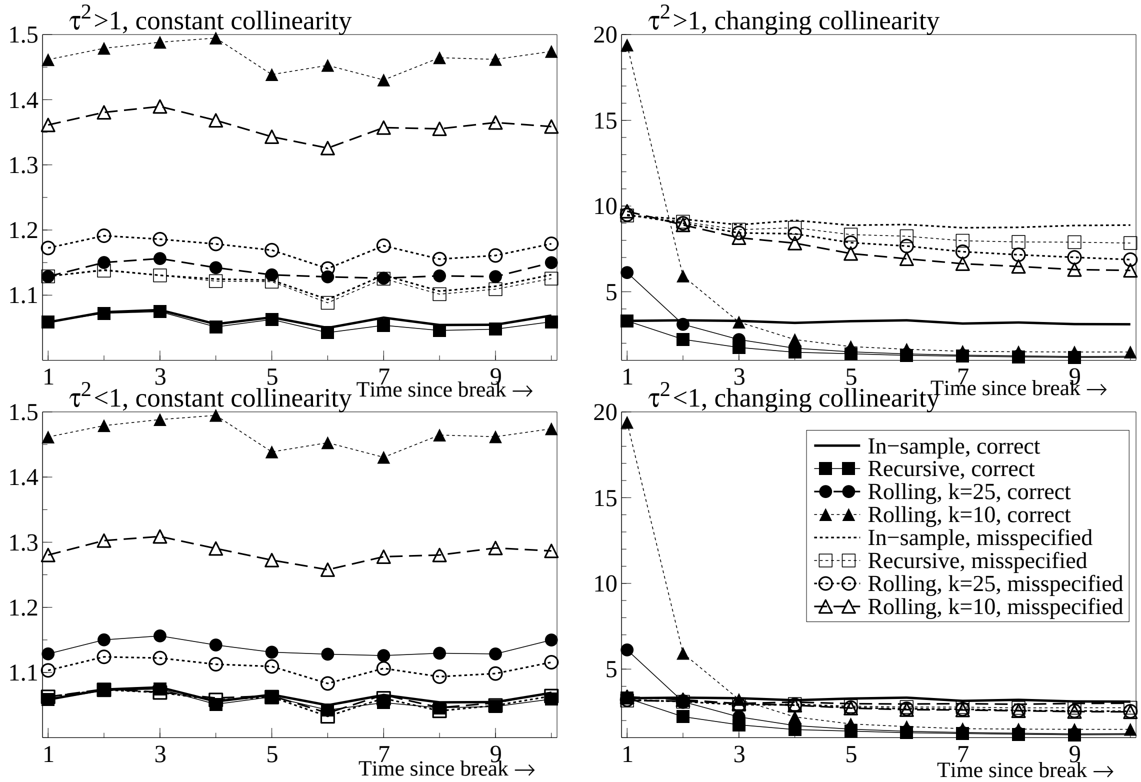

Figure 5: Interactions of collinearity with mis-specification for a large break

albeit the gains are marginal for the full sample and recursive estimates: this demonstrates that parameter estimation uncertainty is of less importance here. Under changing collinearity, all forecasts outperform the un-updated correct in-sample model after 3 steps, although correctly-specified models that update still outperform mis-specified models.

We also considered smaller changes in collinearity. Figure 6 records the impact of a shift in collinearity given by the smallest λ ratio of 9 in panels a and b, and 0.11 in panels c and d (see table 1 for the corresponding covariance matrices), again with parameters chosen to deliver $\tau^2 = 2$ or 0.5 in-sample. When the mis-specification is large, i.e. $\tau^2 > 1$, the correctly-specified models mostly outperform the mis-specified models, although rolling windows do incur a cost such that mis-specified models can be preferable. When $\tau^2 < 1$ the mis-specified models often have a lower MSFE than the correctly-specified models, in keeping with the above analysis.

Four conclusions follow from this analysis of changing collinearity when using mis-specified models. First, one minimizes $\mathsf{E}[\overline{\epsilon}^2_{T+1|T}]$ by eliminating all regressors where $\tau^2_{\beta_j} < 1$. Secondly $\mathsf{E}[\overline{\epsilon}^2_{T+1|T}]$ is possibly smaller than the MSFE of the estimated DGP equation. Thirdly, one cannot forecast better simply by dropping collinear variables if they are relevant with $\tau^2_{\beta_j} > 1$. Fourthly, updating only reduces the first term in (13), so the mis-specification component increases relatively. Hence it is especially valuable to correctly eliminate or retain variables when $\lambda^*_{(n)}/\lambda_{(n)}$ is large.

4 Forecasting during an internal break

Again we consider the simplest DGP, now given by:

$$y_t = \alpha + \lambda \left[1 - \exp\left(-\psi\left[t - T + 1\right]\right)\right] 1_{\{t \geq T\}} + \epsilon_t \text{ with } \epsilon_t \sim \mathsf{IN}\left[0, \sigma^2_\epsilon\right] \tag{16}$$

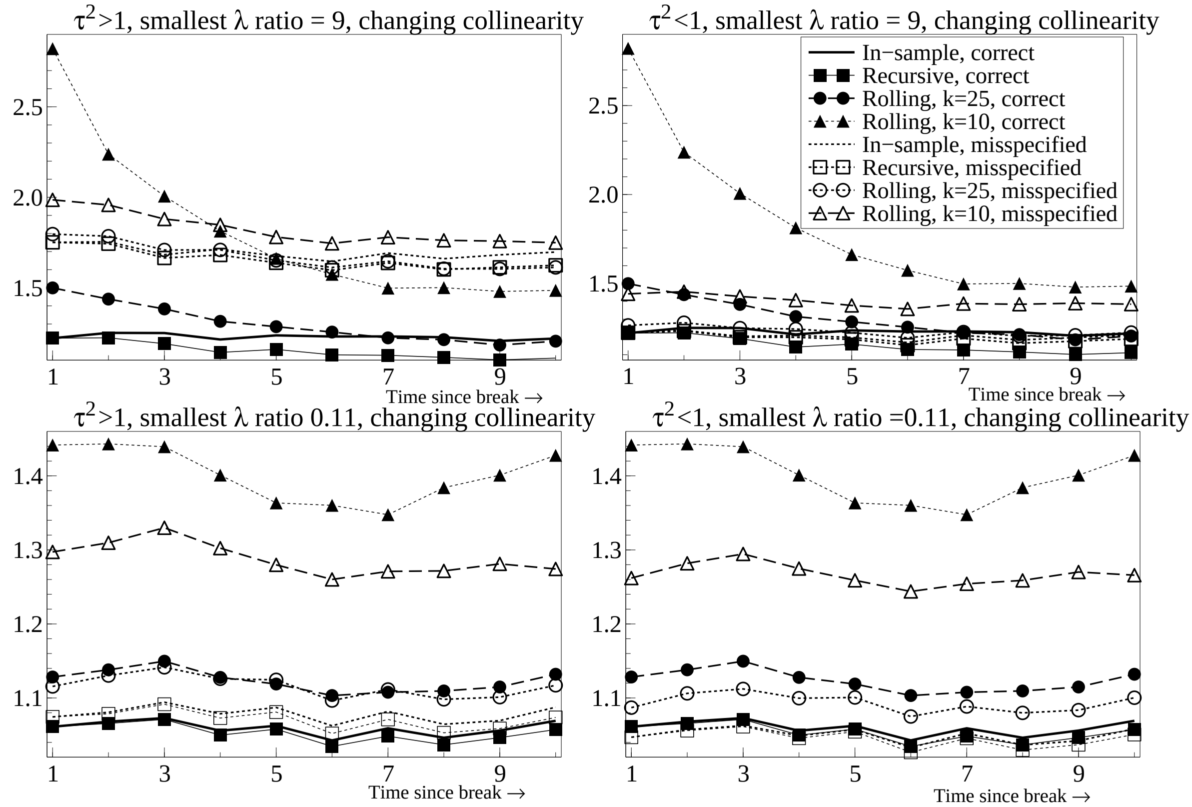

Figure 6: Interactions of collinearity with mis-specification

where $1_{\{t \geq T\}}$ is an indicator function which is zero before time T, and unity thereafter, and $\psi > 0$, although α and λ could have either sign. The existence of a break at time T is known, but the form it takes is not known, to match section 5. Figure 7 shows the plot of an exemplar of (16) when $\lambda = 1$ and $\psi = 0.2$, with the own interest rate (R_S but scaled by 10) following the Banking Act. Despite its simplicity, (16) can be interpreted as having y_t as the deviation of Δx_t from an equilibrium correction, so α denotes the unconditional growth rate, and $\lambda [1 - \exp(-\psi [t - T + 1])] 1_{\{t \geq T\}}$ models the shift in the equilibrium mean.

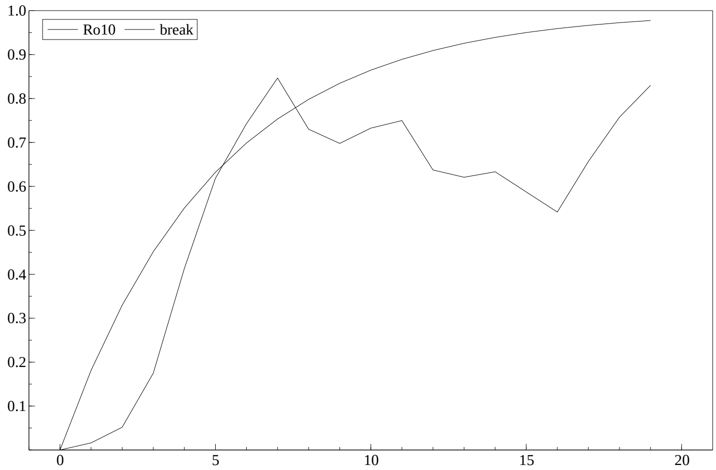

Figure 7: Artificial break function and the own interest rate

At the break:

$$y_T = \alpha + \lambda [1 - \exp(-\psi)] + \epsilon_T$$

and the outcome at $T+1$ from (16) is:

$$y_{T+1} = \alpha + \lambda\left(1 - \exp\left(-2\psi\right)\right) + \epsilon_{T+1}. \tag{17}$$

We consider four alternative 1-step ahead forecasting devices: 4.1 an intercept-corrected model (IC); 4.2 a differenced device; 4.3 an estimated version of (16); and 4.4 ignoring the break, forecasting each of $T+1$ from T, and then $T+2$ from $T+1$ in subsection 4.5.

4.1 Intercept correction

Here the model of (16) is:

$$y_t = \gamma + \delta D_{\{T\}} + v_t \tag{18}$$

where $D_{\{T\}}$ is a dummy variable with the value unity from T onwards, and is zero otherwise. As $D_{\{T\}} = 1$, using the full-sample data, $\widetilde{\gamma}$ is the sample mean over $1 \ldots T-1$, so $\mathsf{E}\left[\widetilde{\gamma}\right] = \alpha$:

$$\mathsf{V}\left[\widetilde{\gamma}\right] = \frac{\sigma_\epsilon^2}{T-1} \tag{19}$$

and:

$$\widetilde{\delta} = y_T - \widetilde{\gamma} = \left(\alpha - \widetilde{\gamma}\right) + \lambda\left(1 - \exp\left(-\psi\right)\right) + \epsilon_T \tag{20}$$

so $\mathsf{E}\left[\widetilde{\delta}\right] = \lambda\left(1 - \exp\left(-\psi\right)\right)$ and:

$$\mathsf{V}\left[\widetilde{\delta}\right] = \mathsf{V}\left[\alpha - \widetilde{\gamma}\right] + \sigma_\epsilon^2 = \sigma_\epsilon^2\left(1 + \left(T-1\right)^{-1}\right). \tag{21}$$

The forecast is $\widetilde{y}_{T+1|T} = \widetilde{\gamma} + \widetilde{\delta}$, so from (17) the error is $\widetilde{\epsilon}_{T+1|T} = y_{T+1} - \widetilde{y}_{T+1|T}$:

$$\widetilde{\epsilon}_{T+1|T} = \left(\alpha - \widetilde{\gamma}\right) + \lambda\left(1 - \exp\left(-2\psi\right)\right) - \widetilde{\delta} + \epsilon_{T+1} = \lambda \exp\left(-\psi\right)\left(1 - \exp\left(-\psi\right)\right) + \Delta\epsilon_{T+1},$$

with:

$$\mathsf{E}\left[\widetilde{\epsilon}_{T+1|T}\right] = \lambda \exp\left(-\psi\right)\left(1 - \exp\left(-\psi\right)\right)$$

and as $\mathsf{C}[\widetilde{\delta}, \widetilde{\gamma}] = -\mathsf{V}\left[\widetilde{\gamma}\right]$:

$$\mathsf{E}\left[\left(\widetilde{\epsilon}_{T+1|T} - \mathsf{E}\left[\widetilde{\epsilon}_{T+1|T}\right]\right)^2\right] = \mathsf{E}\left[\left(\Delta\epsilon_{T+1}\right)^2\right] = 2\sigma_\epsilon^2 \tag{22}$$

so as usual, an IC doubles the forecast error variance. Thus, the MSFE is:

$$\mathsf{M}\left[\widetilde{\epsilon}_{T+1|T}\right] = \lambda^2 \exp\left(-2\psi\right)\left[1 - \exp\left(-\psi\right)\right]^2 + 2\sigma_\epsilon^2. \tag{23}$$

4.2 Differenced device

This is simply:

$$\Delta\overline{y}_{T+1|T} = 0 \tag{24}$$

or $\overline{y}_{T+1|T} = y_T$, so from (17), letting $\overline{\epsilon}_{T+1|T} = y_{T+1} - \overline{y}_{T+1|T}$:

$$\overline{\epsilon}_{T+1|T} = \lambda \exp\left(-\psi\right)\left[1 - \exp\left(-\psi\right)\right] + \Delta\epsilon_{T+1} \tag{25}$$

with $\mathsf{E}\left[\overline{\epsilon}_{T+1|T}\right] = \lambda \exp\left(-\psi\right)\left[1 - \exp\left(-\psi\right)\right]$ and:

$$\mathsf{E}\left[\left(\overline{\epsilon}_{T+1|T} - \mathsf{E}\left[\overline{\epsilon}_{T+1|T}\right]\right)^2\right] = \mathsf{E}\left[\left(\epsilon_{T+1} - \epsilon_T\right)^2\right] = 2\sigma_\epsilon^2.$$

Differencing also doubles the forecast error variance, so the MSFE is:

$$\mathsf{M}\left[\overline{\epsilon}_{T+1|T}\right] = \lambda^2 \exp\left(-2\psi\right)\left[1 - \exp\left(-\psi\right)\right]^2 + 2\sigma_\epsilon^2 \tag{26}$$

which in this setting is identical to IC.

4.3 Estimated DGP

Next, using the sample mean over $1 \ldots T-1$ for $\widehat{\alpha}$ in (16), since λ and ψ cannot be separately estimated at T, the forecast from the estimated DGP is again identical to the IC:

$$\widehat{y}_{T+1|T} = \widehat{\alpha} + \widetilde{\delta} \tag{27}$$

so the MSFE is:

$$\mathsf{M}\left[\widehat{\epsilon}_{T+1|T}\right] = \lambda^2 \exp\left(-2\psi\right)\left[1-\exp\left(-\psi\right)\right]^2 + 2\sigma_\epsilon^2. \tag{28}$$

Thus, (23), (26) and (28) are identical, despite (26) not estimating any parameters. This result is not likely to hold for forecasting y_{T+2} from a forecast origin at $T+1$, as the third method should 'track' the evolving break better than the first two.

4.4 An unadjusted model

Finally, if no action is taken in the face of the break:

$$\overleftarrow{y}_{T+1|T} = \overleftarrow{\alpha} \tag{29}$$

where $\overleftarrow{\alpha} = \sum_{t=1}^{T} y_t/T$, then:

$$\mathsf{E}\left[\overleftarrow{\alpha}\right] = \alpha + \frac{1}{T}\lambda\left(1-\exp\left(-\psi\right)\right) \quad \text{and} \quad \mathsf{V}\left[\overleftarrow{\alpha}\right] = \frac{\sigma_\epsilon^2}{T},$$

so for $\overleftarrow{\epsilon}_{T+1|T} = y_{T+1} - \overleftarrow{y}_{T+1|T}$:

$$\mathsf{E}\left[\overleftarrow{\epsilon}_{T+1|T}\right] = \lambda\left(1-\exp\left(-2\psi\right)\right) - \frac{1}{T}\lambda\left(1-\exp\left(-\psi\right)\right)$$

with MSFE:

$$\mathsf{M}\left[\overleftarrow{\epsilon}_{T+1|T}\right] = \left(\mathsf{E}\left[\overleftarrow{\epsilon}_{T+1|T}\right]\right)^2 + \sigma_\epsilon^2\left(1+\frac{1}{T}\right) \tag{30}$$

The ranking varies with the DGP parameter values, specifically the size of the break relative to σ_ϵ^2, but in general (30) will be larger than (23), (26) and (28) for breaks bigger than σ_ϵ.

4.4.1 Numerical illustration

For the parameter values in figure 7, namely $\lambda = 1$ and $\psi = 0.2$ when $\sigma_\epsilon = 0.1$ and $T = 100$, which is an initial shift of $1.8\sigma_\epsilon$, then (23)=(26)=(28) and (30) have the values 0.0420 and 0.1176, see figure 8. Thus, the 'corrections' for the break outperform relative to taking no action, and would do so even for values of σ_ϵ as large as 0.25.

4.5 2-periods later

Now the outcome is:

$$y_{T+2} = \alpha + \lambda\left(1-\exp\left(-3\psi\right)\right) + \epsilon_{T+2} \tag{31}$$

4.5.1 Intercept correction

Updating using (18), $\widetilde{\gamma}$ remains as before, but there are two possibilities, namely extending $D_{\{T\}}$, or setting $D_{\{T\}} = 1_{\{T\}}$ and adding $D_{\{T+1\}}$. We consider the former here, as the latter yields a similar analysis to 4.1 moved forward one period. Thus:

$$\widetilde{\delta}_2 = \frac{1}{2}\left(y_T + y_{T+1}\right) - \widetilde{\gamma} = \alpha - \widetilde{\gamma} + \frac{1}{2}\left(\lambda\left(1 - \exp\left(-\psi\right)\right) + \lambda\left(1 - \exp\left(-2\psi\right)\right)\right) + \frac{1}{2}\left(\epsilon_T + \epsilon_{T+1}\right) \quad (32)$$

so $\mathsf{E}[\widetilde{\delta}_2] = \lambda - \lambda\exp\left(-\psi\right)\left(1 + \exp\left(-\psi\right)\right)/2$ and:

$$\mathsf{V}\left[\widetilde{\delta}_2\right] = \mathsf{V}\left[\alpha - \widetilde{\gamma}\right] + \frac{1}{2}\sigma_\epsilon^2 = \sigma_\epsilon^2\left(\frac{1}{2} + \left(T-1\right)^{-1}\right). \quad (33)$$

The forecast is $\widetilde{y}_{T+2|T+1} = \widetilde{\gamma} + \widetilde{\delta}_2$, with error $\widetilde{\epsilon}_{T+2|T+1} = y_{T+2} - \widetilde{y}_{T+2|T+1}$ so:

$$\begin{aligned}\widetilde{\epsilon}_{T+2|T+1} &= \left(\alpha - \widetilde{\gamma}\right) + \lambda\left(1 - \exp\left(-3\psi\right)\right) - \widetilde{\delta}_2 + \epsilon_{T+2} \\ &= \frac{\lambda\exp\left(-\psi\right)}{2}\left(1 + \exp\left(-\psi\right)\right) - \lambda\exp\left(-3\psi\right) + \epsilon_{T+2} - \frac{1}{2}\left(\epsilon_T + \epsilon_{T+1}\right)\end{aligned}$$

where $\mathsf{E}\left[\widetilde{\epsilon}_{T+2|T+1}\right] = \lambda\exp\left(-\psi\right)\left(1 + \exp\left(-\psi\right) - 2\exp\left(-2\psi\right)\right)/2$, and hence:

$$\mathsf{E}\left[\left(\widetilde{\epsilon}_{T+2|T+1} - \mathsf{E}\left[\widetilde{\epsilon}_{T+2|T+1}\right]\right)^2\right] = 1.5\sigma_\epsilon^2,$$

so the MSFE is:

$$\mathsf{M}\left[\widetilde{\epsilon}_{T+2|T+1}\right] = \left(\mathsf{E}\left[\widetilde{\epsilon}_{T+2|T+1}\right]\right)^2 + 1.5\sigma_\epsilon^2 \quad (34)$$

Estimating two parameters here to offset the break only adds 50% to the forecast error uncertainty over a known DGP, but 'smoothing' the IC imparts a bias.

4.5.2 Differencing

This forecasting device remains:

$$\overline{y}_{T+2|T+1} = y_{T+1} \quad (35)$$

so letting $\overline{\epsilon}_{T+2|T+1} = y_{T+2} - \overline{y}_{T+2|T+1}$:

$$\overline{\epsilon}_{T+2|T+1} = \lambda\exp\left(-2\psi\right)\left(1 - \exp\left(-\psi\right)\right) + \Delta\epsilon_{T+2} \quad (36)$$

with $\mathsf{E}\left[\overline{\epsilon}_{T+2|T+1}\right] = \lambda\exp\left(-2\psi\right)\left(1 - \exp\left(-\psi\right)\right)$ and $\mathsf{E}[\left(\overline{\epsilon}_{T+2|T+1} - \mathsf{E}\left[\overline{\epsilon}_{T+2|T+1}\right]\right)^2] = 2\sigma_\epsilon^2$, so the MSFE is:

$$\mathsf{M}\left[\overline{\epsilon}_{T+2|T+1}\right] = \left(\mathsf{E}\left[\overline{\epsilon}_{T+2|T+1}\right]\right)^2 + 2\sigma_\epsilon^2. \quad (37)$$

This is generally smaller than (34), as:

$$\mathsf{E}\left[\widetilde{\epsilon}_{T+2|T+1}\right] - \mathsf{E}\left[\overline{\epsilon}_{T+2|T+1}\right] = \lambda\exp\left(-\psi\right)\left(1 - \exp\left(-\psi\right)\right)/2 > 0$$

so the 2-period averaged IC should be dominated by the differenced model for forecasting $T+2$.

4.5.3 Estimated DGP

Now λ and ψ can be estimated by non-linear optimization, albeit with considerable uncertainty, from

$$\min_{\lambda,\psi} \sum_{k=0}^{1} \epsilon_{T+k}^{2}.$$

Using the form in (16) allows 'extrapolation' to the outcome in the next period, namely:

$$\widehat{y}_{T+2|T+1} = \widehat{\alpha} + \widehat{\lambda}\left(1 - \exp\left(-3\widehat{\psi}\right)\right) \tag{38}$$

where $\widehat{\alpha}$ remains as before. The forecast error $\widehat{\epsilon}_{T+2|T+1} = y_{T+2} - \widehat{y}_{T+2|T+1}$ from (38) is:

$$\widehat{\epsilon}_{T+2|T+1} = (\alpha - \widehat{\alpha}) + \left(\lambda - \widehat{\lambda}\right) - \lambda \exp\left(-3\psi\right) + \widehat{\lambda} \exp\left(-3\widehat{\psi}\right) + \epsilon_{T+2}$$

so $\mathsf{E}\left[\widehat{\epsilon}_{T+2|T+1}\right] \simeq 0$; and $\mathsf{E}[\widehat{\epsilon}^{2}_{T+2|T+1}]$ depends on the (co)variances of all the estimated parameters. The MSFE need not exceed $\mathsf{M}\left[\overline{\epsilon}_{T+2|T+1}\right]$, due to eliminating the bias by projecting an additional period ahead, but at 2 periods later for this 'ogive' break, estimating the DGP seems unlikely to outperform given the potentially large variance effects. However, as the post-break sample increases, its performance relative to mechanistic corrections should improve.

4.5.4 An unadjusted model

If no action is taken after the break:

$$\overleftarrow{y}_{T+2|T+1} = \overleftarrow{\alpha} \tag{39}$$

where $\overleftarrow{\alpha} = \sum_{t=1}^{T+1} y_t / T + 1$, so for $\overleftarrow{\epsilon}_{T+2|T+1} = y_{T+2} - \overleftarrow{y}_{T+2|T+1}$:

$$\mathsf{E}\left[\overleftarrow{\epsilon}_{T+2|T+1}\right] = \lambda\left(1 - \exp\left(-3\psi\right)\right) - \frac{1}{T+1}\left[\lambda\left(1 - \exp\left(-\psi\right)\right) + \lambda\left(1 - \exp\left(-2\psi\right)\right)\right]$$

with MSFE:

$$\mathsf{M}\left[\overleftarrow{\epsilon}_{T+2|T+1}\right] = \left(\mathsf{E}\left[\overleftarrow{\epsilon}_{T+2|T+1}\right]\right)^2 + \sigma_\epsilon^2\left(1 + \frac{2}{T+1}\right). \tag{40}$$

4.6 Simulation results

The results in sections 4.1–4.5 were checked in a simulation in which the DGP is generated by (16) for the parameter values in 4.4.1: $\lambda = 1$, $\psi = 0.2$, $\sigma_\epsilon = 0.1$ and $T = 100$. $M = 10,000$ replications were undertaken, and results for up to 4 periods after the break for the 4 alternative forecast devices are reported in figure 8 which summarises the results.[3] The results demonstrate that estimating the 'ogive' break does yield a lower MSFE than the IC, even just 2 periods after the break, due to a reduction in bias, but not relative to the differencing device where the MSFEs are similar even up to 4 periods later. All updating devices do substantially better than ignoring the break. Thus, taking no action does not seem advisable unless the shift is small.

[3]The parameters for the 2–4 period forecasts for the estimated DGP were estimated by non-linear least squares for fixed $\widehat{\alpha}$, with initial conditions of $\lambda_0 = 1$ and $\psi_0 = 0.25$.

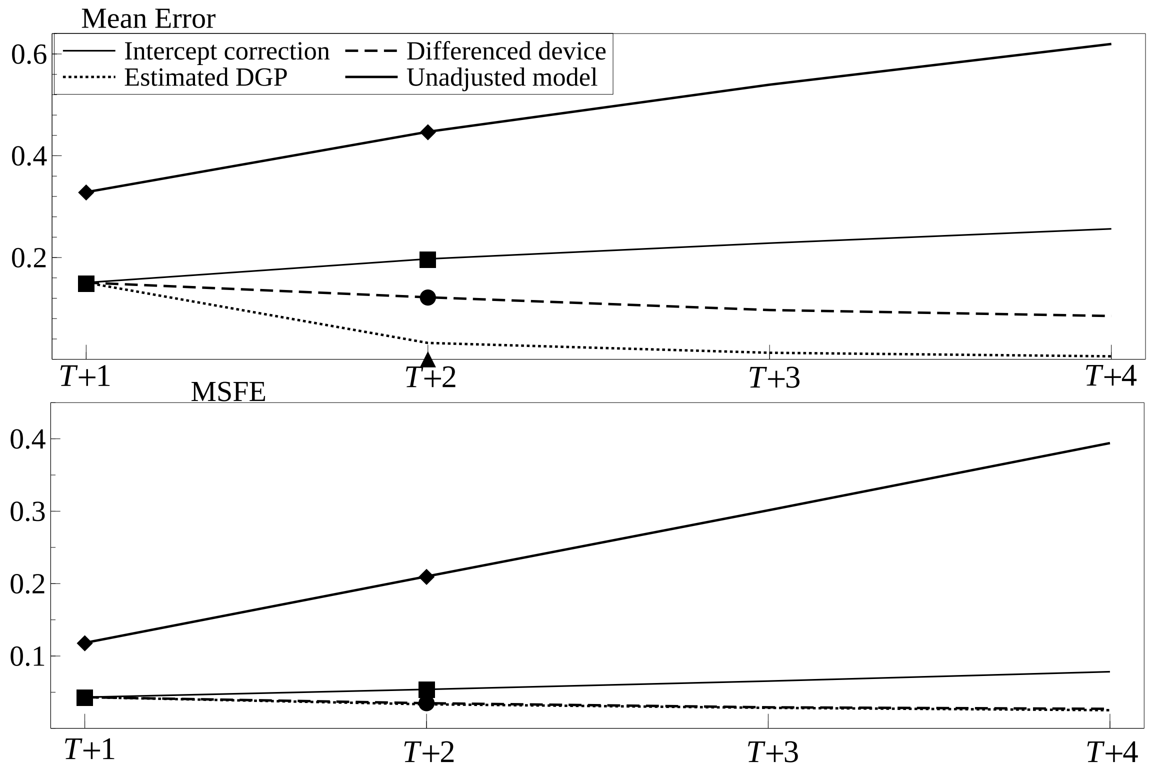

Figure 8: Mean error and MSFE for up to 4 periods after the break for the non-linear exponential DGP. Symbols indicate values for $T+1$ and $T+2$ from the theoretical formulae above.

5 Empirical illustration

The phenomena discussed above are illustrated in an application to the data on money demand in the UK, using the time series defined in section 2. We use the models in Hendry and Mizon (1993) and Hendry and Doornik (1994) (also see Hendry, 1979, Hendry and Ericsson, 1991, Boswijk, 1992, Boswijk and Doornik, 2004 and Hendry, 2006), and express the variables as a cointegrated system. The theoretical basis is a model which links demand for real money, $m-p$ (lowercase denoting logs) to (log) income i (transactions motive) and the interest rate R_{LA} measuring the opportunity cost of holding money.

The Banking Act of 1984 altered the opportunity cost, since it simultaneously allowed banks to pay interest on checking accounts (which is captured by R_S, the interest rate on sight deposits) but required them to report interest income payments to the Inland Revenue. The latter change affected individuals who had hitherto not paid income tax on interest income (previously banks were not obliged to report). By switching wealth to checking accounts, though, they were able to earn R_S interest on deposits, whilst previous non-payment of interest income could be attributed to its absence on such accounts. In combination, these forces provided both the demand and supply for a shift in assets from non-M1 deposits to M1.

The break caused by the Banking Act is relevant to the framework developed above, in two respects. First, the dramatic shift in the opportunity cost shown in figure 9, panel (b) (as $R_D = R_{LA} - R_S$) is a significant non-linearity that needs to be modelled correctly to avoid the systematic forecast failure shown in figure 1 for a model using R_{LA}. Secondly, since the cause of the change was legislative, and knowledge of it did not spread immediately across the economy, the case offers potential insights into forecasting during a structural change through modeling the learning process, a third variant as in (41).

The baseline system is from Hendry and Mizon (1993), which uses R_{LA} as the opportunity cost measure, and we contrast this with the learning-adjusted measure used by Hendry and Doornik (1994). The process by which agents in the economy learn of the structural change is captured by a weighting function w_t of the form analyzed in section 4:

$$w_t = \begin{cases} (1 + \exp[\alpha - \beta(t - t^* + 1)])^{-1} & \text{for } t \geq t^* \\ 0 & \text{otherwise} \end{cases} \tag{41}$$

with $\alpha,\ \beta > 0$, and $t^* = 1984(3)$. The weight is applied to the interest rate on sight deposits R_S, and the resulting series $R_{0,t} = R_{LA} - w_t R_{S,t}$ is added to the system. In estimating the parameters α and β in the weight function, and then using the function itself in the system, we can address two important issues:

1. How quickly did agents respond to the shift in the opportunity cost, and at what point would an econometrician have been able to learn about this?

2. Provided that they succeeded in modeling the adjustment process, how soon could an econometrician have used this updated knowledge to forecast better than the alternatives of no adjustment or a robust method?

The first sheds light on the adjustment process in the aftermath of the break. We can address that question by using the single-equation specification in Hendry and Ericsson (1991) to obtain estimates for α and β in a non-linear regression.[4] Over the whole sample period, 1964(3) to 1989(2), the parameter estimates obtained for α and β are 3.16 and 0.74 respectively, consistent with previous studies in the literature: Hendry and Ericsson (1991) found estimates of 5.0 and 1.2 for the same coefficients using US data.[5] Using these estimates, we construct a weighted series, applying the full-sample estimates at all points following the break at $t^* = 1984(3)$, which is shown as the solid line in panel (a) of figure 9. The weight is zero before 1984(3), climbs to approximately 0.6 within four quarters of t^* and is close to unity after eight quarters. Thus, adjustment was moderate, with nearly complete transition to the interest rate differential $R_{D,t} = R_{LA,t} - R_{S,t}$ after two years, although the fact that it took one year to adjust half way suggests some sluggishness.

From the perspective of an observer at the time, we can establish how quickly plausible estimates of α and β could have been obtained, through recursive estimation of the non-linear equation. Since there are two parameters, the first viable estimates are available in 1985(1), the third quarter following the break, but these estimates, $\widehat{\alpha} = 85.8$ and $\widehat{\beta} = 28.9$, are wild, so for the first three periods after the break, we impose $w_t = 0$. Thereafter, parameter estimates are stable, allowing plausible values of w_t to be constructed, as shown by the dotted line in panel (a) of figure 9. Thus, by the second quarter of 1985 – within a year of the break taking place – an econometrician with knowledge of R_S (which was in the public domain) could have obtained a realistic estimate of the adjustment process in (41).

To answer the second question, we embed the resulting learning-adjusted interest rate $R_{0,t} = R_{LA,t} - w_t R_{S,t}$ shown in panel (b) of figure 9 into the four-variable system discussed above (i.e., replacing $R_{LA,t}$ with $R_{0,t}$). Then, using an expanding estimation window from 1984(2), we can compare the root MSFEs (RMSFEs) of the set of eight dynamic forecasts from the end of each recursive estimation point. Figure 10 shows the RMSFE series for the baseline model incorporating R_{LA} (denoted $\mathsf{M}_{R_{LA}}$) and the adjusted model with R_0 (M_{R_0}). As panel (a) demonstrates, once a learning adjustment is incorporated into the forecast – at 1985(2) – there is a dramatic improvement in forecast accuracy, with the RMSFE falling by 55%, which Morgan–Granger–Newbold tests of comparative forecast accuracy show to be statistically

[4] In doing so, we can apply, and test, the restriction that the coefficient on the $w_t R_{s,t}$ term is of the same magnitude but opposite sign to that on R_{LA}. Statistically, this restriction is accepted, and in terms of economic theory, this formulation is consistent with $(R_{LA,t} - w_t R_{s,t})$ measuring the opportunity cost.

[5] All estimation is conducted using Oxmetrics: see Doornik and Hendry (2006).

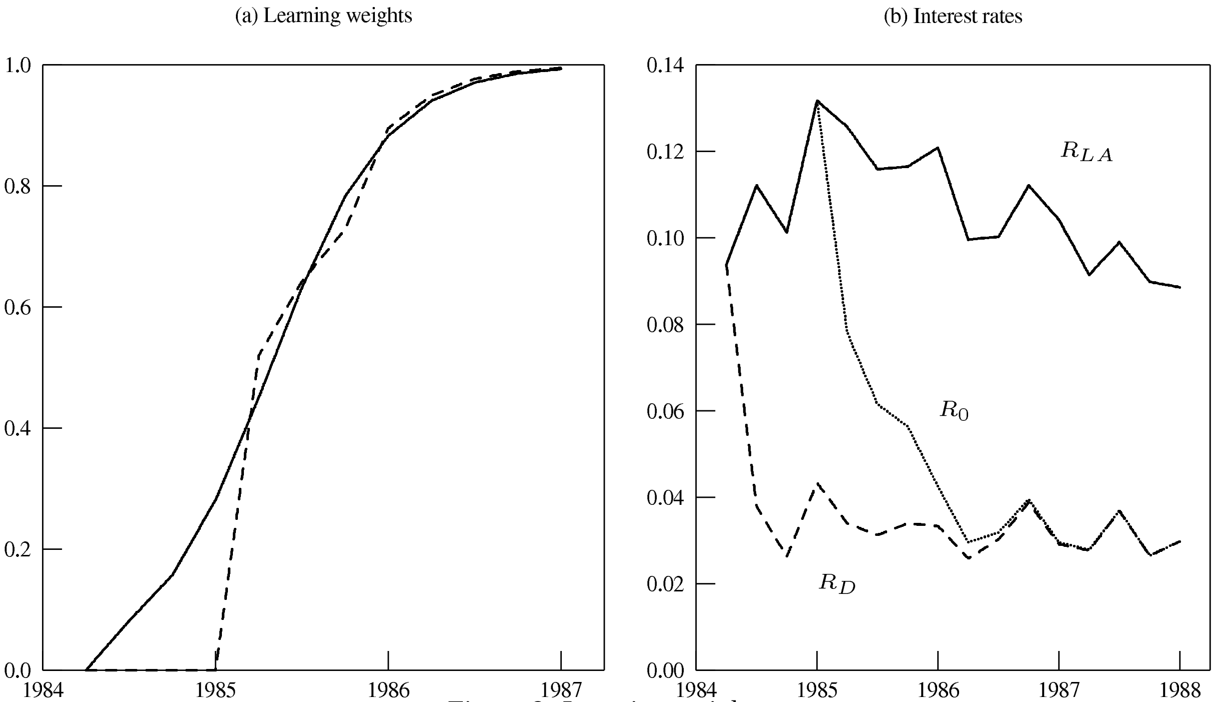

Figure 9: Learning weights

significant (see Morgan, 1940). Even before the weight w_t reaches one (for instance, in 1985(2), when $w_t = 0.519$), there is still a benefit from including an adjusted interest rate series. By 1986(2), nearly two years after the break, this benefit is even larger (w_t is close to unity by this stage), although the bulk of improvement in accuracy occurs within the first year.

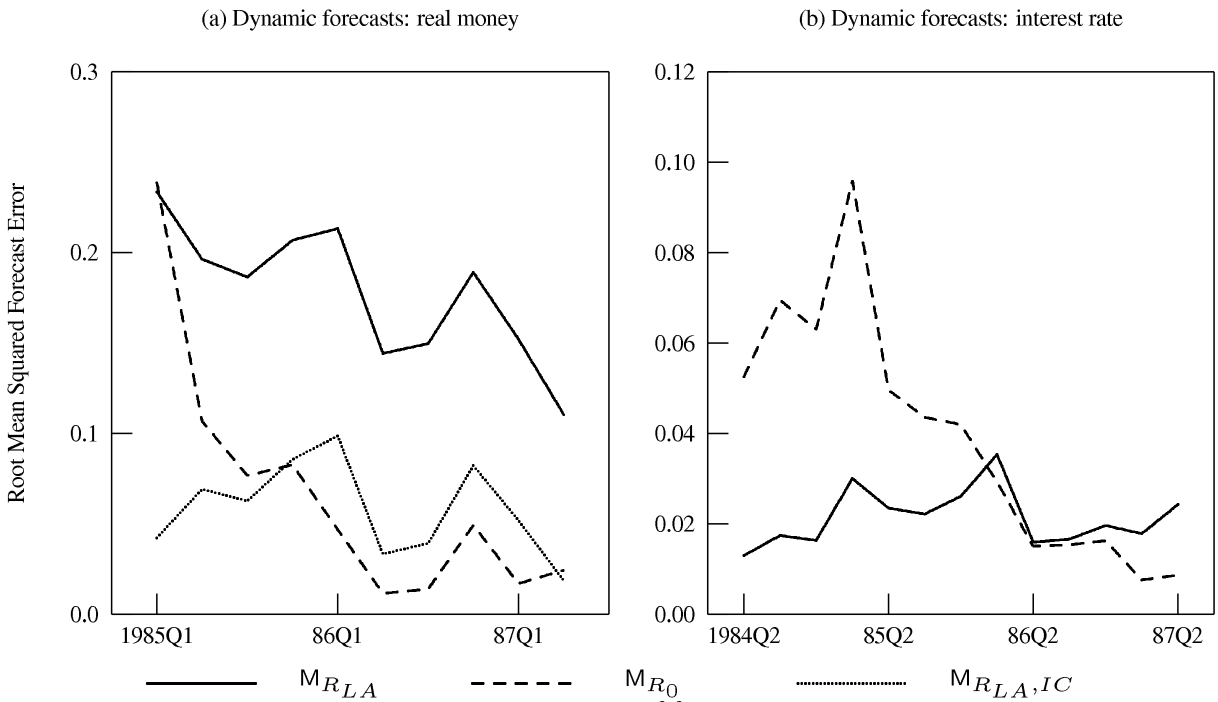

Figure 10: Comparisons of forecast accuracy

The early success of this learning adjustment in improving accuracy is itself beaten by intercept-corrected forecasts, which appear as a dotted line in panel (a) of figure 10. These were obtained by adding an intercept correction term like that in (18), to an $\mathsf{I}(0)$ representation of the cointegrated system.

The IC dummy in this case is equal to one in 1985(1) and 1985(2), and found to be significant in only the real money equation (and therefore the IC-adjusted model is omitted from panel (b) of figure 10).[6]

Using the IC to set the forecasts of money growth back on track, and then integrating these growth rates to provide forecasts in levels, there is an immediate improvement, confirming the benefit of this approach that was identified in section 4. After three periods, though, the ogive-adjusted model dominates, suggesting a role for explicit modeling of the break process.

Noting that RMSFEs are measured relative to the model's own interest rate (for $\mathsf{M}_{R_{LA}}$ this is R_{LA}, and for M_{R_0} it is R_0), panel (b) in figure 9 reveals an interesting point about the structural break (in the form of a step shift) that occurs in the interest rate series $R_{0,t}$: since there is an (*ex ante* unpredictable) fall in the interest rate, forecasting models such as M_{R_0} suffer from predictive failure because they attempt to forecast post-break observations using pre-break data. Since this shift did not affect R_{LA}, we can now explain why $\mathsf{M}_{R_{LA}}$ actually forecasts its own interest rate better than M_{R_0} does. Once the learning-adjusted model succeeds in forecasting its interest rate more accurately, there is a concomitant improvement in accuracy for the real money forecasts as well.

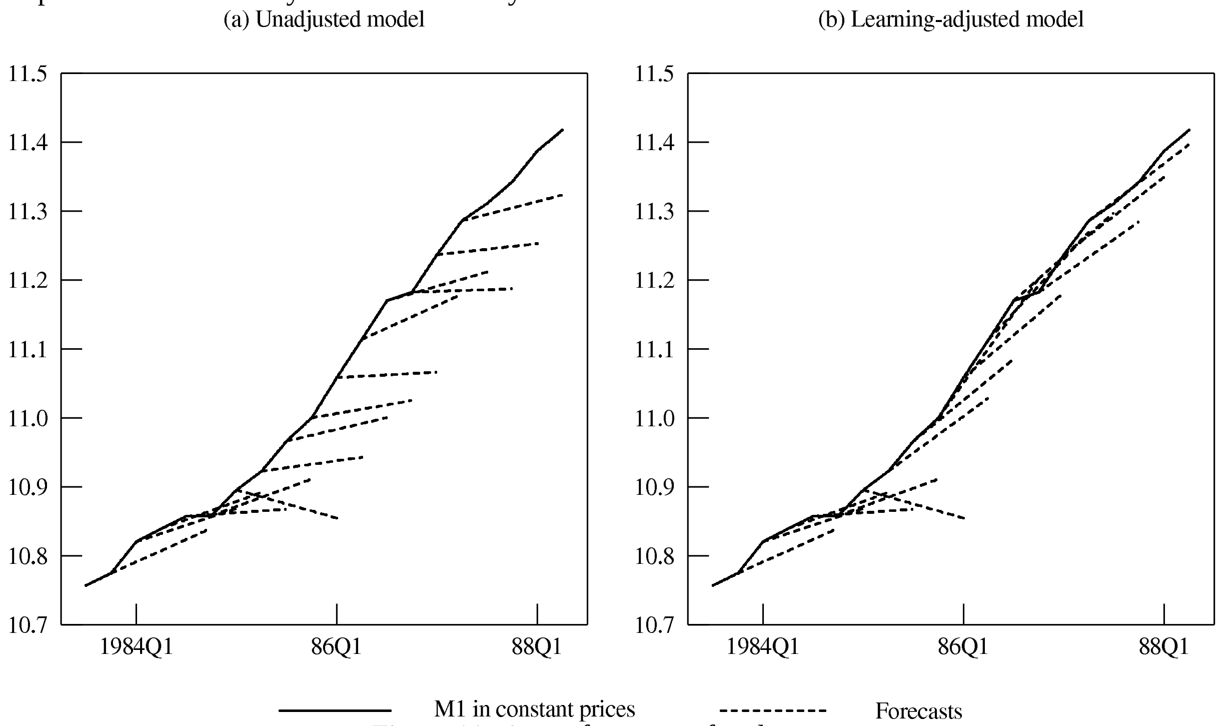

Figure 11: 4-step forecasts of real money

We can see the importance of a correct equilibrium mean when forecasting with VEqCM models by comparing panels (a) and (b) of figure 11. Both show a sequence of four-step forecasts of the level of real money before, during and after the break in the opportunity cost. Prior to the change, both the learning-adjusted and unadjusted models share identical forecasts, and demonstrate some success in forecasting the rate of increase in real money. However, when the Banking Act alters the cointegrating equilibrium mean, the unadjusted model forecasts start to undershoot real money persistently. Underlying this is the fact that inside the cointegrated system, the growth in real money responds partly to any long-run disequilibrium in the system. When this changes in reality, but not in the forecast model, then problems ensue, as at each point the forecasts point to a correction back to the *incorrect* equilibrium mean. By using a learning-adjusted model in panel (b), the change in the mean is captured successfully, which explains why this model can forecast more accurately after 1985(2), when learning is incorporated.

[6]Prior to 1985(1), the IC-adjusted and unadjusted models are therefore identical, so panel (a) only reports the RMSE comparisons for subsequent periods.

6 Conclusions

Changes in equilibrium means create serious forecasting problems for all members of the large class of equilibrium-correction models (EqCMs), including those based on cointegration. We have considered several approaches to alleviating forecast failure following such location shifts, including updating, intercept corrections, differencing, and estimating the future impact of a break during its progress.

We find that updating can help when collinearites are changed by an 'outside' break but the EqCM itself remains constant. Dropping collinear variables does not help, even when such variables are not very significant. Updating the pre-break model helps offset forecast failure for location shifts in the EqCM, but at a cost of loss of cointegration. Mechanistic corrections help in that setting as well, compared to just retaining a pre-break estimated model. A model of the break process should be able to outperform, but seems to require at least three periods into an evolving break to do so.

The much-studied example of EqCMs for UK M1 illustrated the analysis, and showed some gains from updating a learning function as the break evolves. That empirical example suggests that even if a break is not predicted, its course can sometimes be tracked quite well relatively soon after its occurrence, and economic analysis based restrictions had clear value added after a learning modification.

We conclude that progress in developing forecasting models with an ability to adapt to breaks and forecast during the progress of a break is feasible, and could be invaluable in cointegrated processes.

References

Baba, Y., Hendry, D. F., and Starr, R. M. (1992). The demand for M1 in the U.S.A., 1960–1988. *Review of Economic Studies*, **59**, 25–61.

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

Boswijk, H. P., and Doornik, J. A. (2004). Identifying, estimating and testing restricted cointegrated systems: An overview. *Statistica Neerlandica*, **58**, 440–465.

Castle, J. L., and Hendry, D. F. (2008). Forecasting UK inflation: Empirical evidence on robust forecasting devices. In Rapach, D. E., and Wohar, M. E. (eds.), *Forecasting in the Presence of Structural Breaks and Model Uncertainty*, pp. 41–92. Amsterdam: Elsevier. Forthcoming.

Chow, G. C. (1964). A comparison of alternative estimators for simultaneous equations. *Econometrica*, **32**, 532–553.

Clements, M. P., and Hendry, D. F. (1998). *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. (2002). Explaining forecast failure in macroeconomics. In Clements, M. P., and Hendry, D. F. (eds.), *A Companion to Economic Forecasting*, pp. 539–571. Oxford: Blackwells.

Clements, M. P., and Hendry, D. F. (2005). Guest editors' introduction: Information in economic forecasting. *Oxford Bulletin of Economics and Statistics*, **67**, 713–753.

Clements, M. P., and Hendry, D. F. (2006). Forecasting with breaks in data processes. In Elliott, G., Granger, C. W. J., and Timmermann, A. (eds.), *Handbook of Econometrics on Forecasting*, pp. 605–657. Amsterdam: Elsevier.

Davidson, J. E. H., Hendry, D. F., Srba, F., and Yeo, J. S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal*, **88**, 661–692.

Doornik, J. A., and Hendry, D. F. (2006). *Empirical Econometric Modelling using PcGive: Volume I*. 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.

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

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.

Hendry, D. F. (2000). On detectable and non-detectable structural change. *Structural Change and Economic Dynamics*, **11**, 45–65.

Hendry, D. F. (2004). The Nobel Memorial Prize for Clive W.J. Granger. *Scandinavian Journal of Economics*, **106**, 187–213.

Hendry, D. F. (2006). Robustifying forecasts from equilibrium-correction models. *Journal of Econometrics*, **135**, 399–426.

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 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 Juselius, K. (2000). Explaining cointegration analysis: Part I. *Energy Journal*, **21**, 1–42.

Hendry, D. F., and Juselius, K. (2001). Explaining cointegration analysis: Part II. *Energy Journal*, **22**, 75–120.

Hendry, D. F., and Mizon, G. E. (1993). Evaluating dynamic econometric models by encompassing the VAR. In Phillips, P. C. B. (ed.), *Models, Methods and Applications of Econometrics*, pp. 272–300. Oxford: Basil Blackwell.

Hendry, D. F., and Neale, A. J. (1991). A Monte Carlo study of the effects of structural breaks on tests for unit roots. In Hackl, P., and Westlund, A. H. (eds.), *Economic Structural Change, Analysis and Forecasting*, pp. 95–119. Berlin: Springer-Verlag.

Hendry, D. F., and Santos, C. (2005). Regression models with data-based indicator variables. *Oxford Bulletin of Economics and Statistics*, **67**, 571–595.

Morgan, W. A. (1939–1940). A test for the significance of the difference between the two variances in a sample from a normal bivariate population. *Biometrika*, **31**, 13–19.

Perron, P. (1989). The Great Crash, the oil price shock and the unit root hypothesis. *Econometrica*, **57**, 1361–1401.

Pesaran, M. H., Pettenuzzo, D., and Timmermann, A. (2006). Forecasting time series subject to multiple structural breaks. *Review of Economic Studies*, **73**, 1057–1084.

Phillips, A. W. H. (1954). Stabilization policy in a closed economy. *Economic Journal*, **64**, 290–333.

Reade, J. J. (2007). Macroeconomic modelling and forecasting in the face of non-stationarity. Unpublished thesis, Economics Department, Oxford University.

Sargan, J. D. (1964). Wages and prices in the United Kingdom: A study in econometric methodology (with discussion). In Hart, P. E., Mills, G., and Whitaker, J. K. (eds.), *Econometric Analysis for National Economic Planning*, Vol. 16 of *Colston Papers*, pp. 25–63. London: Butterworth Co.