ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

MULTI-STEP FORECASTING IN UNSTABLE ECONOMIES: ROBUSTNESS ISSUES IN THE PRESENCE OF LOCATION SHIFTS

Gullaume Chevillon

Number 257

February 2006

Manor Road Building, Oxford OX1 3UQ

Multi-step Forecasting in Unstable Economies: Robustness Issues in the Presence of Location Shifts

Guillaume Chevillon*

University of Oxford, Economics Department,

and

Institute of Political Studies (Sciences-Po), OFCE.

November, 2005

Abstract

To forecast at several, say h, periods into the future, a modeller faces two techniques: iterating one-step ahead forecasts (the IMS technique) or directly modelling the relation between observations separated by an h-period interval and using it for forecasting (DMS forecasting). It is known that unit-root non-stationarity and residual autocorrelation benefit DMS accuracy in finite samples. We analyze here the effect of structural breaks as observed in unstable economies, and show that the benefits of DMS stem from its better appraisal of the dynamic relationships of interest for forecasting. It thus acts in between congruent modelling and intercept correction. We apply our results to forecasting the South African GDP over the last thirty years as this economy exhibits significant unstability. We analyze the forecasting properties of 31 competing models. We find that the GDP of South Africa is best forecast, 4 quarters ahead, using direct multi-step techniques, as with our theoretical results.

Keywords: Multi-step forecasting, Structural breaks, South Africa.
JEL *Classification*: C32, C53, E3.

*Corresponding e-mail: guillaume.chevillon@sciences-po.fr. Corresponding address: FNSP-OFCE, 69 quai d'Orsay, 75340 Paris cedex 07, France. I am grateful to David Hendry, Janine Aron, John Muellbauer and Michael Clements for helpful comments and suggestions.

1 Introduction and framework for analysis

When a forecaster uses a model with a given periodicity but wishes to forecast at several, say $h > 1$, periods into the future, she is faced with a choice between iterating one-step ahead forecasts or *directly* modelling the relation between the end-of-sample observation and its hth successor in order to forecast the latter. This direct technique was first suggested by Cox (1961) in the context of exponential smoothing; Klein (1971) applied it to dynamic forecasting and Johnston (1974) put a temporary end to the analysis by concluding that, when a quadratic loss function is used as criterion for both estimation and forecast accuracy, the former constitutes a "reliable indicator of prediction efficiency "and hence *direct* methods are inefficient for forecasting. Findley (1983) and then Weiss (1991) re-examined the previous results in the context of misspecified models and found asymptotic relevance in matching estimation and forecast efficiency criteria, as with the direct multi-step method. The renewed interest these authors brought to the topic spurred new theoretical analyze by, inter alia, Tiao and Xu (1993), Clements and Hendry (1996) for simulated ARMA processes, Bhansali (1996, 1997) for long memory processes, Haywood and Tunnicliffe-Wilson (1997) in the frequency domain, Schorfheide (2002) for asymptotically vanishing misspecification, Findley, Potscher, and Wei (2004) for very general settings and most recently Chevillon and Hendry (2005) for misspecified ARIMA processes.

Despite the number of theoretical results in favour of DMS, few empirical analyses exhibited successful uses of direct multi-step methods, except notably Tiao and Tsay (1994) on the U.S. monthly price index for food, Tsay (1993) on the quarterly U.S. unemployment rate, Liu (1996) on the monthly unemployment rate in Taiwan. Results by other authors were more mixed, as in Lin and Tsay (1996) and Kang (2003) who found evidence both in favour and against; the major analysis of 171 U.S. monthly macroeconomic time series by Marcellino, Stock, and Watson (2005) strongly favoured iterative forecasting techniques. It should be noted that the authors who found evidence against direct multi-step, or DMS, mostly used post-war U.S. data, although over extended periods. The United States has not experienced in this era so significant shifts as those witnessed in less well developped countries; furthermore, by aggregation over such a large economy, some stability is expected. Yet, several of the theoretical analyses—as in Peña (1994) for breaks, Bhansali (1997) whose framework is close to that of fractionally integrated processes, and Chevillon and Hendry (2005) for occasional location shifts that induce negative serial correlation—point towards deterministic shifts as a key to the success of DMS in finite samples.

We purpose in this paper to analyze the theoretical properties of DMS forecasting in the presence of breaks and assess our results by observing the empirical performance in unstable economies. As shown in Clements and Hendry (1999), there is no hope to render a forecasting model systematically robust to future breaks in the data generating process: we must therefore focus on breaks occurring before the forecasting origin. Clements and Hendry also note that the class of breaks most detrimental to the forecast accuracy of econometric models is that of deterministic shifts, and the most pernicious are location, or intercept, shifts which are often modelled by impulse or step dummies. This result has been confirmed by Pesaran and Timmermann (2005) who consider the impact of location shifts and breaks on autoregressive coefficients: the latter induce no systematic bias in conditional and unconditional forecasts. Since we intend to assess the properties of direct multi-step forecasting, and given that we intend to generate 'post-break' forecasts, a specific case comes naturally: when the shifts occur a few periods before the forecast origin, so that, depending on which technique is used, their influences on estimation vary. Pesaran and Timmermann derive optimal one-step ahead estimation windows for autoregressive models in this framework. In the following, we analyze the particular case of a data generating process which can be written as the n–variate VAR(1):

$$\mathbf{x}_t = \boldsymbol{\gamma} + \boldsymbol{\Gamma}\mathbf{x}_{t-1} + \boldsymbol{\epsilon}_t, \quad \text{for } t = 1, ..., T, \tag{1}$$

where the disturbances, $\boldsymbol{\epsilon}_t$, are covariance stationary, with zero mean and variance-covariance matrix $\boldsymbol{\Omega}_\epsilon$. We allow for the eigenvalues of $\boldsymbol{\Gamma}$ to be less than or equal to unity in modulus, thus inducing integration and cointegration. The vector $\mathbf{x}_t$ can be quite general, and potentially comprises a time component so that the series may exhibit deterministic trends. We assume that we are interested in the univariate process $\{y_t\}$, where there exists $\mathbf{P} \in \mathbb{R}^n$, such that

$$y_t = \mathbf{P}'\mathbf{x}_t.$$

In this context, intercept breaks can be modelled by an innovative shift from $\boldsymbol{\gamma}$ to $\boldsymbol{\gamma}^* = \boldsymbol{\gamma} + \nabla\boldsymbol{\gamma}$, at the date $T_0 = T - k$, where $k \geq 0$. We denote by $\{\mathbf{x}_t^*\}$ the process when a break occurs and $\{\mathbf{x}_t\}$ the underlying, no-break, process. Hence, for $i \geq 0$, letting $\boldsymbol{\Gamma}^{\{i\}} = \sum_{j=0}^{i-1} \boldsymbol{\Gamma}^j$, (and $\boldsymbol{\Gamma}^{\{i\}} = \mathbf{0}$, for $i \leq 0$), then:

$$\mathbf{x}_{T_0+i}^* = \boldsymbol{\Gamma}^{\{i+1\}}\nabla\boldsymbol{\gamma} + \mathbf{x}_{T_0+i}. \tag{2}$$

To assess the impact of the break on estimation and forecasting, we partition the process into the stationary and non-stationary components $\{\boldsymbol{\beta}'\mathbf{x}_t, \boldsymbol{\alpha}_\perp'\mathbf{x}_t\}$, with $\boldsymbol{\Gamma} - \mathbf{I}_n = \boldsymbol{\alpha}\boldsymbol{\beta}'$ ($\boldsymbol{\alpha}$ and $\boldsymbol{\beta}'$ are $n \times r$ matrices of full rank, see Johansen (1996)) which satisfy:

$$\begin{bmatrix} \boldsymbol{\beta}'\mathbf{x}_t \\ \boldsymbol{\alpha}_\perp'\mathbf{x}_t \end{bmatrix} = \begin{bmatrix} (\mathbf{I}_r + \boldsymbol{\beta}'\boldsymbol{\alpha}) & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(n-r)\times r} & \mathbf{I}_{n-r} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}'\mathbf{x}_{t-1} \\ \boldsymbol{\alpha}_\perp'\mathbf{x}_{t-1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\beta}'\boldsymbol{\epsilon}_t \\ \boldsymbol{\alpha}_\perp'\boldsymbol{\epsilon}_t \end{bmatrix}, \tag{3}$$

where r is the order of cointegration and the eigenvalues of $(\mathbf{I}_r + \boldsymbol{\beta}'\boldsymbol{\alpha})$ are all strictly within the unit circle. The impact of a shift on (3) is for $i > 0$:

$$\begin{bmatrix} \boldsymbol{\beta}'\mathbf{x}_{T_0+i}^* \\ \boldsymbol{\alpha}_\perp'\mathbf{x}_{T_0+i}^* \end{bmatrix} = \begin{bmatrix} (\mathbf{I}_r + \boldsymbol{\beta}'\boldsymbol{\alpha})^{\{i+1\}} \boldsymbol{\beta}'\nabla\boldsymbol{\gamma} \\ (i+1)\,\boldsymbol{\alpha}_\perp'\nabla\boldsymbol{\gamma} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\beta}'\mathbf{x}_{T_0+i} \\ \boldsymbol{\alpha}_\perp'\mathbf{x}_{T_0+i} \end{bmatrix}.$$

For the stationary component, $\boldsymbol{\beta}'\mathbf{x}_t$, the effect of $\nabla\boldsymbol{\gamma}$ varies with the period and, asymptotically:

$$\lim_{i\to\infty} \left[\boldsymbol{\beta}'\mathbf{x}_{T_0+i}^* - \boldsymbol{\beta}'\mathbf{x}_{T_0+i}\right] = -\left[\boldsymbol{\beta}'\boldsymbol{\alpha}\right]^{-1}\boldsymbol{\beta}'\nabla\boldsymbol{\gamma},$$

so that the intercept shift constitutes a 'progressive' location shift. By contrast, the impact of $\nabla\boldsymbol{\gamma}$ is to add a deterministic trend to $\boldsymbol{\alpha}_\perp'\mathbf{x}_t$, in the directions where $\boldsymbol{\alpha}_\perp'\nabla\boldsymbol{\gamma}$ is non-zero. The innovative model (2) therefore provides a location shift only in the stationary directions—denoted by $sp(\boldsymbol{\beta})$—and a trend shift in the non-stationary—i.e. $sp(\boldsymbol{\alpha}_\perp)$—unless there exists $\mathbf{G}$ such that $\nabla\boldsymbol{\gamma} = \boldsymbol{\alpha}\mathbf{G}$. Given that $sp(\boldsymbol{\beta}, \boldsymbol{\alpha}_\perp) = \mathbb{R}^n$, we consider, in the following, that $\mathbf{x}_t = [\mathbf{z}_t', \mathbf{w}_t']'$, where the r-vector $\mathbf{z}_t$ is stationary and $\mathbf{w}_t$ is $(n-r)$-dimensional and integrated of order 1, but not cointegrated. Thus, we write:[1]

$$\mathbf{x}_t = \begin{bmatrix} \mathbf{z}_t \\ \mathbf{w}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma}_z & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-r} \end{bmatrix} \begin{bmatrix} \mathbf{z}_{t-1} \\ \mathbf{w}_{t-1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_{z,t} \\ \boldsymbol{\epsilon}_{w,t} \end{bmatrix},$$

[1] Such a partition may not exist for any vector $\mathbf{x}_t$, but, for any $\mathsf{I}(1)$ process, there is a linear combination, represented by the matrix $\mathbf{R}$, such that $\mathbf{R}\mathbf{x}_t$ can be decomposed into $[\mathbf{z}_t', \mathbf{w}_t']'$.

where the eigenvalues of $\mathbf{\Gamma}_z$ are less than one in modulus. Similarly $\nabla\boldsymbol{\gamma}' = [\nabla\boldsymbol{\gamma}'_z, \nabla\boldsymbol{\gamma}'_w]$. Notice that the series of interest, y_t, could potentially be modelled, via $\mathbf{P}$, by an unbalanced equation involving both $\mathbf{z}_t$ and $\mathbf{w}_t$, but we do not consider this situation as it should be avoided in empirical analyses.

In this article, we compare four forecasting techniques. Two are based on the VAR(1), but with different estimation methods: when (1) is estimated by ordinary least-squares, the corresponding multi-step parameter estimates and forecasts as referred to as iterated multi-step (IMS). Another method consists in using least-squares estimation on the *direct multi-step model* (DMS, see Clements and Hendry (1996)) or given $h > 1$:

$$\mathbf{x}_t = \boldsymbol{\gamma}_h + \mathbf{\Gamma}_h \mathbf{x}_{t-h} + \boldsymbol{\upsilon}_t, \quad \text{for } t = h, ..., T, \tag{4}$$

where $\boldsymbol{\upsilon}_t$ is not modelled, but still assumed to have zero mean. The parameter estimates and resulting forecasts of $\mathbf{x}_{T+h}$—from a forecast origin at T—are referred to as DMS. Two additional methods will be examined. Their aim is to put the forecast 'back on track' by adding to the forecasts from the two previous models the differences between the forecast origins and their fitted values from the estimated models (see Hendry (2005) for an analysis of its use in forecasting with econometric models). We refer to these methods as one-step and multi-step intercept corrections depending on the sample residual used. One-step intercept correction necessarily uses the IMS model (and we refer to it as IC), whereas multi-step can relate to either of IMS or DMS (and hence IMSIC or DMSIC).

The plan of this paper is as follows: we first observe, in section 2, the properties of multi-step VAR forecasting when the DGP undergoes a location shift and if we abstract from estimation issues. In the following sections, we then focus on the influence of the break on estimation (section 3) and therefore on forecasting with a misspecified model (section 4). We provide results from simulation in section 5. In section 6, we proceed to an empirical application to multi-step forecasts of the South African GDP: after reviewing the political and economic context of the last 30 years, we derive 31 competing forecasting models whose performance we assess. Finally we conclude about the use of multi-step methods in unstable economies in section 7.

2 Forecasting in the presence of breaks

We first assume that the presence of a structural break does not affect estimation significantly and that the model used for forecasting is therefore the pre-break one (which follows Clements and Hendry in their work, although this is criticized by Pesaran and Timmermann). This hypothesis corresponds to cases where k is small relative to the sample size and the model is well-specified for the data generating process in the absence of the break.

2.1 Multi-step forecasting

If it is desired to forecast h periods ahead from an end-of-sample forecast origin T, we denote by $\widehat{\mathbf{x}}^*_{T+h|T}$ the IMS forecast resulting from the wrong—assumed constant—model, as given by:

$$\widehat{\mathbf{x}}^*_{T+h|T} = \widehat{\mathbf{\Gamma}}^{\{h\}}\widehat{\boldsymbol{\gamma}} + \widehat{\mathbf{\Gamma}}^h \mathbf{x}^*_T,$$

where $\mathbf{\Gamma}^{\{h\}} = \sum_{i=0}^{h-1} \mathbf{\Gamma}^i$. Neglecting parameter estimation as a first analysis, we express the values with respect to the no-break (unobserved after T_0) process, $\mathbf{x}_t$, generated by (1). From (2), $\widehat{\mathbf{x}}^*_{T+h|T}$ decomposes into:

$$\widehat{\mathbf{x}}^*_{T+h|T} = \mathbf{\Gamma}^{\{h\}}\boldsymbol{\gamma} + \mathbf{\Gamma}^h \mathbf{x}^*_T = \left(\mathbf{\Gamma}^{\{h\}}\boldsymbol{\gamma} + \mathbf{\Gamma}^h \mathbf{\Gamma}^{\{k+1\}} \nabla \boldsymbol{\gamma}\right) + \mathbf{\Gamma}^h \mathbf{x}_T,$$

where the component $\mathbf{\Gamma}^h \mathbf{\Gamma}^{\{k+1\}} \nabla \boldsymbol{\gamma}$ reflects the misspecification effect in the forecasting model. Now, the actual realized value is: $\mathbf{x}^*_{T+h} = \mathbf{\Gamma}^{\{h\}}\boldsymbol{\gamma}^* + \mathbf{\Gamma}^h \mathbf{x}^*_T + \sum_{i=0}^{h-1} \mathbf{\Gamma}^i \boldsymbol{\epsilon}_{T+h-i}$, with resulting forecast error:

$$\widehat{\mathbf{e}}^*_{T+h|T} = \mathbf{x}^*_{T+h} - \widehat{\mathbf{x}}^*_{T+h|T} = \mathbf{\Gamma}^{\{h\}} \nabla \boldsymbol{\gamma} + \sum_{i=0}^{h-1} \mathbf{\Gamma}^i \boldsymbol{\epsilon}_{T+h-i}. \tag{5}$$

$\mathbf{\Gamma}^h \mathbf{\Gamma}^{\{k+1\}} \nabla \boldsymbol{\gamma}$ does not appear in (5) since it is the true post-break forecast origin which is used by the forecaster; yet the forecast is biased:

$$\mathsf{E}\left[\widehat{\mathbf{e}}^*_{T+h|T}\right] = \mathbf{\Gamma}^{\{h\}} \nabla \boldsymbol{\gamma}, \quad \text{and} \quad \mathsf{V}\left[\widehat{\mathbf{e}}^*_{T+h|T}\right] = \sum_{i=0}^{h-1} \mathbf{\Gamma}^i \mathbf{\Omega}_\epsilon \mathbf{\Gamma}^{i\prime} + 2 \sum_{i=0}^{h-1} \sum_{j=i+1}^{h-1} \mathbf{\Gamma}^i \mathbf{\Theta}_{|i-j|} \mathbf{\Gamma}^{j\prime}, \tag{6}$$

where, we define $\mathbf{\Theta}_m = \mathsf{E}\left[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}'_{t-m}\right]$ when we assume that $\{\boldsymbol{\epsilon}_t\}$ is covariance stationary. The DMS forecasts are:

$$\widetilde{\mathbf{x}}_{T+h|T} = \boldsymbol{\gamma}_h + \mathbf{\Gamma}_h \mathbf{x}^*_T, \tag{7}$$

which differ from the IMS only insofar as $\boldsymbol{\gamma}_h \neq \mathbf{\Gamma}^{\{h\}}\boldsymbol{\gamma}$ and $\mathbf{\Gamma}_h \neq \mathbf{\Gamma}^h$. Hence, analysis of their properties requires that we focus on model estimation. But beforehand, we can observe how intercept correction works.

2.2 Intercept Correction

A modeller could suspect that a break has occurred and intercept-correct the model to produce a more robust forecast. If she has no hypotheses about the date of the break and simply uses the IMS model, then the correction from the latest observation is $\boldsymbol{\delta}_{IC} = \mathbf{x}^*_T - \widehat{\mathbf{x}}^*_{T|T-1} = \nabla \boldsymbol{\gamma} + \boldsymbol{\epsilon}_T$, and, added to the IMS forecast, this yields the forecast error:

$$\widehat{\mathbf{e}}^{*IC}_{T+h|T} = \mathbf{x}^*_{T+h} - \left(\widehat{\mathbf{x}}^*_{T+h|T} + \boldsymbol{\delta}_{IC}\right) = \left(\mathbf{\Gamma}^{\{h\}} - \mathbf{I}_n\right) \nabla \boldsymbol{\gamma} + \sum_{i=0}^{h-1} \mathbf{\Gamma}^i \boldsymbol{\epsilon}_{T+h-i} - \boldsymbol{\epsilon}_T.$$

so that the forecasts are, now, unbiased for $h = 1$, but not so at longer horizons—unless $\mathbf{\Gamma} = \mathbf{0}_{n \times n}$ (forecasts unbiased at all horizons), or $\mathbf{\Gamma} = -\mathbf{I}_n$ (unbiasedness at odd horizons). The corresponding forecast error variance is given by:

$$\mathsf{V}\left[\widehat{\mathbf{e}}^{*IC}_{T+h|T}\right] = \sum_{i=0}^{h-1} \mathbf{\Gamma}^i \mathbf{\Omega}_\epsilon \mathbf{\Gamma}^{i\prime} + 2 \sum_{i=0}^{h-1} \sum_{j=i+1}^{h-1} \mathbf{\Gamma}^i \mathbf{\Theta}_{|i-j|} \mathbf{\Gamma}^{j\prime} + \mathbf{\Omega}_\epsilon - 2 \sum_{i=0}^{h-1} \mathbf{\Gamma}^i \mathbf{\Theta}_{|h-i|}.$$

The forecast errors of the IMS and IC models do not depend on the date of the break. So no information about this event is yet necessary; but if the modeller has some belief about whether and when this may have happened, she could try to take advantage of it and use multi-step methods. When using an intercept correction, with DMS forecasting:

$$\begin{aligned}\boldsymbol{\delta}_{DMSIC} &= \mathbf{x}_T^* - \widetilde{\mathbf{x}}_{T|T-h}^* = \mathbf{x}_T^* - \left(\boldsymbol{\gamma}_h + \boldsymbol{\Gamma}_h \mathbf{x}_{T-h}^*\right) \\ &= \left(\boldsymbol{\Gamma}^{\{k+1\}} - \boldsymbol{\Gamma}_h \boldsymbol{\Gamma}^{\{k-h+1\}}\right) \nabla\boldsymbol{\gamma} + \sum_{i=0}^{h-1} \boldsymbol{\Gamma}^i \boldsymbol{\epsilon}_{T-i},\end{aligned}$$

the forecast error results in:

$$\begin{aligned}\widetilde{\mathbf{e}}_{T+h|T}^{*DMSIC} &= \mathbf{x}_{T+h}^* - \left(\widetilde{\mathbf{x}}_{T+h|T}^* + \boldsymbol{\delta}_{DMSIC}\right) \\ &= \left(\boldsymbol{\Gamma}^{\{h\}} - \boldsymbol{\Gamma}^{\{k+1\}} + \boldsymbol{\Gamma}_h \boldsymbol{\Gamma}^{\{k-h+1\}}\right) \nabla\boldsymbol{\gamma} + \left(1 - L^h\right) \sum_{i=0}^{h-1} \boldsymbol{\Gamma}^i \boldsymbol{\epsilon}_{T+h-i}\end{aligned}$$

with variance

$$\mathsf{V}\left[\widetilde{\mathbf{e}}_{T+h|T}^{*DMSIC}\right] = 2\sum_{i=0}^{h-1} \boldsymbol{\Gamma}^i \boldsymbol{\Omega}_\epsilon \boldsymbol{\Gamma}^{i\prime} + 4\sum_{i=0}^{h-1}\sum_{j=i+1}^{h-1} \boldsymbol{\Gamma}^i \boldsymbol{\Theta}_{|i-j|} \boldsymbol{\Gamma}^{j\prime} - 2\sum_{i=0}^{h-1}\sum_{j=0}^{h-1} \boldsymbol{\Gamma}^i \boldsymbol{\Theta}_{|h+i-j|} \boldsymbol{\Gamma}^{j\prime}.$$

Whereas the variance of the intercept corrected forecast is larger than that of the IMS in (5), except for specific values of $\{\boldsymbol{\Theta}_m\}$, there exists a possibility that the bias may be lower. With known parameters, DMS and IMS coincide and $\boldsymbol{\Gamma}_h = \boldsymbol{\Gamma}^h$. Then

$$\boldsymbol{\Gamma}^{\{h\}} - \boldsymbol{\Gamma}^{\{k+1\}} + \boldsymbol{\Gamma}^h \boldsymbol{\Gamma}^{\{k-h+1\}} = \begin{cases} \mathbf{0}_n, & \text{if } h \leq k+1, \\ \boldsymbol{\Gamma}^{\{h\}} - \boldsymbol{\Gamma}^{\{k+1\}}, & \text{if } h > k+1. \end{cases} \tag{8}$$

so that the DMSIC forecast bias is zero as long as the forecast horizon is 'low' ($h < k+2$) and, even beyond $k+1$, it is lower than obtained by the other methods. This result obviously comes at the cost of higher variance, although not necessarily so, depending on the autocorrelation function of the disturbance process $\{\boldsymbol{\epsilon}_t\}$.

Now if we partition the error into stationary and integrated components, (8) implies that

$$\mathsf{E}\left[\left. \begin{array}{c} \widetilde{e}_{z,T+h|T}^{*DMSIC} \\ \widetilde{e}_{w,T+h|T}^{*DMSIC} \end{array} \right| x_T \right] = \mathbf{0}_n, \quad \text{if } h \leq k+1. \tag{9}$$

For $h > k+1$, in the stationary directions: $\boldsymbol{\Gamma}_z^{\{h\}} = \left(\mathbf{I}_r - \boldsymbol{\Gamma}_z\right)^{-1} \left(\mathbf{I}_r - \boldsymbol{\Gamma}_z^h\right)$ and

$$\boldsymbol{\Gamma}_z^{\{h\}} - \boldsymbol{\Gamma}_z^{\{k+1\}} = \left(\mathbf{I}_r - \boldsymbol{\Gamma}_z\right)^{-1} \left(\boldsymbol{\Gamma}_z^{k+1} - \boldsymbol{\Gamma}_z^h\right),$$

where, for $h > k+1$, $\left|\boldsymbol{\Gamma}_z^h\right| < \left|\boldsymbol{\Gamma}_z^{k+1}\right|$. As for the common trends, in $sp\left(\boldsymbol{\alpha}_\perp\right)$, the second line on the right-hand side of (8) becomes $\mathbf{I}_r{}^{\{h\}} - \mathbf{I}_r{}^{\{k+1\}} = (h-k-1)\,\mathbf{I}_r$. Asymptotically, for $\nabla\boldsymbol{\gamma}_z \neq \mathbf{0}_r, \nabla\boldsymbol{\gamma}_w \neq \mathbf{0}_{n-r}$ and $h > k+1$

$$\begin{aligned}\mathsf{E}\left[\left.\widetilde{e}_{z,T+h|T}^{*DMSIC}\right| x_T\right] &\to \begin{cases} \left(\mathbf{I}_r - \boldsymbol{\Gamma}_z\right)^{-1} \boldsymbol{\Gamma}_z^{k+1} \nabla\boldsymbol{\gamma}_z, & \text{if } h \to \infty, \text{ and } k \text{ constant}, \\ \mathbf{0}_r, & \text{if } k, h \to \infty, \end{cases} \\ \left\|\mathsf{E}\left[\left.\widetilde{e}_{w,T+h|T}^{*DMSIC}\right| x_T\right]\right\| &\to +\infty, \quad \text{if } h - k \to \infty, \ (k \text{ constant or not}).\end{aligned}$$

In stationary directions and for fixed k, the forecast bias is therefore zero for low horizon h, but converges to a fixed value as h increases; but this value is decreasing in k the time interval between the break date and the forecast origin. In the non-stationary directions, the forecast bias would explode to infinity with $(h-k-1)\nabla\boldsymbol{\gamma}_w$ if the model were never re-specified after the break had occurred.

The fact that (9) is valid—even asymptotically—is crucial for the interest in intercept correcting: this technique is beneficial when the time interval between the shift and the forecast origin is larger than the horizon.

We have shown that, in the context of known pre-break parameters, multi-step methods can be useful, but only as an ad hoc correction for the shift. We must analyze the case of estimated parameters to determine whether DMS can improve the accuracy of the forecasts. This is what we explore next.

3 Estimation in the presence of a location shift

The purpose of this section is to analyze how, in the presence of an intercept shift, the parameter estimators can be thought of as weighted averages of those obtained over the pre- and post-break subsamples.

3.1 Estimation over two regimes

In the framework above and in the context of an intercept shift, we assume that the parameters $(\boldsymbol{\gamma}^*, \boldsymbol{\Gamma})$ are estimated by ordinary least-squares. As we only focus, here, on location shifts and given that the slope is constant, we neglect to write explicitly the series in deviation from the sample means in the formula for the slope estimator which can, thus, be seen as a weighted average of two (pre- and post-break) regimes since:

$$\widehat{\boldsymbol{\Gamma}} = \left(\sum_{t=1}^{T} \mathbf{x}_{t-1}^{*} \mathbf{x}_{t-1}^{*\prime} \right)^{-1} \left(\sum_{t=1}^{T} \mathbf{x}_{t-1}^{*} \mathbf{x}_{t}^{*\prime} \right)$$

and the numerator can be decomposed into $\sum_{t=1}^{T_0-1} \mathbf{x}_{t-1}\mathbf{x}_t' + \sum_{t=T_0}^{T} \mathbf{x}_{t-1}^{*}\mathbf{x}_t^{*\prime}$. We define the relative sum of squares of the post-break period as

$$\boldsymbol{\lambda} \equiv \left(\sum_{t=1}^{T} \mathbf{x}_{t-1}^{*} \mathbf{x}_{t-1}^{*\prime} \right)^{-1} \left(\sum_{t=T_0}^{T} \mathbf{x}_{t-1}^{*} \mathbf{x}_{t-1}^{*\prime} \right), \tag{10}$$

so that, with $\widehat{\boldsymbol{\Gamma}}^{(\,)}$ denoting the pre-break slope estimator, using the sample up to $T_0 - 1$ and $\widehat{\boldsymbol{\Gamma}}^{(*)}$ using the remaining observations:[2,3]

$$\widehat{\boldsymbol{\Gamma}} = (\mathbf{I}_n - \boldsymbol{\lambda})\, \widehat{\boldsymbol{\Gamma}}^{(\,)} + \boldsymbol{\lambda} \widehat{\boldsymbol{\Gamma}}^{(*)}. \tag{11}$$

We let $\widehat{\boldsymbol{\Delta}}_{\Gamma} = \widehat{\boldsymbol{\Gamma}}^{(*)} - \widehat{\boldsymbol{\Gamma}}^{(\,)}$ so that $\widehat{\boldsymbol{\Gamma}} = \widehat{\boldsymbol{\Gamma}}^{(\,)} + \boldsymbol{\lambda}\widehat{\boldsymbol{\Delta}}_{\Gamma}$; if the expectation exists, then it is zero: $\mathsf{E}\left[\widehat{\boldsymbol{\Delta}}_{\Gamma}\right] = 0$, or as T and k tend to infinity (where $k = o(T)$),

$$\widehat{\boldsymbol{\Gamma}}^{(*)} - \widehat{\boldsymbol{\Gamma}}^{(\,)} \underset{T,k \to \infty}{\to} 0. \tag{12}$$

[2] The issue of the value of $\mathbf{x}_{t-1}^{*}$ at $t = T_0$, which is simply $\mathbf{x}_{T_0-1}$—thus corresponding to the pre-break regime, yet being used in the post-break estimator $\widehat{\boldsymbol{\Gamma}}^{(*)}$—is left aside since $\mathbf{x}_{t-1}^{*}$ is, then, the initial value of the second subsample. It is well known how initial observations affect estimation. Here, least-squares may not be preferred to maximum likelihood.

[3] Notice that this would also be the case in companion form, but then $\nabla\boldsymbol{\Gamma}$ would be non-zero.

From an estimate of the slope, the estimator of the intercept is given by $\widehat{\boldsymbol{\gamma}}^* = \{\overline{\mathbf{x}}^*\}_T^1 - \widehat{\boldsymbol{\Gamma}}\{\overline{\mathbf{x}}^*\}_{T-1}^0$,where the notation $\{\overline{\mathbf{x}}^*\}_T^1$ refers to the average of $\{\mathbf{x}_t^*\}$ over the sample $t = 1, ..., T$. We show in the appendix that

$$\mathsf{E}\left[\widehat{\boldsymbol{\gamma}}^*\right] \simeq \boldsymbol{\gamma}+\frac{k}{T}\nabla\boldsymbol{\gamma}. \tag{13}$$

The accuracy of this approximation actually depends on the process under consideration and whether the DGP coincides (with the exception of break omission) with the model used; we will return to this later. For now, $\widehat{\boldsymbol{\gamma}}^*$ decomposes into a sum of components amongst which is a weighted average of estimators from the two regimes (see the appendix):

$$\widehat{\boldsymbol{\gamma}}^* \simeq \frac{T-k}{T}\left(\mathbf{I}_n - \boldsymbol{\lambda}\right)\widehat{\boldsymbol{\gamma}}^{(\,)} + \frac{k}{T}\boldsymbol{\lambda}\widehat{\boldsymbol{\gamma}}^{(*)} \tag{14}$$

In univariate cases the weights corresponding to the intercept do not add up to one but to $1 - k/T\,(1+2\lambda) - \lambda = 1 - k/T - \lambda\,(1 - 2k/T)$. For instance, if $\lambda = k/T$, then the weights are $(1-\lambda)^2$ and λ^2 and their sum is lower than unity.

For stationary processes, letting

$$\varpi_k\left(\Gamma\right) = \frac{\left(1-\Gamma\right)^2}{1 - \dfrac{\Gamma\left\{\left(2+\Gamma\right)\left(1-\Gamma^{k-1}\right)+\Gamma^{2k-1}\right\}}{k\left(1-\Gamma^2\right)}},$$

the expectation of the weight is approximately:

$$\mathsf{E}\left[\lambda\right] \simeq \frac{k}{T}\left(1 - \left(\frac{k}{T} + \varpi_k\left(\Gamma\right)\frac{\sigma_x^2}{\left(\nabla\gamma\right)^2}\right)^{-1}\right).$$

Thus, the correction from k/T is roughly $\left(\nabla\gamma/\sigma_x\right)^2$, if the squared shift and variance are of similar order, or it is T/k if $\left(\nabla\gamma\right)^2 >> \sigma_x^2$.

When $\{x_t\}$ is integrated of order 1, with $\gamma = 0$, and if we assume that $k/T \to m$ as $T \to \infty$, then

$$\lambda \Rightarrow \frac{\int_m^1\left[W\left(r\right)\right]^2 dr}{\int_0^1\left[W\left(r\right)\right]^2 dr},$$

With the same assumption as before regarding the separability of the numerator and denominator in (10), and the existence of the expectation, we can derive an approximation to λ, where σ is defined such that $T^{-1/2}x_T \Rightarrow \sigma W\left(1\right)$,

$$\mathsf{E}\left[\lambda\right] \simeq \frac{2k}{T}\frac{\left(1 - \frac{k-1}{2T} + \frac{1}{6T}\left(\frac{\nabla\gamma}{\sigma}\right)^2\left(k+1\right)\left(2k+1\right)\right)}{1 + \frac{1}{T} + \frac{1}{3T^2}\left(\frac{\nabla\gamma}{\sigma}\right)^2 k\left(k+1\right)\left(2k+1\right)} = \frac{2k}{T} + O\left(\left(\frac{k}{T}\right)^2\right).$$

Finally, if $\{x_T\}$ exhibits a deterministic trend, then $\mathsf{E}[\lambda] \simeq \frac{3}{T}\left(1 + \frac{k(k-1)+4}{T}\right) + O\left(T^{-3}\right)$, so that in this case only, λ is not proportional to k/T. Yet, for all three types of processes, $\mathsf{E}[\lambda]$ is increasing in k.

3.2 Direct multi-step estimation

The case of multi-step estimation is very similar to one-step. Indeed, the estimators are defined as

$$\widetilde{\mathbf{\Gamma}}_h = \left(\sum_{t=h}^{T} \mathbf{x}^*_{t-h} \mathbf{x}^{*\prime}_{t-h} \right)^{-1} \left(\sum_{t=h}^{T_0-1} \mathbf{x}_{t-h} \mathbf{x}'_t + \sum_{t=T_0}^{T} \mathbf{x}^*_{t-h} \mathbf{x}^{*\prime}_t \right).$$

Thus, in this type of estimation, the size of the post-break subsample is k, irrespective of the horizon h, whereas the number of observations in the pre-break subsample decreases with h. The loss of information essentially occurs by dropping the first observations of the sample. The efficiency loss generally attributed to DMS estimation may not be so acute here, given that there is a focus on the more relevant observations. It is therefore possible to infer that

$$\widetilde{\mathbf{\Gamma}}_h = (\mathbf{I}_n - \boldsymbol{\lambda}_h) \widetilde{\mathbf{\Gamma}}^{(\ )}_h + \boldsymbol{\lambda}_h \widetilde{\mathbf{\Gamma}}^{(*)},$$

where $\boldsymbol{\lambda}_h$ will now depend on the relative values of h and k. For stationary processes, letting the effective sample $T_h = T - h + 1$:

$$\mathsf{E}[\lambda_h] \simeq \begin{cases} k/T_h, & \text{if } k \leq h, \\ \left(k + \dfrac{1 - k/T_h}{(\sigma_x/\nabla\gamma)^2 + 1/T_h} \right) /T_h, & \text{if } k = h+1, \\ \left(k + \dfrac{(k-h)(1-k/T_h)}{(\sigma_x/\nabla\gamma)^2 \varpi_{k-h}(\Gamma) + (k-h)/T_h} \right) /T_h, & \text{if } k > h+1, \end{cases} \tag{15}$$

and in the non-stationary cases, the same approximations as made for one-step estimation apply, with the restriction that the sample size is now T_h rather than T. Thus the λ_h coefficient is increasing in h; and λ_h is approximately proportional to $k/T_h = k/T + k(h-1)/T^2 + O(T^{-3})$. When $\{x_t\}$ is stationary, as seen in (15), $\lambda_h - k/T_h \in (0, (1-k/T_h)/T_h)$, when $h < k-1$. The weight attached—on estimation—to the post-break subsample is therefore larger when DMS methods are used. The same remarks as for one-step estimation prove also valid here (see the appendix), and hence

$$\widetilde{\boldsymbol{\gamma}}^*_h \simeq \frac{T_h - k}{T_h} (\mathbf{I}_n - \boldsymbol{\lambda}_h) \widetilde{\boldsymbol{\gamma}}^{(\ )}_h + \frac{k}{T_h} \boldsymbol{\lambda}_h \widetilde{\boldsymbol{\gamma}}^{(*)}_h, \tag{16}$$

and $\widetilde{\gamma}^*_h \simeq (1-\lambda_h)^2 \widetilde{\gamma}^{(\ )}_h + \lambda_h^2 \widetilde{\gamma}^{(*)}_h$ in the univariate case. Now that we have shown that the least squares estimators behave approximately as weighted averages of the pre- and post-break estimators, and that the weight attached to the post-break estimator is increasing in h, we can turn to the resulting implications for forecasting.

3.3 Comparing multi-step estimators

Given that we are interested in estimation of the models in order to produce forecasts, our parameters of interest are not the one-step ahead but the implied multi-step. Thus, assuming that $\widehat{\mathbf{\Gamma}}^{(\ )}$ and $\boldsymbol{\lambda}\widehat{\mathbf{\Delta}}_\Gamma$ commute, we can write, without loss of generality, that

$$\widehat{\mathbf{\Gamma}}^h = \left\{ \widehat{\mathbf{\Gamma}}^{(\ )} \right\}^h + h \left\{ \widehat{\mathbf{\Gamma}}^{(\ )} \right\}^{h-1} \boldsymbol{\lambda} \widehat{\mathbf{\Delta}}_\Gamma + O_p(\boldsymbol{\lambda}^2),$$

so that in terms of multi-step parameters, the IMS weight is $\boldsymbol{\lambda}_{\{h\}} = h \left\{ \widehat{\boldsymbol{\Gamma}}^{(\ )} \right\}^{h-1} \boldsymbol{\lambda}$. Again, we resort to a univariate processes to compare more accurately $\lambda_{\{h\}}$ and λ_h. In the non-stationary directions:

$$\left\{ \widehat{\Gamma}_w^{(\ )} \right\}^{h-1} = \left\{ 1 - \delta_{\Gamma_w}^{(\ )} \right\}^{h-1} = 1 - (h-1)\, \delta_{\Gamma_w}^{(\ )} + o_p\left(\delta_{\Gamma_w}^{(\ )} \right),$$

and $\lambda_{\{h\}} = h \left[1 - (h-1) \left(\iota_{\Gamma_w} + \delta_{\Gamma_w}^{(\ )} \right) \right] \lambda \simeq h\lambda$, except if the unit-root is inaccurately estimated as in the presence of, say, an omitted negative moving average component. The post-break subsample weight for the DMS estimator is therefore larger than that of the one-step ahead when the process is weakly stationary. By contrast, for $\left| \widehat{\Gamma}_z^{(\ )} \right| < 1$, $\lambda_{\{h\}}$ can be either larger or lower than λ. The value s_h, such that if and only if $\left| \widehat{\Gamma}_z^{(\ )} \right| > s_h$, then $\lambda_{\{h\}} > \lambda$ is given by $s_h = 1/\sqrt[h-1]{h} > 0.5$ (and s_h tends to unity as h increases). If $\left| \widehat{\Gamma}_z^{(\ )} \right| \simeq s_h$ then, depending on the other parameters, given that $\lambda_h > \lambda$, either of $\lambda_{\{h\}}$ and λ_h could be larger than the other.

The multi-step intercept is $\gamma_h^* \equiv \Gamma^{\{h\}} \gamma^*$. From the analysis concerning $\widehat{\Gamma}^h$, it seems clear that $\widehat{\gamma}_h^*$ is a mix of the four pre- and post-break intercept and slope estimators; it is thus likely to be badly biased. It is difficult to provide any general statement about the behaviour of this estimator, we therefore use Monte Carlo experiments in the following section to observe its pattern.

Comparing IMS and DMS slope estimates, the conclusion is that the weight attributed to the post-break subsample is larger in the iterated multi-step case. This IMS estimator is itself more accurate than that which is obtained by DMS when the model is well-specified in both subsamples. However, if the model is not congruent, DMS estimates are likely to prove more accurate. Similarly, the intercept estimator will be contaminated in the IMS framework to a larger extent than in the DMS.

4 Forecasting

The issue which concerns us here is whether it is preferable, when a shift has occurred, to forecast using estimates close to the new parameters or whether we would prefer robust estimates. We showed above that $\lambda_h > \lambda$—i.e. DMS estimators are less robust and closer to their post-break values. On the other hand IMS being obtained by non-linear combinations of one-step parameters, they are much more affected by the shift, this is reflected by the fact that, in general, $\lambda_{\{h\}} > \lambda_h$.

The answer to this puzzle is that anything potentially goes: depending on the problem under examination and the properties of the DGP, it may be preferable to be more or less robust. Assume, for instance, that the pre-break process is stationary, and that the corresponding estimates are close to their true population values. In this context, a strong location shift will be equivalent to a transitional unit-root, and it is preferable to have slope estimates close to unity for forecasting, which is the case with DMS. For a small break, the IMS estimator of the drift, being a combination of two inaccurate estimators (which, as we have seen, are closer to their pre-break values than their DMS counterparts) is more likely, than the DMS, to be badly biased (with respect to the shifted DGP), and hence, useless for forecasting.

When the process is I(1), breaks shift the estimated slopes towards positive values: if this improves the estimation of the unit-root, it proves beneficial for forecasting; but since an intercept shift creates a deterministic trend, it is essential, as in Chevillon and Hendry (2005), to accurately estimate it, and again, DMS tends to perform better then. In the case of negative shifts, it may yet be preferable to use robust estimates in small samples since, in general, the unit-roots are under-estimated and the forecast errors positive: not noticing the break may prove a more accurate forecasting strategy than using an estimated model with a low slope. A location shift may benefit DMS too, since it improves the estimation of the unit-root, which is a positive outcome since the underlying process has not changed and post-break forecasting uses a shifted forecast origin (see section 2). Yet, if the location shift implies estimated acceleration—i.e. if the slope estimator overshoots and settles above unity, which is the case when the process exhibits a deterministic trend in the same direction as the shift—then the forecasts will be very inaccurate. Similarly, a location shift in the direction opposite that of the trend would mimic either a unit-root or, as it seems more likely, stationarity, which may prove lethal to the forecasts.

In order to clarify these issues, we derive, below, forecast error taxonomies which decompose the forecast errors into their influential components.

4.1 A two regime taxonomy

We retain the assumptions made previously: the estimators are weighted averages of those obtained over two regimes. When analyzing a system of equations in vector form, there are no extraneous breaks—i.e. breaks in the system which do not occur in the equation of interest, but see Clements and Hendry (2004) for a formal definition—we, therefore, focus here for clarity on univariate processes. This allows us to see the essential features of the various types of forecasting methods. We thus assume that:

$$\begin{aligned}\widetilde{\Gamma}_h &= \left(1-\lambda_h^{\Gamma}\right)\widetilde{\Gamma}_h^{(\,)} + \lambda_h^{\Gamma}\widetilde{\Gamma}_h{}^{(*)} = \widetilde{\Gamma}_h^{(\,)} + \lambda_h^{\Gamma}\widetilde{\Delta}_{\Gamma_h},\\ \widetilde{\gamma}_h^* &= (1-\lambda_h^{\gamma})\,\widetilde{\gamma}_h^{(\,)} + \lambda_h^{\gamma}\widetilde{\gamma}_h^{(*)} = \widetilde{\gamma}_h^{(\,)} + \lambda_h^{\gamma}\widetilde{\Delta}_{\gamma_h},\end{aligned}$$

but we, now, allow λ_h^{γ} and λ_h^{Γ} to differ, and do not assume that they are positive. The h-step ahead IMS forecasts are, hence, given by:

$$\widehat{x}_{T+h|T} = \widehat{\Gamma}^{\{h\}}\widehat{\gamma}^* + \widehat{\Gamma}^h x_T^*.$$

Define the following shifts and biases: $\widehat{\delta}_{\gamma}^* = \gamma^* - \widehat{\gamma}^{(*)}$, $\widehat{\delta}_{\Gamma} = \Gamma - \widehat{\Gamma}$, $\widehat{\delta}_{\Delta_\gamma} = \nabla\gamma^* - \widehat{\Delta}_{\gamma}$, $\widehat{C}_i = \Gamma^i - \widehat{\Gamma}^i$, and $\widehat{C}^{\{i\}} = \Gamma^{\{i\}} - \widehat{\Gamma}^{\{i\}}$. From this, we can derive a taxonomy of the IMS forecast error, which we decompose, as in Table 1, into the eight components affecting it. By contrast, the DMS forecasts are such that the error $\widetilde{e}_{T+h|T} = x_{T+h} - \widetilde{x}_{T+h|T}$ is:

$$\widetilde{e}_{T+h|T} = \Gamma^{\{h\}}\gamma^* + \Gamma^h x_T^* + \sum_{i=0}^{h-1}\Gamma^i\epsilon_{T+h-i} - \widetilde{\gamma}_h^* - \widetilde{\Gamma}_h x_T^*,$$

and this leads to Table 2.

Comparing both IMS and DMS forecast error taxonomies allows to determine which elements may contribute to the success of either. First, as regards the presence of a regime change and its amplitude, the (i) components have exactly the same impact on the forecast errors. And these components are obviously zero if and only if there is no break and they, otherwise, draw the forecast error in the direction of the shift. The omission of the break in modelling translates into components (iia), which, again, behave in a similar fashion for both IMS and DMS, but, contrary to (i), with a negative coefficient. These components therefore partially offset (i) when

Table 1: Taxonomy of IMS forecast errors in the case of an intercept shift, with two regime approximation.

$\widehat{e}_{T+h\mid T} =$	$+\ \Gamma^{\{h\}}\nabla\gamma^{*}$	(i)	Regime change
	$-\ \lambda^{\gamma}\Gamma^{\{h\}}\nabla\gamma^{*}$	(iia)	Break omission
	$-\ (1-\lambda^{\gamma})\,\Gamma^{\{h\}}\widehat{\delta}_{\Delta_{\gamma}}$	(iib)	Break estimation
	$+\ \Gamma^{\{h\}}\widehat{\delta}_{\gamma}^{*}$	$(iiia)$	Intercept estimation
	$-\ \widehat{C}_{h}x_{T}^{*}+\widehat{C}^{\{h\}}\gamma^{*}-(1-\lambda^{\gamma})\,\widehat{C}^{\{h\}}\nabla\gamma^{*}$	$(iiib)$	Slope estimation
	$-\ \widehat{C}^{\{h\}}\widehat{\delta}_{\gamma}^{*}$	$(iiic)$	Slope/intercept estimation
	$+\ (1-\lambda^{\gamma})\,\widehat{C}^{\{h\}}\widehat{\delta}_{\Delta_{\gamma}}$	$(iiid)$	Slope/break estimation
	$+\ \sum_{i=0}^{h-1}\Gamma^{i}\epsilon_{T+h-i}$	(iv)	Error accumulation

Table 2: Taxonomy of DMS forecast errors in the case of an intercept shift, with two regime approximation.

$\widetilde{e}_{T+h\mid T} =$	$\nabla\gamma_{h}^{*}$	(i)	Regime change
	$-\ \lambda_{h}^{\gamma}\nabla\gamma_{h}^{*}$	(iia)	Break omission
	$-\ (1-\lambda_{h}^{\gamma})\,\widetilde{\delta}_{\Delta_{\gamma_{h}}}$	(iib)	Break estimation
	$+\ \widetilde{\delta}_{\gamma_{h}}^{*}$	$(iiia)$	Intercept estimation
	$+\ \widetilde{C}_{h}x_{T}^{*}$	$(iiib)$	Slope estimation
	$+\ \sum_{i=0}^{h-1}\Gamma^{i}\epsilon_{T+h-i}$	(iv)	Error accumulation

λ_{h}^{γ} or λ^{γ} are positive and even more so, the larger the weights are. This, in this case, benefits DMS since the coefficient increases with h. Yet, in absolute value, this component is lower for IMS. Break estimation is the only other common feature for both forecast error taxonomies which solely concerns the shift itself. If the break is under-estimated, then $\widetilde{\delta}_{\Delta_{\gamma_{h}}}$ is negative, and the contribution, positive and decreasing in the weight λ_{h}^{γ}. Potentially anything could happen, but because of the two-regime nature and the fact that in the case of an innovative intercept shift, the transition is smoothened, it seems likely that the break amplitudes could be underestimated. In this case, $\widetilde{\delta}_{\Delta_{\gamma_{h}}}$ and $\nabla\gamma^{*}$ have same signs, which implies that (iib) and (i) counterbalance one another. Then, the less robust the multi-step intercept estimator is (i.e. the larger λ_{h}^{γ}), the more (iia) offsets the shift, but the less (iib) does so. However, when the break is underestimated, a larger λ^{γ} (or λ_{h}^{γ}) means that (i) , (iia) and (iib) together are closer to zero. Indeed, the coefficient in $-(1-\lambda^{\gamma})$ in (iib) improves the forecast when λ^{γ} is lower, but this, in turn, worsens the forecast via (iia); if $\nabla\gamma^{*}$ is larger than $\widetilde{\delta}_{\Delta_{\gamma_{h}}}$ in absolute value, this is especially beneficial to DMS since, as we have seen, it provides more robust an estimator of the *multi-step* intercept and yet $\lambda_{h}^{\gamma}>\lambda^{\gamma}$.

As regard parameter estimation, first, $(iiia)$—the multi-step intercept estimation impact—has a similar structure for both IMS and DMS. It corresponds to *post-break* estimation. As previously noted, DMS should perform better on this, yielding a smaller absolute bias. The issue of whether the shift and intercept biases have same signs, which would mean that indeed DMS be more accurate in forecasting, actually depends on the size of the shift and on the horizon h. Yet, in a two regime framework, with positive weighting parameter λ^{γ}, it is the case. So, DMS performs better in this respect. The slope estimation component $(iiib)$ is the first, so far, which is noticeably different for IMS and DMS. Given that the IMS estimate of the multi-step intercept is a nonlinear combination of the one-step slope and intercept, the added IMS bias reflects this. The other part is similar for either method and again DMS, by minimizing the in-sample multi-step parameters does not suffer from powering up and should prove more accurate. The main issue is whether the added part is detrimental to IMS or not. This, in turn, depends on the specific series considered. For instance, a pure unit-root will be underestimated and hence if the—positive—intercept estimator is shifted upwards by the break, then (*iiib*) will be lowered for IMS. Chevillon and Hendry (2005), in a different setting, show that $\widehat{C}^{\{h\}}$ is of order $O_p\left(h^2\right)$ when the process is integrated, and DMS should, therefore, forecast better at larger horizons.

The next two components—(*iiic*) and (*iiid*)—are specific to IMS and again depend on the process and break. As noted earlier, a small positive shift tends to move the estimators in opposite directions, yielding a positive (*iiic*) which worsens the forecast; (*iiid*) counterpoises this if the slope estimator is shifted towards unity, but not if it is lowered. Finally, (*iv*)—the purely stochastic part of the forecast error, for which there is no hope of amelioration—is identical for both methods.

To conclude, DMS should perform better on average, if estimation leads to IMS and DMS biases of the same magnitude. The main contribution which could benefit IMS is the slope estimation—(*iiib*) and possibly (*iiic*) and (*iiid*). These properties also translate to the MSFE. Notice that the error accumulation (*iv*) has only zero expectation when a shift is indeed present.

If we conduct the same analysis of forecast errors, when using a method of intercept correction, we obtain, for instance, when a DMS estimation method is used with a multi-step correction:

$$\delta_{DMSIC} = \widetilde{\delta}_{\gamma_h} + \left(1 - \lambda_h^{\gamma}\right) \nabla \gamma_h^* - \left(1 - \lambda_h^{\gamma}\right) \widetilde{\delta}_{\Delta\gamma_h} + \widetilde{C}_h x_{T-h} + \widetilde{C}_h \Gamma^{\{k-h+1\}} \nabla \gamma^*$$

so that the forecast error is much simpler:

$$\widetilde{e}_{T+h|T}^{DMSIC} = \widetilde{C}_h \gamma_h - \widetilde{C}_h \left(1 - \Gamma_h\right) x_{T-h} + \left(1 + \widetilde{C}_h L^h\right) \sum_{i=0}^{h-1} \Gamma^i \epsilon_{T+h-i}.$$

Contrary to the no-estimation case from section 2, the forecast is biased here, but there are fewer components which contribute to the forecast error and the main feature of DMSIC forecasting is that the forecast bias is lower than for the other techniques provided that slope estimation is not too off the mark. Indeed, the intercept bias does not affect forecast accuracy.

We have attempted to highlight in this section the main factors which may contribute to an improved forecast accuracy by using techniques of direct multi-step estimation when the data generating process undergoes a location shift. We have shown that there is some potential for DMS in our framework. Before validating this by Monte Carlo simulation and an empirical application, we turn to some extensions of our models.

4.2 Extensions

Our analysis has, so far, focused on specific types of breaks and economic series. We now consider some additional cases and observe how the results are affected.

4.2.1 Additive breaks

We have modelled the location shift above via an innovative step dummy. We can also see whether the results are affected if the DGP shifts as:

$$\mathbf{x}^{**}_{T_0+i} = \mathbf{D}_0 + \mathbf{x}_{T_0+i}, \quad \text{for } i \geq 0. \tag{17}$$

where $\mathbf{x}_t$ is the underlying no-break process. In such a setting, the shift implies either a mean shift (in the case of a stationary $\mathbf{x}_t$) with an apparent transitional unit-root, or a trend break when $\mathbf{x}_t$ is integrated.

Given our framework with two subsamples, and since the second in chronological order contains always k observations because it is the pre-break sample whose size depends on h, the bias should reasonably be similar for both one-step estimation and h-step DMS. Yet, it is no longer possible to decompose the estimation into two regimes in the additive location shift framework since it implies that the pseudo-true parameters of the post-break VAR(1) change every period.

In univariate models, the slope estimator is shifted from $\widetilde{\rho}_h$ to $\widetilde{\rho}^*_h$ by a factor that depends on the underlying process. For stationary series, long but straightforward analytic decompositions show that:

$$\mathsf{E}\left[\widetilde{\rho}^*_h|\widetilde{\rho}_h\right] = \widetilde{\rho}_h + \frac{1}{T}\left\{h - (k+1)\left(\widetilde{\rho}_h + 1\right)\right\}\frac{D_0^2}{\mathsf{V}\left[x_t\right]} + O_p\left(T^{-2}\right)$$

whereas if x_t is integrated of order 1, $\mathsf{E}\left[\widetilde{\rho}^*_h|\widetilde{\rho}_h\right] = \widetilde{\rho}_h + \frac{6}{T^2\sigma^2_\epsilon}\left\{h - (k+1)\left(\widetilde{\rho}_h + 1\right)\right\} + O_p\left(T^{-3}\right)$, and if the data exhibit a linear trend as in $x_t = \beta t + \epsilon_t$, then

$$\widetilde{\rho}^*_h = \widetilde{\rho}_h + 6\frac{k+1}{\beta}\frac{D_0}{T^2} + O_p\left(T^{-2}\right).$$

For h large enough (and at any rate if $h > 2k$), the slope estimate is shifted toward positive values, and there can exist cases for large k such that the DMS slope estimator is lower than its no-shift value (except in the presence of a linear trend).

When a break which can be modelled as an additive location shift occurs, we have seen in section 2 that the forecasts, being generated *after* T_0 if $k > 0$, are unbiased when misestimation is neglected. But when it is taken into account, the differences between IMS and DMS become significant. Indeed, the parameter estimates are derived via techniques which react differently to breaks. In the case of stationary or integrated time series exhibiting a location shift, the slope estimator, $\widetilde{\rho}_h$, is shifted downwards for low h, and upwards for h larger than some $h^*_k \leq 2(k+1)$. For forecasting purposes, the DMS estimates are directly usable, but the one-step are non-linearly combined and powered-up in IMS. Hence, if $0 < \widehat{\rho}^*_1 < \rho < \widetilde{\rho}^*_h$, then $\left(\widehat{\rho}^*_1\right)^h$ is even lower than $\widehat{\rho}^*_1$ and the former can, in fact, be close to zero. In the limiting case of $\widehat{\rho}^*_1 = 0$ and $\widetilde{\rho}^*_h = 1$, the intercept estimates are given by:

$$\begin{aligned}\widehat{\tau}^*_1 &= \overline{x}^*_0,\\ \widetilde{\tau}^*_h &= \overline{x}^*_0 - \overline{x}^*_{-h}, \text{ for } h \geq 1,\end{aligned}$$

and their expectations, when $\mathsf{E}[\overline{x}_t] = 0$ without loss of generality, are $\mathsf{E}[\widehat{\tau}^*_1] = \dfrac{k}{T}D_0$ and $\mathsf{E}[\widetilde{\tau}^*_h] = \dfrac{k}{T_h}D_0$. The corresponding forecasting procedure are then:

$$\begin{aligned}\mathsf{F}_{IMS} &: \quad \widehat{x}_{T+h|T} = \frac{k}{T}D_0,\\ \mathsf{F}_{DMS} &: \quad \widetilde{x}_{T+h|T} = \frac{k}{T-h+1}D_0 + x^*_T.\end{aligned}$$

These stylized models emphasize the main differences between the two forecasting techniques: the multi-step intercept used by IMS is smaller than the DMS, and the multi-step IMS slope is lower than the DMS. IMS therefore takes less advantage of the process history than DMS and the latter provides a better assessment of the post-break mean. DMS can prove more accurate because it makes better use of the forecast origin, x_T^*, which is measured *after the break* has occurred. This method is, thus, quite close in practice to some form of intercept correction, re-centering the forecast on the latest available observation.

4.2.2 Blips

The breaks above were permanent and were statistically modelled by step dummies. It seems natural to ask what happens if, for instance, the shift is only temporary and could be represented by an impulse dummy variable. This is the limiting case of a process which is regularly subject to shifts or regime switches, and, for which, each of them can only be seen as being sustained for a short period.

In this framework, if a blip occurs at T_0, it will be made permanent in non-stationary directions and its effect will progressively vanish in the stationary ones. Moreover it will appear $\min\{p+1, k+1\}$ times in one-step estimation of an AR(p) (on the LHS at x_{T_0}, and on the RHS as lags) but only $\min\{p+2-h, k+2-h\}$ times in DMS estimation. Thus, if $p=1$ and $h \geq k+1$, the break is counted only once as a larger disturbance. The DMS estimates are therefore closer to their pre-break values. It is a desirable feature in the stationary directions since the shift impact is only temporary. In the integrated directions, our previous analysis (from subsection 4.1) still applies for the determinants of forecast accuracy, but in this case, the roles of IMS and DMS are now reversed since the latter technique is more robust than the former.

4.2.3 Multiple breaks

When the process undergoes multiple breaks at various dates, we can still draw on our previous framework. Depending on the time intervals between the breaks relative to the sample size, three main cases can be distinguished. When the shifts are few and the series have time to settle in between, we can assume that only the latest break matters and that the pre-break DGP is composite and made of past history; the results from subsection 4.1 then apply. When the breaks are more frequent but still occasional, the blip framework seems valid since, then, each shift can be seen as having only temporary effects in stationary direction; in the non-stationary directions,previous analyses still hold. Finally, when the shifts are very frequent, the series is better modelled as integrated and the breaks can be represented by a stochastic trend, potentially with some residual autocorrelation as in Chevillon and Hendry (2005).

5 Monte Carlo

This section records the results from Monte Carlo simulation of a univariate process undergoing a location shift represented by an innovative step dummy. Three main settings of ARMA processes were considered: stationary time-series, integrated or trend-stationary. For each experiment 5,000 replications were conducted using the Ox programming language for OxEdit and GiveWin 2. The model used for estimation was an AR(1) where we estimated the intercept and the slope over small samples of 26 observations (the initial observation at $t = 0$ was set to zero); we generated location shifts at dates ranging from $T_0 = 20$ to 25. The forecasting horizon varied between 1 and 6 periods, from a forecast origin at $T = 25$. This choice of framework allowed us to report the joint influence of the breaks and of various kind of potential misspecification in small samples. Thus, the DGP used for the Monte Carlo is:

$$\mathsf{DGP}: \quad x_t = \tau + 1_{\{t \geq T_0\}} \Delta\tau + \rho x_t + \phi t + \zeta_t + \theta \zeta_{t-1}, \tag{18}$$

where $\zeta_t \sim \mathsf{IN}(0,1)$. We, now, observe the relative accuracy of the IMS and DMS techniques for multi-step forecasting. As before, we let the forecast horizon and the post-break sample size vary and record the Monte Carlo estimator means for various types of series.

5.1 Stationary processes

Simulations show that the gain from DMS versus IMS is quite significant when $h < k$ and the amplitude of the shift is small and θ is negative. This gain is also decreasing in the AR coefficient ρ. Moreover, DMS outperforms IMS when the shift is very recent ($k = 0$ or 1 here). For large breaks and low absolute value of the non-zero moving average coefficient, DMS $\succ$ IMS (we use this symbol to denote a better forecast performance in terms of MSFE) at larger horizons, and even more so, when $\theta > 0$ rather than negative. Intercept corrections bring improvements when $k = h - 1$ and the break amplitude is low, in which case in can be the most accurate if $\theta < 0$. For other parameter combinations, IC performs better when the break is more recent (low k) and when there is little unmodelled residual autocorrelation.

From our simulations, we see that, as expected, DMS benefits from longer post-break sample sizes (k). Omitted negative moving averages and low slope coefficients also favour DMS. By observing the parameter estimates, it seems that our interpretation, in subsection 4.2.1, regarding the fact that DMS acts as some form of intercept correction by shifting the slope estimate towards unity is probably valid, since it can explain both the cases when DMS outperforms IMS and when it does not. The taxonomy of forecast errors which we presented above seems to represent in a satisfactory way the different impacts on the forecast accuracy, especially insofar as the gain from DMS is not monotonic in h. Indeed, as h increases relative to k and when $h > k$, the DMS slope estimate drops to zero, thus offsetting the previous gains. When comparing the relative forecast accuracy of IMS and DMS to that of intercept correction we find that if DMS $\succ$ IMS, IC mainly outperforms DMS for $k = 0$ or for small breaks and $k = h - 1$. Otherwise intercept correction tends to be the most accurate only if DMS is itself dominated by IMS.

In order to summarize the simulation results and validate our analysis, a response surface was fitted to 14,256 simulated combinations of parameters:

$$\begin{aligned}
\log\left(\frac{\mathsf{MSFE}_{DMS}}{\mathsf{MSFE}_{IMS}}\right) &= \underset{[0.0142]}{0.36} + \underset{[0.0054]}{0.17}\sqrt{h-1} + \underset{[0.0033]}{0.15}\sqrt{|\Delta\tau|} \\
&\quad - \underset{[0.0102]}{0.82}\,\mathsf{I}_{k=0} + \underset{[0.0102]}{0.41}\,\mathsf{I}_{k=h-1} - \underset{[0.0113]}{0.58}\,\mathsf{I}_{k=h} - \underset{[0.0105]}{0.79}\,\mathsf{I}_{k>h} \\
&\quad + \underset{[0.0057]}{0.11}\,\theta + \varepsilon_t, \\
R^2 &= 0.670;\ \widehat{\sigma} = 0.352;\ SC = -2.081,
\end{aligned} \tag{19}$$

were $\mathsf{I}_{k=0}$ denotes an indicator variable taking the value 1 if $k = 0$ and zero otherwise. We notice from this equation that the main factors benefitting DMS are $k = 0$ or $k \geq h$. From this response surface, we also see that DMS performs better relative to IMS in the presence of a negative moving average coefficient and for breaks of small amplitude. By contrast, at larger forecast horizons, the gain from using IMS increases. Finally, the intercept coefficient does not enter the equation, and this confirms our hypothesis of its irrelevance in the stationary case.

5.2 Integrated processes

Here, we assume that the DGP can be represented by an integrated ARMA$(1,1)$ process where $\rho = 1$ in (18). When the deterministic shift occurs at the forecast origin and is significantly larger than the intercept, the forecast accuracy gain from using DMS is not very large and necessitates that the drift be quite low: indeed for $\tau = .7$, IMS MSFE-dominates DMS. When $k > 0$, positive shifts are more beneficial to DMS than negative ones (this is probably specific to our set of Monte Carlo). When the drift coefficient increases, for DMS to present a forecast accuracy gain relative to IMS, the break date has to, somehow, relate to the drift; the larger τ, the larger k for which DMS $\succ$ IMS. For such matching (k, τ), the gain is increasing in the forecast horizon and decreasing in the moving-average coefficient θ. In an experiment where we set the drift to unity and the break, $\Delta\tau$, takes values 4, 6, 8 and 10, we observed that $\log\left(\frac{\mathsf{MSFE}_{DMS}}{\mathsf{MSFE}_{IMS}}\right)$ was non-monotonic in $\Delta\tau$: for instance when $k = 3$ and $h = 4$, it first decreases then increases with $\Delta\tau$; from the case $k = 4$, we saw that the minimum $\log\left(\frac{\mathsf{MSFE}_{DMS}}{\mathsf{MSFE}_{IMS}}\right)$ was achieved for a value of $\Delta\tau$ itself decreasing in k. Again, though, the larger the horizon is, or the lower θ, the larger the gain from using DMS. But it was often so that the post-shift slope estimator was above unity, thus proxying the trend with some acceleration. As regards IC, simulations showed that it is a method which requires large breaks to perform well, and even then, only so at short horizons.

Finally, for $\tau = 0$, when the break is both large and recent ($k = 1$ or 2), DMS tends to outperform IMS at high horizons, $h \geq 4$, essentially for $k = 1$ when $\theta < 0$ and for $k = 2$ when $\theta > 0$.

Although the results delineated above are understandable in the light of our preceding analyses, it must be admitted that it seems difficult to establish exactly what combination $(k, \tau, \Delta\tau)$ will *a priori* benefit DMS. A response surface presenting the key features was fitted to 30,096 simulated combinations of parameters, setting $\rho = 1$ ($\tau \in [0,1]$; $\theta \in (-1,1)$; $T = 25, 30$; $h = 1, ..., 4$; $k = 0, ..., 5$; $\Delta\tau \in [-10, 10]$):

$$\begin{aligned}
\log\left(\frac{\mathsf{MSFE}_{DMS}}{\mathsf{MSFE}_{IMS}}\right) &= \underset{[0.0098]}{-0.08} - \underset{[0.0103]}{0.44}\,\mathsf{I}_{\Delta\tau=0} - \underset{[0.0120]}{0.19}\,\mathsf{I}_{\Delta\tau<0} - \underset{[0.00090]}{0.025}\,\Delta\tau \\
&\quad + \underset{[0.0068]}{0.19}\,\mathsf{I}_{k=h-1} - \underset{[0.0077]}{0.20}\,\mathsf{I}_{k>h+1} + \underset{[0.0040]}{0.20}\,\sqrt{k} + \underset{[0.0043]}{0.30}\,\sqrt{h-1} \\
&\quad - \underset{[0.0071]}{0.17}\left(\tau - \frac{1}{\sqrt{T}}\right)^2 + \underset{[0.0023]}{0.25}\left\{(h-1)\,\theta + \theta^2\right\} + \varepsilon_t, \\
R^2 &= 0.48;\ \widehat{\sigma} = 0.423;\ SC = -1.717,
\end{aligned} \tag{20}$$

were $\mathsf{I}_{k=0}$ denotes an indicator variable taking the value 1 if $k = 0$ and zero otherwise. The fit was not very good ($R^2 = 0.48$) but the equation is useful to understand average behaviours. The constant is negative and factors benefitting DMS are an omitted negative moving average (impact increasingly significant with h), the presence of a drift close to $1/\sqrt{T}$. Recent breaks (the $\sqrt{k}$ component) are also beneficial to DMS, with an additional effect related to the forecast horizon, namely that it is preferable to forecast using direct estimation at an horizon $h < k - 1$. Finally, the effect of the break amplitude is nonlinear: negative breaks of small amplitude and large positive breaks favour DMS

6 Forecasting the South African GDP

In order to complement our previous analysis, we conduct an empirical analysis of 31 competing forecasting methods in the presence of structural breaks. Our interest lies in assessing the multi-step model which was developed by Aron and Muellbauer (2002) for forecasting the South African GDP. Contrary to Marcellino, Stock, and Watson (2005) who differenced all the integrated series to ensure that they were working with stationary variables, we focus on accuracy in forecasting the level of the variable since differencing corrects for some location shifts at the cost of higher variance. Thus, Aron and Muellbauer established a relationship which allowed them, in spite of the many breaks experienced by the economy, to forecast the annual change in the quarterly GDP series over the last twenty years. Because it directly targets the endogenous variable several periods ahead of the forecast origin, their method belongs to the class of direct multi-step forecasts. Unfortunately, Aron and Muellbauer were unable to assess the accuracy of their model by comparing it to alternative techniques and we proceed to this below.

6.1 Background

6.1.1 Thirty years of breaks

South Africa has undergone a profound transition in the last thirty years and, hence, any model of its economy would be subjected to frequently occurring breaks. From 1976 and with the government's policy of Apartheid, the country began suffering from increasing international isolation which culminated, between late 1985 and the 'free' elections of 1994, by a period of almost no access to international capital. These factors, combined to the high degree of reliance of the economy to mineral exports might explain some of the shocks and large variations observed in the economic variables (see the articles by Aron and Muellbauer for an extended analysis of the South African context).

Following Aron and Muellbauer, we can distinguish three main monetary regimes since the 1960s. Until the late 1970s, there existed quantitative controls on interest rates and credit and the main criteria used for monetary policy were liquid asset ratios, while the corrective effect of interest rates was largely neglected by the regulatory authorities. Financial liberalization and transition towards a more flexible, cash-reserves based, system took place over the first half of the 1980s. From 1986 onwards, the monetary authorities made use of the discount rate to influence the interbank overnight refinancing market in order to achieved pre-declared monetary targets. The credit growth which followed the financial liberalization soon lessened the usefulness of monetary targets, thus leading, from 1998, to a new regime where the South African Reserve Bank (SARB) offers some amount at the daily tender for repurchase transactions, thus signalling its preferences on short-term interest rates via an auction mechanism. In early 2000, inflation targeting was reinstated as part of the medium-term monetary objectives.

Following the developments of monetary policy, the foreign exchange market experienced various regimes pointing towards greater flexibility. From US dollar or pound Sterling pegs combined with restrictions on resident and nonresident capital flows, the system moved, in 1979, to a regime of dual currency. Most non-resident transactions operated at the floating 'financial' exchange rate, while a 'commercial' rate was instated and announced in line with market forces. The latter became market determined in 1983 and the dual rates were soon re-unified. A debt crisis and the collapse of the Rand provoked a return to the dual currency system after 1985. In 1995, unification of the dual currency was initiated under a managed rate which has become fully floating at the introduction of inflation targeting.

6.1.2 The AM model

Aron and Muellbauer developed a model for forecasting the annual change of the South African real GDP (Y, or y in log) via a solved-out equilibrium correction equation which depends on a set of variables $\{X_i\}$:

$$\Delta_4 y_{t+4} = \gamma \left(\alpha_0 + \mu_t + \sum_{i=1}^{n} \alpha_i X_{i,t} - y_t \right) + \sum_{i=1}^{n} \sum_{s=1}^{k} \beta_{i,s} \Delta X_{i,t-s} + \epsilon_t,$$

where ϵ_t is assumed white noise, but may be modelled by some moving average component and μ_t is a smooth $\mathsf{I}(2)$ stochastic trend which aims to capture the underlying production capacity of the economy. AM find a stable equation using as regressors the real prime interest rate (*RPRIME*) and its annual change, the ratio of current account surplus to current GDP (*RCASUR*), the government surplus to GDP ratio (*RGSUR*), the long-term growth of terms of trade (*TOT*), the difference in a financial liberalization indicator (a spline indicator variable *FLIB*), a monetary regime shift dummy (for the 1983(2)–85(4) period, denoted by N) interacting with *RPRIME* and its difference, and finally an indicator dummy variable for the 1991(3)–92(2) drought (*DUM92*).

To assess the forecasting power of this equation, we develop alternative techniques in the following section. But several dummy variables which enter the equation could only be defined ex-post; hence, for ex-ante forecast comparisons, these should be omitted from the equation (at least *DUM92*, and potentially N and *FLIB*, although these could be kept since they were 'predictable' from the government's decisions). Finally, there remains the issue of the $\mathsf{I}(2)$ trend: indeed, for computational ease, we do not estimate the state-space model for each subsample but store the whole-sample trend as it was estimated by STAMP and use it as a regressor, assuming, as seems likely, that its value at some date $t < T$ from an estimation over 0,..., T does not strongly differ from the value obtained at t from estimation over $0, ..., t$. However, in order to avoid this issue, we also replace it, in an alternative model, with a deterministic trend.

Thus, the AM technique and its variants are not seen as a well-specified model of the process generating the South African GDP, but rather as a misspecified DMS forecasting tool.

6.2 Competing forecasts

The forecasting equation developed by Aron and Muellbauer includes a few exogenous variables which ought to be modelled for a proper IMS technique to be used. Unfortunately, doing so would increase the degree of misspecification, which would, in turn, be detrimental to our assessment of forecast accuracy. Hence, we resort to two simpler multivariate models which, once solved out, should provide the possibility for both IMS and DMS forecasting. The first model follows a small monetary system for South Africa, the second uses the main variables of the AM equation.

Another issue arises regarding the variable to be forecast: the AM technique provides forecasts of $\Delta_4 y_{T+4}$ from an end of sample T. This annual difference does not fit the definition of the IMS forecast as usually referred to. Indeed, if a model in differences provides some $\Delta\widehat{\mathbf{x}}_{T+1} = \widehat{\mathbf{B}}\Delta\mathbf{x}_T$, with $y_t = \mathbf{P}\mathbf{x}_t$, then the iterated forecast is given by $\Delta\widehat{\mathbf{x}}_{T+4} = \widehat{\mathbf{B}}\Delta\widehat{\mathbf{x}}_{T+3} = \widehat{\mathbf{B}}^4\Delta\mathbf{x}_T$, and hence $\Delta_4\widehat{\mathbf{x}}_{T+4} = \widehat{\mathbf{x}}_{T+4} - \mathbf{x}_T = \left(\sum_{i=1}^{4}\widehat{\mathbf{B}}^i\right)\Delta\mathbf{x}_T$, so that the iterated estimator becomes complex. Indeed, if, as for AM, the model is $\Delta\widehat{\mathbf{x}}_{T+1} = \widehat{\mathbf{A}}\mathbf{x}_T$, then:

$$\Delta\widehat{\mathbf{x}}_{T+2} = \widehat{\mathbf{A}}\widehat{\mathbf{x}}_{T+1} = \widehat{\mathbf{A}}\left(\Delta\widehat{\mathbf{x}}_{T+1} + \mathbf{x}_T\right) = \widehat{\mathbf{A}}\left(\widehat{\mathbf{A}} + \mathbf{I}\right)\mathbf{x}_T. \tag{21}$$

Thus, as this very choice of target seems to benefit direct multi-step estimation, we assume that the forecasts to be evaluated concern the levels of the GDP and not the annual differences. Hence from the AM equation $\Delta_4\widehat{y}_{T+4} = \widehat{c}_1 y_T + \widehat{\mathbf{C}}\mathbf{x}_T$, we retrieve $\widehat{y}_{T+4} = (1+\widehat{c}_1)\, y_T + \widehat{\mathbf{C}}\mathbf{x}_T$.

6.2.1 A small monetary model

The variables included in our model comprise the M1 narrow money aggregate (denoted by M), the consumer price index (CPI), the 3–month treasury bill interest rate (per annum, R) and the South African Rand/US dollar exchange rate which were obtained from the International Financial Statistics database provided by the International Monetary Fund. Unfortunately, the IFS series for the real Gross Domestic Product do not seem reliable and hence were discarded and we used data provided by Aron and Muellbauer. Following Jonsson (1999) we conduct a cointegration analysis of five variables:[4,5] the log of real narrow money $m - cpi$, the log of real exchange rate rer, the log of real GDP y, the nominal treasury interest rate R and inflation Δcpi. Augmented Dickey-Fuller tests with 4 lags do not reject the hypothesis that these variables are all integrated of order 1. Real narrow money experienced large contractions in the late 1970s and mid 1980s during periods of higher inflation and the continuous depreciation of the Rand with respect to the US dollar. In addition to the five economic variables, we allow for a constant and a trend to enter the cointegration space. We restrict our attention to a VAR(2) as tests showed that lags beyond 2 are not significant. The VAR in levels, estimated over the 1966(1)–2001(2) period, seems to fit the data reasonably well as shown in figure 1. Indeed, except for inflation, the vector of variables is well explained.

FIGURES 1 and 2 ABOUT HERE

A corresponding test summary is presented on Table 3 which records statistical information about the VAR, namely the equation residual standard errors ($\widehat{\sigma}$); single-equation evaluation statistics for no serial correlation (F_{ar}, against 5th-order residual autoregression); no ARCH (F_{arch}, against fourth-order); no heteroscedasticity (F_{het}, see White, 1980); and a test for normality (χ^2_{nd}, see Doornik and Hansen, 1994). Analogous system (vector) tests are labelled as ^{v} and, finally, * and ** denote significance at, respectively, the 5 per cent and 1 per cent levels. Hence the Normality tests fail here for most of the variables and the system; this reflects what can be seen from the graphs on the third column of figure 1, namely the presence of large outliers. Given our interest in

[4]The variables modelled by Jonsson (1999) were the interest rates, real income, exchange rate, broad money and prices.

[5]Computations and tests were conducted using PcGive and GiveWin.

breaks, and noticing that, but for these, the densities seem close to Normal, we hence retain our model.

TABLES 3 and 4 ABOUT HERE

Finally, we observe the significant aspect of parameter constancy. Figure 2 records the 1↑ and N↓ (BreakPoint) Chow constancy tests for the individual equations and the VAR system (see Chow, 1960), together with the 1% boundary. We can be reasonably confident in the overall constancy of the relations, although they exhibit occasional breaks and we therefore retain this system for further analysis. Assuming, now, that the VAR is reasonably specified, we would normally investigate the cointegration properties of the five variables. The cointegration statistics reported in table 4—where a constant and a trend entered the cointegration space respectively unrestrictedly and restrictedly—support the hypothesis that there are two cointegrating relations (see Johansen, 1996).

Unfortunately, as shown by Clements and Hendry (1999), the use of cointegrating relationships which experience breaks tends to worsen the accuracy of the forecasts. Given that South Africa provides such an example—indeed, when estimating the model over various subperiods, the trace statistic provides justifications for varying numbers of cointegrating vectors—we decide to exclude the cointegrating vectors from the VAR. Hence the mapping of the VAR in levels to a parsimonious stationary VAR consists simply in differencing. This provides the first of the two IMS forecasting models.

6.2.2 An IMS version of AM

To analyze the properties of the *multi-step* AM forecasting procedure, we develop below an IMS equivalent. It must be noticed that the parsimonious version of the AM only uses few variables: besides the GDP itself and dummies, the regressors entering the equilibrium correction mechanism are *RPRIME* and its differences, *RCASUR*, *RGSUR*, the log of *TOT* and an I(2) trend (see definitions in section 6.1.2 above). Hence it seems natural to develop a VAR model which includes these variables. As regards the *N* and *FLIB* indicators, they could be included or not, according to how acute the forecaster's perception of the economic environment could be assumed to have been at the time. Finally, concerning the I(2) trend, which seems an essential element of the model, several strategies can be envisaged: for simplicity and computational ease when forecasting from a past origin, we resort to either omitting it altogether and replacing it with a deterministic trend, or to storing the whole-sample estimated trend and using it as a regressor.

Augmented Dickey-Fuller tests reject the presence of a unit root in the *RCASUR* and *RGSUR* series, but fail to reject it as regards y, *RPRIME* and *TOT*. Thus a cointegration analysis should include the latter three variables, and a trace test for reduced rank claims the existence of one cointegrating relationship between the three, when a trend is allowed to enter the cointegration space. By contrast, if we allow for *N* and *FLIB* to enter unrestrictedly in the system, then we only marginally reject the hypothesis that the matrix is of full rank and hence that the variables do not cointegrate (*p-value* of 4%, as given by PcGive from estimation over 1960(3)–2000(2), for a VAR(2) model—the lags beyond 2 being statistically insignificant). This small model allows us to generate IMS—and DMS—forecasts which we can compare to the solved out AM equation.

6.2.3 Univariate methods

We follow the results from the research by Clements and Hendry and use alternative forecasting techniques. These essentially belong to one of two classes: those of differencing or of intercept correcting. Thus, if we wish to forecast y_{T+4} from a forecast origin T, it has been shown that the models:

$$\mathsf{DV}: \quad \Delta y_t = \zeta_{1t}; \quad \mathsf{DDV\_IMS}: \quad \Delta\Delta y_t = \zeta_{2t}; \quad \mathsf{DDV\_DMS}: \quad \Delta_4\Delta_4 y_t = \zeta_{4t}$$

where the ζ_{it} are assumed independent white noises, can exhibit some degrees of robustness to breaks. They respectively lead to forecasts $\widehat{y}_{T+4}$ given by: $\widehat{y}_{T+4}^{\mathsf{DV}} = y_T$, $\widehat{y}_{T+4}^{\mathsf{DDV\_IMS}} = y_T + 4\left(y_T - y_{T-1}\right)$, and $\widehat{y}_{T+4}^{\mathsf{DDV\_DMS}} = y_T + \left(y_T - y_{T-4}\right)$.

By contrast, intercept correcting constitutes an adjustment to an existing model, such that if $\widehat{y}_{T+4} = \widehat{\Psi}^4 y_T$ and $\widetilde{y}_{T+4} = \widetilde{\Psi}_4 y_T$ are forecasts from the IMS and DMS models, the intercept corrected forecasts become:

$$\begin{aligned} \mathsf{IMSIC} \quad &: \quad \widehat{y}_{T+4} = \widehat{\Psi}^4 y_T + \left(y_T - \widehat{\Psi}^4 y_{T-4}\right), \\ \mathsf{DMSIC} \quad &: \quad \widetilde{y}_{T+4} = \widetilde{\Psi}_4 y_T + \left(y_T - \widetilde{\Psi}_4 y_{T-4}\right). \end{aligned}$$

IMSIC could alternatively be defined as $\widehat{\Psi}^4 y_T + \widehat{\Psi}^4\left(y_T - \widehat{\Psi} y_{T-1}\right)$, or variations thereof, but these are very unlikely to improve the accuracy and are, hence, left aside.

6.3 Forecast comparison

6.3.1 Techniques

We proceed to a comparison of ex-ante forecast accuracy as resulting from the various methods delineated above. These amount to 31 techniques and are labelled as follows: the three 'difference' operator forecasts are DV, DDV_IMS and DDV_DMS. We, then, use six models derived from the AM framework, these are the original AM as defined previously, AM_trend where the stochastic I(2) trend is replaced with a deterministic one, AM_noDUM92 where the 1992 drought is not accounted for, AM_noDUM92_trend which combines both features from the previous two techniques; and finally, the last two techniques ignore *FLIB*, *DUM92* and *N* altogether: AM_nodum and AM_nodum_trend (dealing as before with the stochastic trend). In order to assess the forecasting power of the small monetary model which we briefly analyzed above, three additional techniques were used to produce forecasts using IMS, DMS and IMSIC procedures, we refer to them as M1_IMS, M1_DMS and M1_IC, and to their equivalent from estimating the model in difference as DM1_IMS and DM1_IC—see (21)—noticing that DMS was not computed here, as it would have involved estimating and combining models at all horizons between 1 and 4. Moreover, the VAR model derived from AM was also used, with the dummy variables in levels and with or without them in differences, thus yielding the seven forecasts VARAM in levels, DVARAM in differences, with suffices as before. We also computed forecasts from the M1 and VARAM suites, estimated in levels, adding a deterministic trend as a regressor(suffix _trend). Finally, two pooled forecasts from the 6 and 15 overall most accurate methods were also included (Pool 6 and Pool 15).

6.3.2 Forecast accuracy

We present the empirical mean-square forecast errors and derive modified Diebold–Mariano test statistics (see Diebold and Mariano, 1995, and Harvey, Leybourne, and Newbold, 1997 who allow for errors to follow a moving average, as in multi-step forecasting) for testing the equality of the MSFEs over various subsamples of the data.

First, as regards overall forecast accuracy, we notice on table 5 that the original AM method is the most accurate as it yields the—statistically significantly—lowest MSFE. An interesting feature for direct multi-step estimation lies in the overall accuracy of the AM variants: indeed, that with the lowest ranking in the forecasting exercise is still more accurate (statistically significantly for the best five, see the appendix for the statistics) than the best element from any other class of techniques, even when we remove the dummies from the model and replace the stochastic trend with a deterministic. The next best forecast is obtained by using the small monetary VAR in differences and in levels. This is reassuring since it shows that the purely statistical techniques (DV and DDV) are dominated, although they rank next and before the VARAM model in levels or in differences when we keep the dummies. We notice also that, as regards VAR forecasts (VARAM and M1), the models in differences perform better than the levels; this is in line with the results by Clements and Hendry. In terms of iterated versus direct forecasting, we notice that, when the technique is very inaccurate, DMS is less biased than IMS (see VARAM), but the converse is true when the accuracy of either seems reasonable (see M1). As for IC, it tends to be close to—yet worse than—IMS.

Table 5 also provides equivalent statistics as computed over smaller samples. Following subsection 6.1.1, we split the dataset into two periods which exhibit different features: in the 1973(4)–1986(4) subset, the South African economy went through many changes and the breaks were rather frequent, whereas from 1986 to 1995, the country was not very involved in the international economy and the system (i.e. the legal-political environment) did not evolve as fast (which is not to say that the economy did not suffer, say from the 1992 drought). From 1995 to 1998, after the democratic elections, the economic environment (banking and financial sectors) was relatively stable and deregulation took place again afterwards. We, thus, find that forecast accuracy from the econometric models (M1 and VARAM) is improved in the second period, whereas the statistical techniques tend to perform better in the first. Finally, we find that M1 is more accurate in differences in the whole sample and in the first–less stable–period but that the levels are more accurate in the last era (and can even rival some variants of AM) .

From table 5 we see that direct multi-step estimation computed from a dedicated model such as that derived by Aron and Muellbauer is by far the most accurate technique for this dataset. Econometric models come second followed by statistical ones. DMS estimation of a misspecified VAR (M1, here) does not necessarily improve accuracy. Finally the VAR model which uses the same variables as AM did not perform so well for two reasons: first, the lag length is necessarily reduced by multivariate estimation, and then we did not include a deterministic trend in the VAR in levels. Doing otherwise would have improved accuracy as we show in table 6, but it would not have altered our conclusions.

Indeed, table 6 presents the empirical MSFEs for the same samples as above when a deterministic trend is also used as a regressor, thus providing eight new forecasts. We notice that, as regards the overall accuracy, although the VARAM variants now perform better than beforehand—whether in levels or differences—they yet do not match AM and its derived techniques. In fact, adding a trend improve the accuracy of all models in the first subsample, but in the second it deteriorates that of M1, which was almost on a par with AM. Hence the trend renders the multi-step forecasts more robust in the unstable era as expected from our analysis in the previous sections.

Now, we present on table 7 the empirical MSFEs of the pooled forecasts of the best 15 and 6 estimators. The pooled estimate of the AMs, DM1s, and M1s with and without a trend, which is denoted by Pool 15 performs worse than any of the AMs (except for AM_nodum in the last subperiod) but is more accurate than all the others. Moreover, its empirical MSFE is close in both subperiods. If we restrict pooling to the 6 AMs, then the average forecast ranks third overall and in both subsamples.

TABLES 5 TO 7 ABOUT HERE

6.4 Conclusions to the empirical forecasting exercise

The direct—misspecified—equation derived by AM has impressive forecasting power, whether in troubled periods or more quiet eras. As regards DMS forecasting from alternative models in levels, the results are more mixed: it is hard to recommend one of IMS, DMS or IC as a rule, as their respective performance rankings alter. However, although direct estimation tends not to lead to be best forecast, it does not generally provide the worst either; and it presents the interesting feature of 'decent' accuracy almost everywhere (which is not the case for the other two as VARAM witnesses on table 5).

Direct multi-step estimation is therefore a technique which has the interesting property of being occasionally very close, rarely very far from its target; and its accuracy is increasing in the unpredictability of the economic variable (via regime and deterministic instability). Contrary to Marcellino, Stock, and Watson (2005) who focused on stationary transforms of U.S. macroeconomic series and found no evidence benefitting DMS, we see here that this technique can prove efficient when used for forecasting non-stationary variables in unstable economies, such as those of emerging countries. Yet the performance of the AM variants does advocate for a mix of economic analysis and robust forecasting techniques.

7 Conclusions

The aim of this paper was to show that, when economic series undergo location shifts, there exists a rationale in using direct multi-step estimation for forecasting purposes. We have demonstrated that *recent* breaks matter and not only if they occur at the forecast origin. Indeed, as the post-shift sample size alters, so do the forecasting properties of DMS and IMS. Depending on which of the inclusion of an innovative or of an additive dummy variable is the adequate model for the break, we saw that the DMS technique provides more or less robust estimates. We also showed how this can be included in a forecast error taxonomy which proves a convenient tool for the analysis of forecast accuracy. In general, the main contribution of direct multi-step estimation is to be found insofar as it takes better account of the inter-temporal dependence of the series on its own history. Thus, in the context of a location shift, DMS may act as an intermediate strategy between congruent dynamic specification and intercept correction, therefore proving most valuable when the shift has occurred recently. This analysis sheds light on why the multi-step estimation literature has found that DMS seems a good strategy for forecasting long-memory processes such as represented by Fractionally Integrated ARMA models. One of the underlying strengths of DMS lies, once again, in its capacity to provide more accurate estimates of the relevant multi-step dynamic—essentially autoregressive—parameters. And, its main flaw is, therefore, in its less precise appreciation of the mean of the process. It, hence, proves an interesting technique when its drawbacks are overshadowed by its advantages, as in this paper.

Finally, concerning the issue of whether it might be preferable to model the shift, it should be stressed that breaks (or mismeasurements) which occur close to the end of the sample can be difficult to detect, and even more so to date. Intercept correction is one "model" of such end of sample breaks and this was examined above; more generally ICs are known to perform well mainly in the absence of measurement errors and do less well if the series exhibit unmodelled autocorrelation or in-sample breaks. Hence, congruent modelling may seem an utopian aim for current and recent data, and, given the frequent empirical occurrence of shifts, DMS can prove a less ambitious—if more mundane—yet efficient tool for forecasting.

References

Aron, J. and J. Muellbauer (2002). Interest rate effects on output: evidence from a GDP forecasting model for South Africa. *IMF Staff Papers* **49**, 185–213.

Bhansali, R. J. (1996). Asymptotically efficient autoregressive model selection for multistep prediction. *Annals of the Institute of Statistical Mathematics* **48**, 577–602.

Bhansali, R. J. (1997). Direct autoregressive predictors for multistep prediction: order selection and performance relative to the plug-in predictors. *Statistica Sinica* **7**, 425–449.

Chevillon, G. and D. F. Hendry (2005). Non-parametric direct multi-step estimation for forecasting economic processes. *International Journal of Forecasting* **21**, 201–18.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica* **28**, 591–605.

Clements, M. P. and D. F. Hendry (1996). Multi-step estimation for forecasting. *Oxford Bulletin of Economics and Statistics* **58**, 657–683.

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

Clements, M. P. and D. F. Hendry (2004). Pooling of forecasts. *Econometrics Journal* **7**, 1–31.

Cox, D. R. (1961). Prediction by exponentially weighted moving averages and related methods. *Journal of the Royal Statistical Society* **B 23**, 414–422.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**, 253–263.

Doornik, J. A. and H. Hansen (1994). An omnibus test for univariate and multivariate normality. Discussion paper, Nuffield College.

Findley, D. F. (1983). On the use of multiple models for multi-period forecasting. *Proceedings of Business and Economic Statistics, American Statistical Association*, 528–531.

Findley, D. F., B. M. Potscher, and C. Z. Wei (2004). Modeling of time series arrays by multistep prediction or likelihood methods. *Journal of Econometrics* **118**, 151–187.

Harvey, D., S. Leybourne, and P. Newbold (1997). Testing the equality of prediction mean squared errors. *International Journal of forecasting* **13**, 281–291.

Haywood, J. and G. Tunnicliffe-Wilson (1997). Fitting time series model by minimizing multistep-ahead errors: a frequency domain approach. *Journal of the Royal Statistical Society* **B 59**, 237–254.

Hendry, D. F. (2005). Robustifying forecasts from equilibrium-correction systems. *Journal of Econometrics*, forthcoming.

Johansen, S. (1996). *Likelihood-based inference in cointegrated vector auto-regressive models*. Oxford: Oxford University Press.

Johnston, H. N. (1974). A note on the estimation and prediction inefficiency of 'dynamic' estimators. *International Economic Review* **15**, 251–255.

Jonsson, G. (1999). Inflation, money demand and purchasing power parity in south africa. Working paper 122, International Monetary Fund.

Kang, I.-B. (2003). Multi-period forecasting using different models for different horizons: an application to U.S. economic time series data. *International Journal of Forecasting* **19**, 387–400.

Klein, L. R. (1971). *An essay on the theory of economic prediction*. Chicago, IL: Markham.

Lin, J. L. and R. S. Tsay (1996). Co-integration constraint and forecasting: An empirical examination. *Journal of Applied Econometrics* **11**, 519–538.

Liu, S. I. (1996). Model selection for multiperiod forecasts. *Biometrika* **83**(4), 861–873.

Marcellino, M., J. H. Stock, and M. W. Watson (2005). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, forthcoming.

Peña, D. (1994). Discussion: Second-generation time-series model, A comment. *Journal of Forecasting* **13**, 133–140.

Pesaran, M. H. and A. Timmermann (2005). Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics* **129**, 183–217.

Schorfheide, F. (2002). VAR forecasting under local misspecification: Loss function estimation and model selection. Working paper, Economics department, University of Pennsylvania.

Tiao, G. C. and R. S. Tsay (1994). Some advances in non-linear and adaptive modelling in time-series analysis. *Journal of Forecasting* **13**, 109–131.

Tiao, G. C. and D. Xu (1993). Robustness of maximum likelihood estimates for multi-step predictions: the exponential smoothing case. *Biometrika* **80**, 623–641.

Tsay, R. S. (1993). Comment: Adpative forecasting. *Journal of Business and Economic statistics* **11**(2), 140–142.

Weiss, A. A. (1991). Multi-step estimation and forecasting in dynamic models. *Journal of Econometrics* **48**, 135–149.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48**, 817–838.

Table 3: VAR Statistics

Statistic	R	Δcpi	$(m - cpi)$	y	rer	VAR
$\widehat{\sigma}$	8.03%	0.86%	4.84%	1.11%	6.29%	
$F_{ar}\,(5, 127)$	0.517	1.22	2.03^{*}	0.108	1.41	–
$F_{arch}\,(4, 124)$	2.39	0.684	0.715	0.740	0.625	–
$F_{het}\,(22, 109)$	1.80	0.773	0.931	1.25	1.09	–
$\chi^2_{nd}\,(2)$	40.6^{**}	10.8^{**}	0.805	7.48^{*}	10.1^{**}	–
$F^{v}_{ar}\,(125, 511)$	–	–	–	–	–	1.06
$F^{v}_{het}\,(330, 1228)$	–	–	–	–	–	1.06
$\chi^{2v}_{nd}\,(10)$	–	–	–	–	–	66.1^{**}

Table 4: Cointegration statistics: eigenvalues (λ), log-likelihood (l), Trace statistic (Tr) and corresponding pvalue.

rank	λ	l	Tr	*pvalue*
0	–	1457.141	156.95^{**}	[0.000]
1	0.453	1499.968	71.293^{**}	[0.009]
2	0.243	1519.728	31.773	[0.407]
3	0.124	1529.099	13.031	[0.737]
4	0.058	1533.359	4.5108	[0.671]
5	0.031	1535.615	–	–

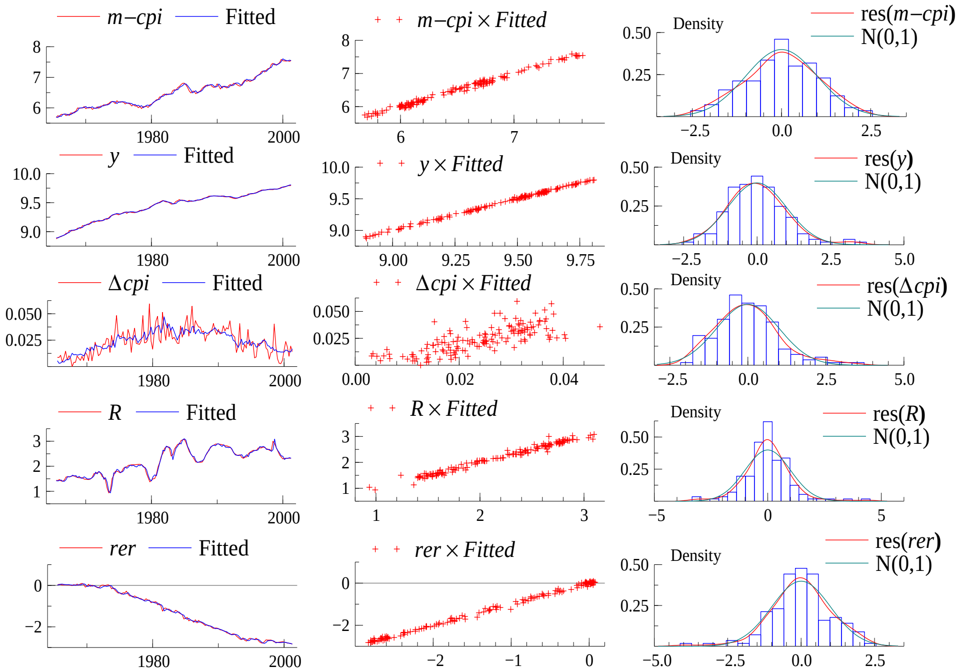

Figure 1: Goodness of fit of the VAR model.

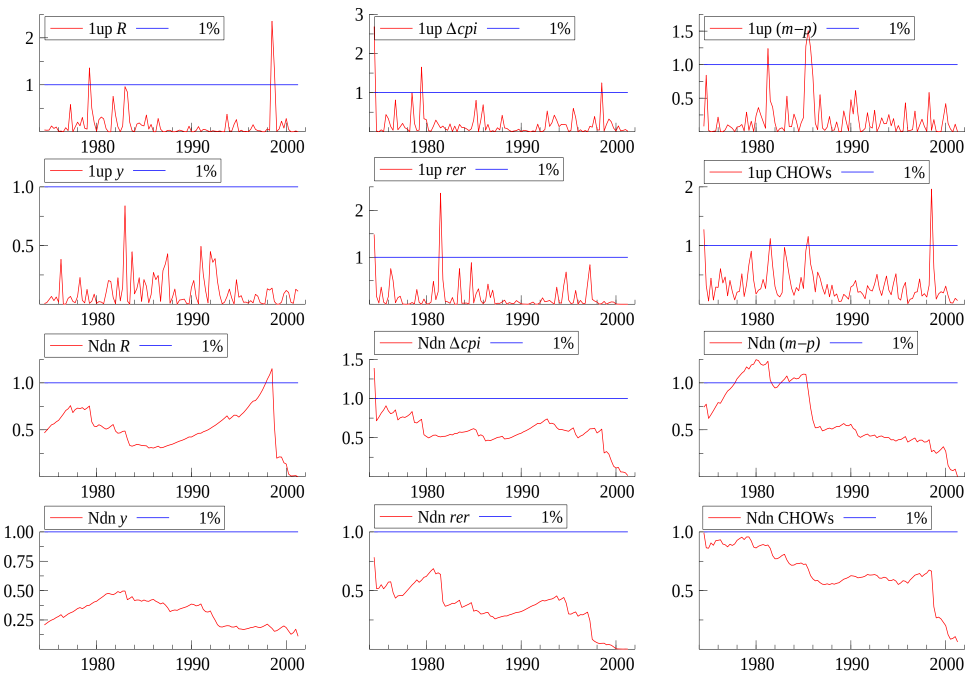

Figure 2: 1↑ and N↓ (BreakPoint) Chow constancy tests for the individual equations and the system.

Table 5: Empirical MSFEs ($\times 10,000$) and their accuracy ranking. Stars after figures in parentheses indicate whether the MSFE is significantly different from that of rank indicated by the figure (which is the next for which significance is achieved) according to the modified Diebold–Mariano test statistic, respectively at the 5% (*) or the 1% level (**).

	Period						
Forecast	73(4)–00(2)	ranking		73(4)–86(4)	ranking	86(1)–00(2)	ranking
DV	11.09	14	(18^*)	15.08	12	6.70	12
DDV_DMS	12.56	15	(18^*)	19.53	17	5.61	9
DDV_IMS	21.68	18		32.70	19	11.42	17
AM	0.12	1	$(2^*, 2^{**})$	0.17	1	0.07	1
AM_trend	1.60	3	$(6^*, 6^{**})$	0.96	3	2.09	3
AM_noDUM92	0.37	2	$(3^*, 3^{**})$	0.22	2	0.53	2
AM_noDUM92_trend	1.68	4	$(6^*, 6^{**})$	1.21	4	2.21	4
AM_nodum	2.62	6	(18^*)	1.44	6	3.77	7
AM_nodum_trend	1.85	5	$(6^*, 9^{**})$	1.37	5	2.38	5
M1_IMS	9.03	10	(18^*)	14.82	11	3.60	6
M1_DMS	10.91	13	(18^*)	15.35	14	7.24	14
M1_IC	9.47	11	(18^*)	15.34	13	3.85	8
DM1_IMS	8.52	8	(18^*)	11.45	9	6.19	11
DM1_IC	8.35	7		10.75	8	5.82	10
VARAM_IMS	168.13	20		323.15	20	29.49	19
VARAM_DMS	21.76	19		27.32	18	29.49	20
VARAM_IC	169.90	21		325.55	21	30.96	21
DVARAM_IMS	13.32	16		18.13	15	10.44	16
DVARAM_IC	14.10	17		19.01	16	11.49	18
DVARAM_IMS_nodum	9.63	12	(18^*)	12.44	10	8.11	15
DVARAM_IC_nodum	8.72	9	(18^*)	10.44	7	7.19	13

Table 6: Empirical MSFEs ($\times 10,000$) of the VAR models when a deterministic trend is included in the model.

	Period		
Forecast	1973(4)–2000(2)	1973(4)–1986(4)	1986(1)–2000(2)
VARAM_IMS_trend	11.04	13.52	8.46
VARAM_DMS_trend	13.11	10.31	15.77
VARAM_IMSIC_trend	10.88	14.90	7.10
VARAM_DMSIC_trend	10.65	11.60	10.51
M1_IMS_trend	8.38	8.50	8.56
M1_DMS_trend	11.44	9.62	13.53
M1_IMSIC_trend	10.50	16.58	4.40
M1_DMSIC_trend	7.63	9.14	6.20

Table 7: Empirical MSFEs ($\times 10,000$) of the Pooled forecast of respectively 15 estimators (the AMs, DM1s, and M1s with and without a trend) and 6 estimators (the AMs).

	Period		
Forecast	1973(4)–2001(2)	1973(4)–1986(4)	1986(1)–2000(2)
Pool 15	2.78	3.33	4.35
Pool 6	0.85	0.61	1.09

8 Appendix

8.1 Two-regime estimation

Equation (14) is derived as follows. From an estimate of the slope, the estimator of the intercept is given by $\widehat{\gamma}^* = \{\overline{\mathbf{x}}^*\}_T^1 - \widehat{\mathbf{\Gamma}}\{\overline{\mathbf{x}}^*\}_{T-1}^0$,where the notation $\{\overline{\mathbf{x}}^*\}_T^1$ refers to the average of $\{\mathbf{x}_t^*\}$ over the sample $t = 1, ..., T$. Let $\mathbf{\Gamma}^{\langle h \rangle} = \sum_{i=1}^h \mathbf{\Gamma}^{\{i\}}$, $\boldsymbol{\delta}_\Gamma^{(\,)} = \mathbf{\Gamma} - \widehat{\mathbf{\Gamma}}^{(\,)}$, and $\boldsymbol{\iota}_{\mathbf{\Gamma}} = \mathbf{I}_n - \mathbf{\Gamma}$. Then:

$$\widehat{\gamma}^* = \widehat{\gamma} - \boldsymbol{\lambda}\widehat{\boldsymbol{\Delta}}_\Gamma \{\overline{\mathbf{x}}\}_{T-1}^0 + \frac{1}{T}\left\{\left(\boldsymbol{\iota}_{\mathbf{\Gamma}} + \boldsymbol{\delta}_\Gamma^{(\,)} - \boldsymbol{\lambda}\widehat{\boldsymbol{\Delta}}_\Gamma\right)\mathbf{\Gamma}^{\langle k \rangle} + \mathbf{\Gamma}^{\{k+1\}}\right\}\nabla\boldsymbol{\gamma}, \tag{22}$$

where $\widehat{\gamma} = \{\overline{\mathbf{x}}\}_T^1 - \widehat{\mathbf{\Gamma}}\{\overline{\mathbf{x}}\}_{T-1}^0$ is defined to render explicit the specific impact of the break. In this formulation, the slope is the same in both regimes; $\widehat{\mathbf{\Gamma}}^{(*)}$ and $\widehat{\mathbf{\Gamma}}^{(\,)}$ are therefore estimators of the same $\mathbf{\Gamma}$, and their values should be very close—although they differ since the sizes of the samples used for estimation may be very different (k versus $T-k$), and so can their biases. When it exists, the expectation of the estimator is thus:

$$\mathsf{E}\left[\widehat{\gamma}^*\right] = \boldsymbol{\gamma} + \frac{1}{T}\left\{\boldsymbol{\iota}_{\mathbf{\Gamma}}\mathbf{\Gamma}^{\langle k \rangle} + \mathbf{\Gamma}^{\{k+1\}}\right\}\nabla\boldsymbol{\gamma}, \tag{23}$$

which, using the $(\mathbf{z}, \mathbf{w})$ partition, simplifies to:

$$\mathsf{E}\begin{bmatrix}\widehat{\gamma}_z^* \\ \widehat{\gamma}_w^*\end{bmatrix} = \begin{cases}\boldsymbol{\gamma}_z + T^{-1}\left[k\mathbf{I}_r - \mathbf{\Gamma}_z\left(\mathbf{I}_r - \mathbf{\Gamma}_z\right)^{-1}\left(\mathbf{I}_r - \mathbf{\Gamma}_z^k\right)\right]\nabla\boldsymbol{\gamma}_z, \\ \boldsymbol{\gamma}_w + \frac{k+1}{T}\nabla\boldsymbol{\gamma}_w,\end{cases}$$

since in this case: $\mathbf{\Gamma}_z^{\langle h \rangle} = \left(\mathbf{I}_r - \mathbf{\Gamma}_z\right)^{-2}\left\{h\left(\mathbf{I}_r - \mathbf{\Gamma}_z\right) - \mathbf{\Gamma}_z\left(\mathbf{I}_r - \mathbf{\Gamma}_z^h\right)\right\}$, and $\mathbf{I}_r^{\langle h \rangle} = \dfrac{h(h+1)}{2}\mathbf{I}_r$. We can therefore write that

$$\mathsf{E}\left[\widehat{\gamma}^*\right] \simeq \boldsymbol{\gamma} + \frac{k}{T}\nabla\boldsymbol{\gamma}. \tag{24}$$

The accuracy of this approximation actually depends on the process under consideration and whether it coincides (with the exception of the break omission) with the model used. We can also decompose $\widehat{\gamma}^*$ in a sum of components amongst which is a weighted average of estimators from the two regimes:

$$\begin{aligned}\widehat{\gamma}^* = {} & \frac{T-k}{T}\left(\mathbf{I}_n - \boldsymbol{\lambda}\right)\widehat{\gamma}^{(\,)} + \frac{k}{T}\boldsymbol{\lambda}\widehat{\gamma}^{(*)} \\ & + \frac{k}{T}\left(\boldsymbol{\iota}_{\mathbf{\Gamma}L} + \boldsymbol{\delta}_{\Gamma L}^{(\,)}\right)\{\overline{\mathbf{x}}\}_T^{T_0+1} \\ & + \boldsymbol{\lambda}\left(\boldsymbol{\iota}_{\mathbf{\Gamma}L} + \boldsymbol{\delta}_{\Gamma L}^{(\,)}\right)\left(\frac{T-k}{T}\{\overline{\mathbf{x}}\}_{T_0}^1 - \frac{k}{T}\{\overline{\mathbf{x}}\}_T^{T_0+1}\right) \\ & - \frac{T-k}{T}\boldsymbol{\lambda}\widehat{\boldsymbol{\Delta}}_\Gamma\{\overline{\mathbf{x}}\}_{T_0-1}^0 + \frac{k}{T}\left(\mathbf{I}_n - \boldsymbol{\lambda}\right)\left(\boldsymbol{\iota}_{\mathbf{\Gamma}L} + \boldsymbol{\delta}_{\Gamma L}^{(\,)}\right)\mathbf{\Gamma}^{\langle k \rangle}\nabla\boldsymbol{\gamma},\end{aligned} \tag{25}$$

where we write $\boldsymbol{\iota}_{\mathbf{\Gamma}L} = \mathbf{I}_n - \mathbf{\Gamma}L$, $\boldsymbol{\delta}_{\Gamma L}^{(\,)} = \mathbf{\Gamma}L - \widehat{\mathbf{\Gamma}}^{(\,)}L$ and, notionally, $L\mathbf{\Gamma}^{\langle k \rangle} = \mathbf{\Gamma}^{\langle k-1 \rangle}$.

In general, in the non-stationary directions, $\{\mathbf{w}_t\}$, $\{\overline{\mathbf{w}}\}_{T_0}^1$ and $\{\overline{\mathbf{w}}\}_T^{T_0+1}$ differ, but $\boldsymbol{\iota}_{\mathbf{\Gamma}_w} = 0$; by contrast, for stationary processes:

$$\mathsf{E}\left[\{\overline{\mathbf{z}}\}_{T_0}^1 - \{\overline{\mathbf{z}}\}_T^{T_0+1}\right] = 0,$$

but the eigenvalues of $\mathbf{\Gamma}_z$ may be very different from unity. There is, thus, a trade-off in the second line of (25). The coefficients in $(T-k)/T$ and k/T do modify these results, but, by partitioning

$\mathbf{x}_t$ into $\mathbf{z}_t$ and $\mathbf{w}_t$, we notice that two cases arise: either (i) the process $\{\mathbf{z}_t\}$ is stationary and we can work in deviation from the mean without loss of generality, hence, setting $\boldsymbol{\gamma}_z = 0$—yet $\nabla\boldsymbol{\gamma}_z \neq 0$—nearly so are the sample averages; or (ii) the process is unit-root non-stationary, or trend-stationary, in which case, if the expectations exist, $\mathsf{E}[\boldsymbol{\iota}_{\boldsymbol{\Gamma}}] = \mathsf{E}\left[\boldsymbol{\delta}^{(\ )}_{\Gamma L}\right] = 0$, or at least $\left(\boldsymbol{\iota}_{\boldsymbol{\Gamma}L} + \boldsymbol{\delta}^{(\ )}_{\Gamma L}\right) \underset{T\to\infty}{\overset{p}{\to}} 0$. In order to derive an approximation, we turn to univariate series. The approximate weights of the pre- and post-break estimators become approximately

$$\begin{aligned}\widehat{\Gamma} &\simeq (1-\lambda)\,\widehat{\Gamma}^{(\ )} + \lambda\widehat{\Gamma}^{(*)}\\ \widehat{\gamma}^{*} &\simeq \frac{T-k}{T}(1-\lambda)\,\widehat{\gamma}^{(\ )} + \frac{k}{T}\lambda\widehat{\gamma}^{(*)},\end{aligned}$$

where we note that, in the case of the intercept, the weights do not add up to one but to $1 - k/T\,(1+2\lambda) - \lambda = 1 - k/T - \lambda\,(1-2k/T)$. For instance, if $\lambda = k/T$, then the weights are $(1-\lambda)^2$ and λ^2 and their sum is lower than unity. Assuming again the existence of expectations for notational ease, (13) rewrites as the weighted sum of the two regime estimators:

$$\begin{aligned}\mathsf{E}[\widehat{\gamma}^{*}] &\simeq \left(1-\frac{k}{T}\right)\gamma + \frac{k}{T}(\gamma+\nabla\gamma) = \left(1-\frac{k}{T}\right)\gamma + \frac{k}{T}\gamma^{*}\\ &= \gamma + \frac{k}{T}\nabla\gamma.\end{aligned}$$

For stationary processes, if the numerator and denominator of (10) can be treated separately and letting

$$\varpi_k(\Gamma) = \frac{(1-\Gamma)^2}{1 - \dfrac{\Gamma\left\{(2+\Gamma)\left(1-\Gamma^{k-1}\right)+\Gamma^{2k-1}\right\}}{k\,(1-\Gamma^2)}},$$

its expectation is, then, approximately:

$$\mathsf{E}[\lambda] \simeq \frac{k}{T}\left(1 - \frac{1}{k/T + \varpi_k(\Gamma)\,\sigma_x^2/(\nabla\gamma)^2}\right) + O\left((k/T)^2\right),$$

and for low Γ, $\varpi_k(\Gamma)$ is close to 1, and close to zero when Γ tends to unity, so that

$$\mathsf{E}[\lambda] \simeq \frac{k}{T}\left(1 - \frac{T(\nabla\gamma)^2}{k(\nabla\gamma)^2 + T\varpi_k(\Gamma)\,\sigma_x^2}\right).$$

The DMS intercept estimator is given by:

$$\begin{aligned}\widetilde{\boldsymbol{\gamma}}_h^{*} =\ & \frac{T_h-k}{T_h}(\mathbf{I}_n-\boldsymbol{\lambda}_h)\,\widetilde{\boldsymbol{\gamma}}_h^{(\ )} + \frac{k}{T_h}\boldsymbol{\lambda}_h\widetilde{\boldsymbol{\gamma}}_h^{(*)}\\ &+\frac{k}{T_h}\boldsymbol{\delta}^{(\ )}_{\Gamma_h L^h}\{\overline{\mathbf{x}}\}_T^{T_0+1} + \boldsymbol{\lambda}_h\boldsymbol{\delta}^{(\ )}_{\Gamma_h L^h}\left(\frac{T_h-k}{T_h}\{\overline{\mathbf{x}}\}_{T_0}^{h} - \frac{k}{T_h}\{\overline{\mathbf{x}}\}_T^{T_0+1}\right)\\ &-\frac{T_h-k}{T_h}\boldsymbol{\lambda}_h\widetilde{\boldsymbol{\Delta}}_{\Gamma_h}L^h\{\overline{\mathbf{x}}\}_{T_0}^{h} + \frac{k}{T_h}(\mathbf{I}_n-\boldsymbol{\lambda}_h)\,\boldsymbol{\delta}^{(\ )}_{\Gamma_h L^h}\boldsymbol{\Gamma}^{\langle k\rangle}\nabla\boldsymbol{\gamma},\end{aligned} \tag{26}$$

where we write $L^h\boldsymbol{\Gamma}^{\langle k\rangle} = \boldsymbol{\Gamma}^{\langle k-h\rangle}$. The results follow from this equation.

8.2 Modified Diebold–Mariano Statistics

We record below the statistics of the modified Diebold–Mariano test for equality of the empirical MSFEs as they were computed for the various forecasting techniques used in this paper. The results are split between tables 8 and 9 for reasons of length.

Table 8: Modified Diebold–Mariano statistics for testing the equality of MSFEs over 1973(4)–2000(2), to be continued on table 9.

Technique	DV	DDV_DMS	DDV_IMS	AM	AM_trend	AM_noDUM92	AM_noDUM92_trend	AM_nodum	AM_nodum_trend	M1_IMS	M1_DMS	M1_IC	DM1_IMS	DM1_IC	VARAM_IMS	VARAM_DMS	VARAM_IC	DVARAM_IMS	DVARAM_IC	DVARAM_IMS_nodum
DV	(−)																			
DDV_DMS	-0.26	(−)																		
DDV_IMS	1.00	2.22	(−)																	
AM	-2.43	-1.38	-2.30	(−)																
AM_detrend	-1.61	-0.84	-1.82	4.91	(−)															
AM_noDUM92	-2.20	-1.25	-2.18	3.48	-3.85	(−)														
AM_noDUM92_trend	-1.55	-0.81	-1.80	6.25	0.36	5.97	(−)													
AM_nodum	-0.97	-0.42	-1.42	5.26	3.95	4.75	2.91	(−)												
AM_nodum_trend	-1.49	-0.76	-1.75	6.30	0.75	6.05	1.96	-2.40	(−)											
M1_IMS	-1.04	-0.79	-2.79	1.23	0.49	1.05	0.45	-0.05	0.39	(−)										
M1_DMS	0.14	0.58	-1.27	2.15	1.48	2.01	1.48	0.96	1.41	2.45	(−)									
M1_IC	-0.88	-0.97	-2.91	1.28	0.56	1.09	0.52	0.02	0.46	0.22	-1.69	(−)								
DM1_IMS	-0.13	0.17	-1.21	2.35	1.47	2.16	1.48	0.82	1.41	0.75	-0.28	0.73	(−)							
DM1_IC	-0.65	0.06	-1.24	3.63	2.18	3.19	2.08	1.09	1.97	0.85	-0.43	0.72	-0.12	(−)						
VARAM_IMS	0.17	0.18	0.14	0.21	0.20	0.21	0.20	0.19	0.20	0.19	0.16	0.19	0.17	0.17	(−)					
VARAM_DMS	2.03	2.09	1.56	2.90	2.71	2.88	2.71	2.54	2.69	2.33	2.09	2.32	2.22	2.22	0.00	(−)				
VARAM_IC	0.18	0.19	0.15	0.23	0.21	0.22	0.21	0.20	0.21	0.20	0.18	0.20	0.18	0.18	1.73	0.01	(−)			
DVARAM_IMS	0.78	2.09	-0.28	2.83	2.17	2.70	2.19	1.66	2.14	2.26	1.10	2.66	1.57	1.15	-0.14	-1.90	-0.15	(−)		
DVARAM_IC	1.22	1.48	0.01	2.95	2.35	2.82	2.33	1.87	2.29	2.12	1.12	2.26	1.15	1.53	-0.13	-1.81	-0.15	0.33	(−)	
DVARAM_IMS_nodum	0.34	0.66	-0.76	3.56	2.49	3.39	2.56	1.67	2.48	1.33	0.24	1.28	1.67	0.76	-0.16	-2.20	-0.17	-0.96	-0.77	(−)
DVARAM_IC_nodum	0.32	0.37	-0.86	4.78	3.29	4.37	3.14	1.95	3.03	1.26	-0.01	1.08	0.30	1.52	-0.16	-2.15	-0.17	-0.78	-1.22	-0.29
VARAM_IMS_trend	0.58	0.71	-0.74	2.73	2.07	2.55	1.98	1.47	1.91	1.66	0.32	1.64	0.55	0.97	-0.15	-1.99	-0.17	-0.46	-0.75	0.09
VARAM_DMS_trend	2.28	1.99	0.76	5.24	4.91	5.10	4.62	4.41	4.49	2.70	1.80	2.77	2.08	3.03	-0.10	-1.43	-0.11	1.08	0.94	1.80
VARAM_IMSIC_trend	0.12	0.36	-1.00	1.91	1.33	1.77	1.29	0.85	1.24	1.12	-0.03	1.06	0.21	0.40	-0.16	-2.07	-0.18	-0.71	-0.96	-0.23
VARAM_DMSIC_trend	1.15	1.31	-0.20	4.68	3.74	4.48	3.68	2.85	3.56	2.17	0.92	2.24	1.21	1.88	-0.14	-1.94	-0.15	0.02	-0.26	0.74
M1_IMS_trend	0.55	0.68	-0.57	5.02	3.51	4.96	3.93	2.45	3.85	1.58	0.41	1.38	1.01	1.16	-0.15	-2.12	-0.16	-0.56	-0.70	0.23
M1_DMS_trend	1.55	1.43	0.34	4.61	3.85	4.61	4.14	3.26	4.11	2.39	1.67	2.09	2.03	2.16	-0.12	-1.64	-0.13	0.72	0.41	1.74
M1_IMSIC_trend	-1.27	-0.35	-1.86	1.55	0.78	1.34	0.73	0.20	0.67	0.44	-0.97	0.25	-0.46	-0.74	-0.18	-2.20	-0.20	-1.39	-1.76	-0.95
M1_DMSIC_trend	-0.18	0.17	-1.22	4.05	2.46	3.63	2.43	1.29	2.29	1.10	-0.40	0.94	0.01	0.20	-0.17	-2.23	-0.18	-1.24	-1.36	-0.74

Table 9: Modified Diebold–Mariano statistics for testing equality of MSFEs over 1973(4)–2000(2), continued.

Technique	DVARAM_IC_nodum	VARAM_IMS_trend	VARAM_DMS_trend	VARAM_IMSIC_trend	VARAM_DMSIC_trend	M1_IMS_trend	M1_DMS_trend	M1_IMSIC_trend	M1_DMSIC_trend
DVARAM_IC_nodum	(−)								
VARAM_IMS_trend	0.48	(−)							
VARAM_DMS_trend	2.84	2.42	(−)						
VARAM_IMSIC_trend	-0.03	-0.90	-2.10	(−)					
VARAM_DMSIC_trend	1.39	0.86	-2.03	1.26	(−)				
M1_IMS_trend	0.61	0.03	-1.86	0.39	-0.69	(−)			
M1_DMS_trend	1.94	1.09	-0.48	1.29	0.78	3.02	(−)		
M1_IMSIC_trend	-1.25	-1.37	-2.56	-0.90	-1.92	-1.20	-2.01	(−)	
M1_DMSIC_trend	-0.48	-0.69	-2.59	-0.25	-1.65	-1.02	-2.19	0.77	(−)