Multi-Step Forecasting in the Presence of Location Shifts

Guillaume Chevillon*

ESSEC Business School

and

CREST–INSEE

December, 2008

Abstract

To forecast at several, say h, periods into the future, a modeler faces two techniques: iterating one step ahead forecasts (the IMS technique) or directly modeling 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 in-sample location shifts on the performances of the two forecasting techniques and relate them to the size of the post-break window and to the forecasting horizon of interest. We also consider two versions of intercept correction (IC), a robust forecasting technique. We show analytically and by simulation that breaks provide a rationale for using other methods than IMS: DMS and IC particular. IMS only performs well relative to the other techniques for shifts occuring towards the end of the sample. In an empirical application, we revisit an oft-used dataset of G7 macroeconomic series and corroborate our theoretical results.

Keywords: Multi-step forecasting, Structural breaks, Intercept Correction.

JEL *Classification*: C32, C53.

---

*Corresponding e-mail: chevillon@essec.fr. Corresponding address: ESSEC Business School, Av. Bernard Hirsch B.P. 50105, 95021 Cergy-Pontoise cedex, France. I am grateful to Michael Clements, David Hendry, Bent Nielsen, Hashem Pesaran and participants at the workshop in Forecasting under Model Instability at the University of Cambridge for helpful comments and suggestions. Research support from the Department of Economics at Oxford University is also gratefully acknowledged.

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 (iterated multi-step of IMS) or *directly* modeling the relation between the end-of-sample observation and its hth successor in order to forecast the latter (direct multi-step, or DMS). The direct technique has a long pedigree but it was originally thought that it brought no benefit, until Weiss (1991) found asymptotic relevance in matching estimation and forecast efficiency criteria, in particular in the presence if model misspecification. From them, many authors have produced theoretical analyses. In particular Tiao and Xu (1993), Clements and Hendry (1996), Chevillon and Hendry (2005) and Proietti (2008) studied misspecified ARIMA processes, Bhansali (1996, 1997), Brodsky and Hurvich (1999) and Bhansali and Kokoszka (2002) analyzed long memory processes, Haywood and Tunnicliffe-Wilson (1997) focused on the frequency domain, Schorfheide (2005) allowed for asymptotically vanishing misspecification. Recently Findley, Potscher and Wei (2004) and Ing (2003, 2004) derived results for very general settings, see Chevillon (2007) for a survey of the literature.

These theoretical studies have spurred a number of empirical analyses of the relative merits of IMS and DMS forecasting. Many have concluded, like Weiss (1991), that the theoretical benefits did not appear clearly in practice. Indeed, Lin and Tsay (1996), Kang (2003), Eklund and Karlsson (2005), Jorda and Marcellino (2007), Schumacher and Breitung (2008) and Proietti (2008) found evidence both in favor and against DMS. Together with the major analysis of 171 U.S. monthly macroeconomic time series by Marcellino, Stock and Watson (2006), which strongly favours iterative forecasting techniques, all these forecast comparisons do not point towards a real improvement brought by DMS. By contrast, Tiao and Tsay (1994) and Tsay (1993) do find significant benefits in using DMS. Yet, many authors who found evidence against direct multi-step used post-war U.S. data (although over extended periods). The United States has not undergone massive shifts in this era and by aggregation over such a large economy, some relative macroeconomic stability is expected. But several theoretical analyses—as in Peña (1994) for breaks, Bhansali (1997) whose framework of long memory can be seen in the light of Perron and Qu (2007) as relevant for regularly occurring breaks, and Chevillon and Hendry (2005) for negative serial correlation that could be induced by occasional location shifts—point towards deterministic shifts as a potential source for the success of DMS in finite samples. This has led **** (2009) to perform a forecast comparison using data for South Africa, an economy that has suffered many regime changes over the last decades. Focusing on the GDP and using 779 different techniques, this author has found substantial evidence sustaining the use of DMS in multivariate models: it outperformed iterated methods, unless the latter were 'intercept corrected' as in Clements and Hendry (1998).

In view of the recent evidence, we propose in this paper to analyze the theoretical properties of DMS forecasting in the presence of breaks. 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 prior to the forecast period. Clements and Hendry (1999, 2006) 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 modeled 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 (2005) derive optimal one-step ahead estimation windows for autoregressive models in this framework. Here, we allow for varying horizons and, in addition, we analyze the improvements that intercept correction brings to forecasting.

The plan of this paper is as follows: we first present the VAR model under analysis and then, in section 3, the properties of multi-step VAR forecasting when the DGP undergoes a location shift and where we abstract from estimation issues. In the following sections, we then focus on the influence of the break on estimation (section 4) and then on forecasting with a misspecified model (section 5). We provide results from simulation in section 6. In section 7, we proceed to an empirical application. Finally we conclude about the use of multi-step methods in the presence of breaks in section 8.

2 The model

In this paper, we analyze the particular case of a data generating process which can be written as the n–variate VAR(1):

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

where the disturbances, $\boldsymbol{\epsilon}_t$, are covariance stationary, with zero mean and variance-covariance matrix $\boldsymbol{\Omega}_\epsilon$ (we will restrict the dynamics of the disturbances below). We allow for the eigenvalues of $\boldsymbol{\Pi}$ to be less than or equal to unity in modulus, thus inducing integration and cointegration. The vector $\mathbf{x}_t$ can be quite general, but we assume that it does not comprise a deterministic component, yet nonzero $\boldsymbol{\gamma}$ and $|\boldsymbol{\Pi} - \mathbf{I}| = 0$ together imply 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 modeled 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 $\boldsymbol{\Pi}^{\{i\}} = \sum_{j=0}^{i-1} \boldsymbol{\Pi}^j$, (and $\boldsymbol{\Pi}^{\{i\}} = \mathbf{0}$, for $i \leq 0$), then:

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

This choice of break is motivated by which is corroborated by Pesaran and Timmermann (2004, 2005): namely that it is location shifts that are the main source of forecast failure. The impact of the break on estimation and forecasting depends on whether $\mathbf{x}_t$ is stationary. It is well known that if $\mathbf{x}_t \sim \mathsf{I}(0)$ then in $\mathbf{\Pi}^{\{i+1\}} \to (\mathbf{I}_n - \mathbf{\Pi})^{-1}$ as $i \to \infty$, i.e. the location shift in (2) is felt progressively and induces an asymptotic mean change. By contrast, for $\mathbf{x}_t \sim \mathsf{I}(1)$, $\mathbf{\Pi}^{\{i+1\}} = O(i+1)$ and $\nabla\boldsymbol{\gamma}$ induces a deterministic trend shift. Since our purpose is location shifts, we mainly focus below on $\mathsf{I}(0)$ processes, but will occasionally comment on integrated variables.

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). Alternatively, DMS consists in OLS estimation of the multi-step model, for $h > 1$

$$\mathbf{x}_t = \boldsymbol{\gamma}_h + \mathbf{\Pi}_h \mathbf{x}_{t-h} + \boldsymbol{v}_t, \quad \text{for } t = h, ..., T, \tag{3}$$

where $\boldsymbol{v}_t$ is not modeled, but still assumed to have zero mean. Parameter estimates are in turn used to compute the forecast of $\mathbf{x}_{T+h}$ from a forecast origin at T. Two additional methods will be examined. Their aim is to put the forecast 'back on track' by adding to the forecasts from the previous two models the differences between the observed forecast origin and its in-sample fitted values from the estimated models (see Hendry (2006) 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).

3 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. 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. Pesaran and Timmermann (2005) show that in such a situation, it can be optimal not to estimate the model uniquely using post-break data owing to the small available sample.

3.1 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:

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

where $\mathbf{\Pi}^{\{h\}} = \sum_{i=0}^{h-1} \mathbf{\Pi}^i$. Neglecting parameter estimation as a first analysis, so $\widehat{\mathbf{\Pi}} = \mathbf{\Pi},$ 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:

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

where the component $\mathbf{\Pi}^h \mathbf{\Pi}^{\{k+1\}} \nabla\boldsymbol{\gamma}$ corrects the misspecification effect in the forecasting model and arises from using the observed, post-break, $\mathbf{x}^*_T$, instead of the hypothesized $\mathbf{x}_T$. Now, the actual realized value is: $\mathbf{x}^*_{T+h} = \mathbf{\Pi}^{\{h\}}\boldsymbol{\gamma}^* + \mathbf{\Pi}^h \mathbf{x}^*_T + \sum_{i=0}^{h-1} \mathbf{\Pi}^i \boldsymbol{\epsilon}_{T+h-i}$, with resulting forecast error:

$$\widehat{\mathbf{e}}^*_{T+h|T} = \mathbf{x}^*_{T+h} - \widehat{\mathbf{x}}^*_{T+h|T} = \mathbf{\Pi}^{\{h\}} \nabla\boldsymbol{\gamma} + \sum_{i=0}^{h-1} \mathbf{\Pi}^i \boldsymbol{\epsilon}_{T+h-i}. \tag{4}$$

$\mathbf{\Pi}^h \mathbf{\Pi}^{\{k+1\}} \nabla\boldsymbol{\gamma}$ does not appear in (4) since it is the true post-break forecast origin which is used by the forecaster; yet the forecast is biased:

$$\mathsf{E}\left[\widehat{\mathbf{e}}^*_{T+h|T}\right] = \mathbf{\Pi}^{\{h\}} \nabla\boldsymbol{\gamma}, \quad \text{and} \quad \mathsf{V}\left[\widehat{\mathbf{e}}^*_{T+h|T}\right] = \sum_{i=0}^{h-1} \mathbf{\Pi}^i \mathbf{\Omega}_\epsilon \mathbf{\Pi}^{i\prime} + 2\sum_{i=0}^{h-1} \sum_{j=i+1}^{h-1} \mathbf{\Pi}^i \mathbf{\Theta}_{|i-j|} \mathbf{\Pi}^{j\prime}, \tag{5}$$

where $\mathbf{\Theta}_m = \mathsf{E}\left[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}'_{t-m}\right]$. The DMS forecasts are:

$$\widetilde{\mathbf{x}}_{T+h|T} = \boldsymbol{\gamma}_h + \mathbf{\Pi}_h \mathbf{x}^*_T, \tag{6}$$

with forecast errors

$$\widetilde{\mathbf{e}}^*_{T+h|T} = \mathbf{\Pi}^{\{h\}} \nabla\boldsymbol{\gamma} + \left(\mathbf{\Pi}^{\{h\}}\boldsymbol{\gamma} - \boldsymbol{\gamma}_h\right) + \left(\mathbf{\Pi}^h - \mathbf{\Pi}_h\right) \mathbf{x}^*_T + \sum_{i=0}^{h-1} \mathbf{\Pi}^i \boldsymbol{\epsilon}_{T+h-i}$$

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

3.2 Intercept Correction

A modeler 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, yields the IC forecast error:

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

so that the forecasts are, now, unbiased for $h = 1$. At longer horizons, the behavior of the forecast bias is similar to the non-corrected version, bounded if $\mathbf{x}_t \sim \mathsf{I}(0)$ and monotonic in h for integrated processes. Yet, the bias of IC is smaller than that of the non intercepted since $\left\|\mathbf{\Pi}^{\{h\}} - \mathbf{I}_n\right\| < \mathbf{\Pi}^{\{h\}}$. Only two situations bring new patterns: first if the dynamics only depends on extraneous variables, i.e. $\mathbf{\Pi} = \mathbf{0}_{n\times n}$, then IC forecasts are unbiased. Second, in the case of over-differenced series were $\mathbf{\Pi} = -\mathbf{I}_n$, then IC brings unbiasedness at odd horizons. These results hold whether IMS or DMS is used since these methods do not differ when the model is well-specified (but for the break) and parameters are known. When $\mathbf{\Pi} = \mathbf{0}_{n\times n}$, the result is in line with the approach in Clements and Hendry (2004). Here the extraneous variables are contained in $\boldsymbol{\epsilon}_t$, they can be correlated and undergo location shifts that will translate into nonzero $\nabla\boldsymbol{\gamma}$.

As seen in Clements and Hendry (1998), the IC forecast error variance is not necessarily larger than in the uncorrected case if $\boldsymbol{\epsilon}_t$ is autocorrelated:

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

The forecast errors of the IMS and IC models do not depend so far on the date of the break. So no information about this event is yet necessary; but if the modeler 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:

$$\boldsymbol{\delta}_{DMSIC} = \mathbf{x}_T^* - \widetilde{\mathbf{x}}_{T|T-h}^* = \mathbf{x}_T^* - \left(\boldsymbol{\gamma}_h + \mathbf{\Pi}_h\mathbf{x}_{T-h}^*\right) \tag{7}$$

which rewrites as $\boldsymbol{\delta}_{DMSIC} = \left(\mathbf{\Pi}^{\{k+1\}} - \mathbf{\Pi}_h\mathbf{\Pi}^{\{k-h+1\}}\right)\nabla\boldsymbol{\gamma} + \sum_{i=0}^{h-1}\mathbf{\Pi}^i\boldsymbol{\epsilon}_{T-i}$. The forecast error results in:

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

with variance

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

Whereas the DMSIC forecast error variance is larger than that of IMS and IC, except for some values of $\{\mathbf{\Theta}_m\}$, there exists a possibility that the bias may be lower. With known parameters, DMSIC and IMSIC coincide and $\mathbf{\Pi}_h = \mathbf{\Pi}^h$. Then

$$\mathsf{E}\left[\boldsymbol{\delta}_{DMSIC}\right] = \begin{cases} \mathbf{0}_n, & \text{if } h \leq k+1, \\ \left(\mathbf{\Pi}^{\{h\}} - \mathbf{\Pi}^{\{k+1\}}\right)\nabla\boldsymbol{\gamma}, & \text{if } h > k+1. \end{cases} \tag{8}$$

so that the DMSIC forecast bias is zero as long as the forecast horizon is ‘low’ compared to the post-break sample size, k. Even at horizons beyond $k+1$, the forecast bias is lower than obtained by the other methods. This result obviously comes at the cost of higher variance if $\boldsymbol{\epsilon}_t \sim i.i.d.$, although not necessarily so, depending on the autocorrelation structure of $\{\boldsymbol{\epsilon}_t\}$.

For $h > k+1$, if the process is stationary $\mathbf{\Pi}_z^{\{h\}} - \mathbf{\Pi}_z^{\{k+1\}} = \left(\mathbf{I}_r - \mathbf{\Pi}_z\right)^{-1}\left(\mathbf{\Pi}_z^{k+1} - \mathbf{\Pi}_z^h\right)$,where, for $h > k+1$, $\left|\mathbf{\Pi}_z^h\right| < \left|\mathbf{\Pi}_z^{k+1}\right|$. As for integrated variables, the second line on the right-hand side of (8) becomes a trend $O\left(h-k-1\right)$. Asymptotically, for $\nabla\boldsymbol{\gamma} \neq \mathbf{0}$ and $h > k+1$

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

We summarize these results in the following proposition.

Proposition 1 *In the known-parameter model (1) with $\boldsymbol{\epsilon}_t$ covariance stationary and mean zero, break date $T_0 = T - k < T$ and where the forecast is intercept corrected using expression (7), denoting by h the forecast horizon, then:*

1. *The forecast bias is zero if the dynamics of $\mathbf{x}_t$ only depends on stationary extraneous variables*

2. At 'low' horizons (i.e. $h \leq k+1$), the forecast bias is zero.

3. At 'longer' horizons:

(a) For $\mathbf{x}_t$ stationary, the forecast bias is finite and tends to zero as both h and k become large.

(b) For $\mathbf{x}_t$ over-differenced ($\mathbf{\Pi} = -\mathbf{I}_n$), the forecast bias is $\frac{(-1)^k-(-1)^{h+1}}{2}\nabla\boldsymbol{\gamma}$

(c) For $\mathbf{x}_t$ integrated of order one, the forecast bias is of the order of $(h-k-1)\nabla\boldsymbol{\gamma}$

If the process is stationary, and for fixed k, the DMSIC 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 situation, the forecast bias would explode to infinity with $(h-k-1)\nabla\boldsymbol{\gamma}$ if the model were never re-specified after the break had occurred. Indeed k only becomes large if the sample size so does but if the modeler does not notice that a break has occurred. This is unlikely in practice unless $\nabla\boldsymbol{\gamma}$ is small compared to $\boldsymbol{\gamma}$. Point 2 in the previous proposition is crucial for the interest in multi-step 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 as corrections 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.

4 Estimation in the presence of a location shift

This section shows how, in the presence of an intercept shift, parameter estimators can be thought of as weighted averages of those obtained over the pre- and post-break subsamples. For this, we let $k/T \to c \in [0,1]$ and define $\mathbf{g}$ such that $\nabla\boldsymbol{\gamma} = \boldsymbol{\gamma}\mathbf{g}$ (if $\boldsymbol{\gamma} \neq 0$, or we let $\nabla\boldsymbol{\gamma} = \mathbf{g}$)

4.1 Iterated multi-step estimation over two regimes

In the framework above and in the context of an intercept shift, we assume that the parameters $(\boldsymbol{\gamma}^*, \mathbf{\Pi})$ are estimated by ordinary least-squares. We denote by $\widehat{\mathbf{\Pi}}^{(\ )}$ the pre-break slope estimator, using the sample up to $T_0 - 1$ and $\widehat{\mathbf{\Pi}}^{(*)}$ using the remaining observations. The overall estimator is $\widehat{\mathbf{\Pi}}$, which can be written as $(\mathbf{I}_n - \boldsymbol{\lambda})\,\widehat{\mathbf{\Pi}}^{(\ )} + \boldsymbol{\lambda}\widehat{\mathbf{\Pi}}^{(*)}$ with weight $\boldsymbol{\lambda} = \left(\widehat{\mathbf{\Pi}} - \widehat{\mathbf{\Pi}}^{(\ )}\right)\left(\widehat{\mathbf{\Pi}}^{(*)} - \widehat{\mathbf{\Pi}}^{(\ )}\right)^{-1}$. Let $\widehat{\mathbf{\Delta}}_{\Pi} = \widehat{\mathbf{\Pi}}^{(*)} - \widehat{\mathbf{\Pi}}^{(\ )}$, then in finite samples $\mathsf{E}\left[\widehat{\mathbf{\Delta}}_{\Pi}\right] \neq \mathbf{0}$ but it vanishes asymptotically and is of order $O_p\left(T^{-1}\right)$ (see below). More generally, as $(k, T) \to (\infty, \infty)$ the three estimators $\widehat{\mathbf{\Pi}}, \widehat{\mathbf{\Pi}}^{(\ )}$ and $\widehat{\mathbf{\Pi}}^{(*)}$ converge to $\mathbf{\Pi}$ and we can assume that $\|\mathsf{E}\left[\boldsymbol{\lambda}_\infty\right]\| \leq \infty$, where $\boldsymbol{\lambda}_\infty = plim_{(k,T)\to(\infty,\infty)}\boldsymbol{\lambda}$. We disregard the case where k is held constant while $T \to \infty$ for which $\|\boldsymbol{\lambda}\|$ diverges. We also let the intercept estimator weight $\boldsymbol{\mu} = \left(\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\gamma}}^{(\ )}\right)\left(\widehat{\boldsymbol{\gamma}}^{(*)} - \widehat{\boldsymbol{\gamma}}^{(\ )}\right)^{-1}$ and $\overline{\mathbf{x}}_i = T^{-1}\sum_{t=1}^{T}\mathbf{x}_{t+i}$, then

$$\widehat{\boldsymbol{\gamma}} = \overline{\mathbf{x}} - \widehat{\mathbf{\Pi}}\overline{\mathbf{x}}_{-1} = \left(1 - \frac{k}{T}\right)\left[\overline{\mathbf{x}}^{(\ )} - \widehat{\mathbf{\Pi}}\overline{\mathbf{x}}_{-1}^{(\ )}\right] + \frac{k}{T}\left[\overline{\mathbf{x}}^{(*)} - \widehat{\mathbf{\Pi}}\overline{\mathbf{x}}_{-1}^{(*)}\right] \underset{T\to\infty}{\to} \boldsymbol{\gamma}\left(\mathbf{I}_n + c\mathbf{g}\right) \tag{9}$$

hence for $T \to \infty$, $\boldsymbol{\mu} \to c\mathbf{I}_n$, which we can rewrite as $\boldsymbol{\mu} \to c$, i.e. the weight is asymptotically scalar.

Our interest lies in the iterated multistep parameters $\left(\widehat{\Pi}^h, \widehat{\boldsymbol{\gamma}}_{\{h\}}\right)$. First the slope estimator, by Taylor expansion of $\left(\widehat{\mathbf{\Pi}}^{(\ )} + \boldsymbol{\lambda}\widehat{\mathbf{\Delta}}_{\Pi}\right)^h$ yields $\boldsymbol{\lambda}_{\{h\}} = h\boldsymbol{\lambda} + o_p\left(\widehat{\mathbf{\Delta}}_{\Pi}\right)$, but using $\left(\widehat{\mathbf{\Pi}}^{(*)} + (\boldsymbol{\lambda} - \mathbf{I}_n)\,\widehat{\mathbf{\Delta}}_{\Pi}\right)^h$ leads to $\left(\mathbf{I}_n - \boldsymbol{\lambda}_{\{h\}}\right) = h\left(\mathbf{I}_n - \boldsymbol{\lambda}\right) + o_p\left(\widehat{\mathbf{\Delta}}_{\Pi}\right)$. By interpolation we obtain $\boldsymbol{\lambda}_{\{h\}} \approx h\boldsymbol{\lambda}\left(\mathbf{I}_n - \boldsymbol{\lambda}\right) + o_p\left(\widehat{\mathbf{\Delta}}_{\Pi}\right)$. This is an approximation that can be shown by induction to hold except around a neighborhood of $k = T/2$. Now, the multi-step intercept estimator is $\widehat{\boldsymbol{\gamma}}_{\{h\}} = \left(\sum_{i=0}^{h-1}\widehat{\Pi}^i\right)\widehat{\boldsymbol{\gamma}}$. We show in the appendix that this rewrites as

$$\widehat{\boldsymbol{\gamma}}_{\{h\}} = \left(\mathbf{I} + \boldsymbol{\lambda}_{\{h\}}\right)\widehat{\boldsymbol{\gamma}}_{\{h\}}^{(\ )} - \boldsymbol{\lambda}_{\{h\}}\widehat{\boldsymbol{\gamma}}_{\{h\}}^{(*)} + O_p\left(\widehat{\mathbf{\Delta}}_{\Pi}\right) \tag{10}$$

The multistep weights $\boldsymbol{\mu}_{\{h\}}$ are therefore defined as $-\boldsymbol{\lambda}_{\{h\}}$.

The conclusion from this analysis is that the intercept estimated over the whole sample is a weighted sum of those over the pre- and post-break subsamples, with respective weights that are non-linear. In terms of IMS estimators, the weights for the slope are linear in h.

In order to obtain a finite sample approximation, we resort to univariate processes since they capture the main message. Yet, Pantula and Fuller (1985), Maekawa (1987), Grubb and Symons

(1987) and several articles by Kiviet and Phillips (1999 and with Schipp, 2005) have derived properties of the autoregressive slope in more general ARX processes which could also be used. Unfortunately, allowing for exogenous variables poses a problem when the aim is forecasting since the modeler must then avail herself of accurate forecasts thereof. Hence we restrict our attention to autoregressive processes and the AR(1) in particular as finite sample approximations exist in the literature. We denote by (γ, π) the univariate parameters.

First regarding the one-step estimators, the approximation by Kendall (1954), which was shown by Nankervis and Savin (1988) to work well for $|\pi| < 0.8$, is

$$\mathsf{E}\left[\widehat{\pi}^{(\,)} - \pi\right] = -\frac{1+3\pi}{T-k} + O\left(\frac{1}{(T-k)^2}\right).$$

This implies that the value of the intercept does not matter for the slope estimator bias. Hence using the corresponding expressions for equivalent approximations concerning $\widehat{\pi}^{(*)}$ and $\widehat{\pi}$

$$\lambda = \frac{k}{T}\frac{k}{2k-T} + O_p\left(T^{-2}\right) \underset{T\to\infty}{\to} \frac{c^2}{2c-1} \tag{11}$$

which is negative if $c < 1/2$ and not defined if $c = 1/2$ since the approximation then yields $\widehat{\mathbf{\Delta}}_\Pi \approx 0$. In practice, the finite sample expectation of the estimator also depends on the initial value of the process in the sample. Kendall's approximation does allow for this and it will in practice depend on x_{T_0}. Yet we use (11), as we will see in section 6 that it is accurate enough to serve our purpose, and it yields:

$$\lambda_{\{h\}} = -\mu_{\{h\}} \underset{T\to\infty}{\to} -h\frac{c^2(1-c)^2}{(2c-1)^2}.$$

This approximation cannot apply for large horizons $h > [Tc]$, but it yields $\lambda_{\{r\}} \approx -Tr\frac{c^2(1-c)^2}{(2c-1)^2}$ if we let $h/T \to r < c$.

4.2 Direct multi-step estimation

The case of multi-step estimation is very comparable to one-step. But, in OLS estimation of (3), 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 most recent observations. We define the DMS slope estimator weights $\boldsymbol{\lambda}_h$ such that $\widetilde{\mathbf{\Pi}}_h = (\mathbf{I}_n - \boldsymbol{\lambda}_h)\,\widetilde{\mathbf{\Pi}}_h^{(\,)} + \boldsymbol{\lambda}_h \widetilde{\mathbf{\Pi}}^{(*)}$. $\boldsymbol{\lambda}_h$ depends on the relative values of h and k. An equivalent of the Kendall (1954) approximation was derived by Marriott and Pope (1954),

$$\mathsf{E}\left[\widetilde{\pi}_h^{(\,)} - \pi^h\right] = -\frac{1}{T-k-h+1}\left[(1+\pi)\,\pi^{\{h\}} + 2h\pi^h\right] + O\left(T^{-2}\right)$$

hence, using equivalent approximations for $\widetilde{\pi}_h^{(*)}$ and $\widetilde{\pi}_h$, yields

$$\lambda_h \approx \frac{k-h}{T-h}\frac{k}{2\,(k-h)-T}$$

so, $\lambda_h/\lambda \approx 1$ for h finite and large (T,k), but when $h/T \to r$, then

$$\lambda_h \approx \frac{c-r}{1-r}\frac{c}{2\,(c-r)-1}$$

Now, $\widetilde{\boldsymbol{\gamma}}_h = \overline{\mathbf{x}}^* - \widetilde{\mathbf{\Pi}}_h \overline{\mathbf{x}}^*_{-h} = \left(1 - \frac{k}{T-h+1}\right)\left[\overline{\mathbf{x}}^{(\,)} - \widetilde{\mathbf{\Pi}}_h \overline{\mathbf{x}}^{(\,)}_{-h}\right] + \frac{k}{T-h+1}\left[\overline{\mathbf{x}}^{(*)} - \widetilde{\mathbf{\Pi}}_h \overline{\mathbf{x}}^{(*)}_{-h}\right]$ and let $\mu_h = \frac{k}{T-h+1}$ so that (see appendix)

$$\begin{aligned}
\widetilde{\boldsymbol{\gamma}}_h &= (1-\mu_h)\,\widetilde{\boldsymbol{\gamma}}_h^{(\,)} + \mu_h \widetilde{\boldsymbol{\gamma}}_h^{(*)} + O_p\left(\widetilde{\boldsymbol{\Delta}}_\Pi\right) \\
&\underset{T\to\infty}{\to} \left(1 - \frac{c}{1-r}\right)\boldsymbol{\gamma}_r + \frac{c}{1-r}\boldsymbol{\gamma}_r^*
\end{aligned}$$

where $\boldsymbol{\gamma}_r = \mathbf{\Pi}^{\{h\}}\boldsymbol{\gamma}$ if $r = 0$ (we assume that h is then held constant) and $\boldsymbol{\gamma}_r = (\mathbf{I}_n - \mathbf{\Pi})^{-1}\boldsymbol{\gamma}$ if $r > 0$. We define $\boldsymbol{\gamma}_r^*$ similarly. Notice that contrary to IMS, the weights for multi-step intercept estimation are scalar and hence identical in all directions. Also, the weight attributed to $\widetilde{\boldsymbol{\gamma}}_h^{(*)}$ is increasing in both the post-break window and the horizon r. Next, we observe the resulting implications for forecasting.

5 Multi-step forecasts

In order to compare the forecasting performances of IMS and DMS, we first decompose their forecast errors across to the two subsamples. This is done in the appendix, where we show that the iterated multi-step forecast error can be written as the weighted sum of those issuing from using the pre- and post-break models, but that an additional component arises from the misspecification. The latter has non-zero expectation, even asymptotically:

$$\begin{aligned}
\widehat{\mathbf{e}}_{T+h|T} &= \left(\mathbf{I}+\boldsymbol{\lambda}_{\{h\}}\right)\widehat{\mathbf{e}}^{(\,)}_{T+h|T}-\boldsymbol{\lambda}_{\{h\}}\widehat{\mathbf{e}}^{(*)}_{T+h|T} \\
&\quad+\left(\mathbf{I}+\boldsymbol{\lambda}_{\{h\}}\right)\left[\boldsymbol{\Pi}^{\{h\}}+\left(\boldsymbol{\Pi}^{h}-\widehat{\boldsymbol{\Pi}}^{(\,)h}\right)\boldsymbol{\Pi}^{\{k+1\}}\right]\nabla\boldsymbol{\gamma} \\
&\quad+O_p\left(\widehat{\boldsymbol{\Delta}}_{\Pi}\right)
\end{aligned} \tag{12}$$

The same can be done for the direct multi-step forecast and the same property arises, with corresponding estimators and weights:.

$$\begin{aligned}
\widetilde{\mathbf{e}}_{T+h|T} &= \left(1-\mu_h\right)\widetilde{\mathbf{e}}^{(\,)}_{T+h|T}+\mu_h\widetilde{\mathbf{e}}^{(*)}_{T+h|T} \\
&\quad+\left(1-\mu_h\right)\left[\boldsymbol{\Pi}^{\{h\}}+\left(\boldsymbol{\Pi}^{h}-\widetilde{\boldsymbol{\Pi}}^{(\,)}_{h}\right)\boldsymbol{\Pi}^{\{k+1\}}\right]\nabla\boldsymbol{\gamma} \\
&\quad+O_p\left(\widehat{\boldsymbol{\Delta}}_{\Pi}\right)
\end{aligned} \tag{13}$$

These results allow to formulate the following proposition.

Proposition 2 *In the model presented in proposition 1 where the parameters are estimated using least-squares, where the regressors and disturbances are such that the slope estimators are unbiased, the forecast errors resulting from iterated and direct multi-step have non-zero expectations and these are functions of the weight attributed to the pre-break model in estimating the intercept:*

$$\begin{aligned}
\mathsf{E}\left[\widehat{\mathbf{e}}_{T+h|T}-\left(\mathbf{I}-\boldsymbol{\mu}_{\{h\}}\right)\widehat{\mathbf{e}}^{(\,)}_{T+h|T}-\boldsymbol{\mu}_{\{h\}}\widehat{\mathbf{e}}^{(*)}_{T+h|T}\right] &= \left(\mathbf{I}_n-\boldsymbol{\mu}_{\{h\}}\right)\boldsymbol{\Pi}^{\{h\}}\nabla\boldsymbol{\gamma} \\
\mathsf{E}\left[\widetilde{\mathbf{e}}_{T+h|T}-\left(1-\mu_h\right)\widetilde{\mathbf{e}}^{(\,)}_{T+h|T}-\mu_h\widetilde{\mathbf{e}}^{(*)}_{T+h|T}\right] &= \left(1-\mu_h\right)\boldsymbol{\Pi}^{\{h\}}\nabla\boldsymbol{\gamma}
\end{aligned}$$

The previous proposition shows that for models that are well-specified over each—pre- and post-break—subsamples, so that $\widehat{\mathbf{e}}^{(\,)}_{T+h|T}$, $\widehat{\mathbf{e}}^{(*)}_{T+h|T}$, $\widetilde{\mathbf{e}}^{(\,)}_{T+h|T}$ and $\widetilde{\mathbf{e}}^{(*)}_{T+h|T}$ have zero expectation, the forecast computed from the model estimated over the whole sample is biased. Its bias depends on the interaction between the break size and the characteristic roots of the lag polynomial presiding over the dynamics of $\mathbf{x}_t$. In addition, irrespective of the relative performance of IMS and DMS forecasting for well-specified and constant models, multi-step forecast biases are function of the weight attributed to the pre-break multi-step intercept estimator.

We showed earlier that these weights are scalar for DMS, but vectorial for IMS. In addition, they are decreasing in both c and r, and hence so are the absolute forecast biases. In the univariate

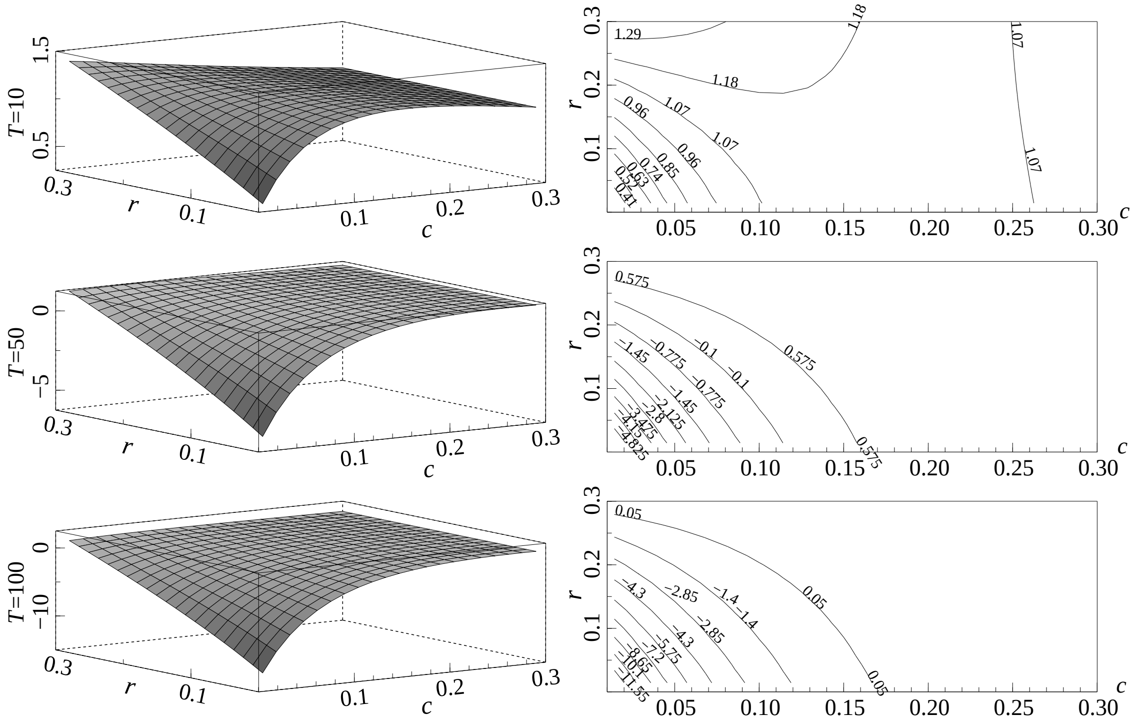

Figure 1: Plot of the function θ defined as in expression (14) for $(r, c) \in [0, 3]^2$ and $T \in \{10, 50, 100\}$ (left column) with corresponding two dimensional projection contours (right column).

and stationary framework considered above, the relative bias is then

$$\frac{1-\mu_{\{h\}}}{1-\mu_h} \approx \frac{(1-r)\left[(2c-1)^2 - Trc^2(1-c)^2\right]}{(1-r-c)(2c-1)^2} = \theta(T, r, c) \tag{14}$$

which is valid for $r < 1-c$, i.e. if DMS is computable on the pre-break subsample. The function θ is drawn in figure 1: at $T = 10$, it only takes positive values, but for larger sample sizes θ is negative when both c and r are low. Hence very small samples, long horizons and not-so-recent breaks favor DMS over IMS according to expression (14).

Proposition 2 concerns the forecast bias, not the mean-square forecast error. But expressions (12) and (13) show that the forecast error variances are approximately that of the well-specified models (yet with differing weights). Since IMS forecasting is well known to yield smaller error variances than DMS in the absence of misspecification, it is hence mainly with respect to the forecast bias that DMS can bring an accuracy benefit (the different $\mu_{\{h\}}$ and μ_h could yet mitigate this).

We do not consider intercept correction here as we show in the next section that the known parameter framework is sufficiently general.

6 Monte Carlo

In this section, we present the results from a Monte Carlo simulation of forecast errors resulting from the previous models.[1] We consider autoregressive processes with *i.i.d.* innovations where we let the sample size vary from T =10 to 90, by steps of 10, with $0 \leq k \leq T-5$ (which implies $T_0 \geq 5$) and $h \leq T/2$. The parameters are $\gamma = 1$, $\nabla\gamma = g \in \{-3, -1, 1, 3\}$, and $\pi \in \{0.5, 1\}$. This results in 53,400 combinations, with 2,000 Monte Carlo replications each.

In the case of stationary processes, i.e. for $\pi = 0.5$, a response surface was fitted to the case $g = 1$ and $r < 1 - c$ which yielded for the 10,299 observations considered:

$$\begin{aligned}\widehat{\log\left|\frac{E\left[\widehat{\mathsf{e}}_{T+h|T}\right]}{E\left[\widetilde{\mathsf{e}}_{T+h|T}\right]}\right|} &= \underset{(8.63\times10^{-3})}{0.769} + \underset{(1.79\times10^{-3})}{0.166}\log\left|\frac{(1-r)\left[(2c-1)^2 - Trc^2(1-c)^2\right]}{(1-r-c)(2c-1)^2}\right| \\ \mathsf{R}^2 &= 45.5\%, \quad \widehat{\sigma} = 0.643 \end{aligned} \qquad (15)$$

where E is the Monte Carlo mean. When estimating the same equation only for the larger sample sizes increases the goodness of fit. According to section 5, the ratio $\frac{\mathsf{E}\left[\widehat{\mathsf{e}}_{T+h|T}\right]}{\mathsf{E}\left[\widetilde{\mathsf{e}}_{T+h|T}\right]}$ of mean forecast errors is independent of g. In order to assess this property, we report in figure 2 the scatterplots of the log ratios and of the Monte Carlo parameters. Panel d shows that the amplitude of the break impacts on that of the ratio: the variability of the ratio is smaller for $g = -1$ than for the other values thereof. This case is atypical in that the break offsets the intercept: the process shifts from an expectation of 2 to zero. Hence the approximation above is limited in that it does not take into account this type of effect. Also, panel e shows that the amplitude of the log ratio increases with the sample size but this corresponds to more experiments in the Monte Carlo and may hence be spurious.

[1] Computations were carried out using Oxmetrics and the Ox programming language, see Doornik (2006). Figure panels are labeled alphabetically from left to right, top to bottom.

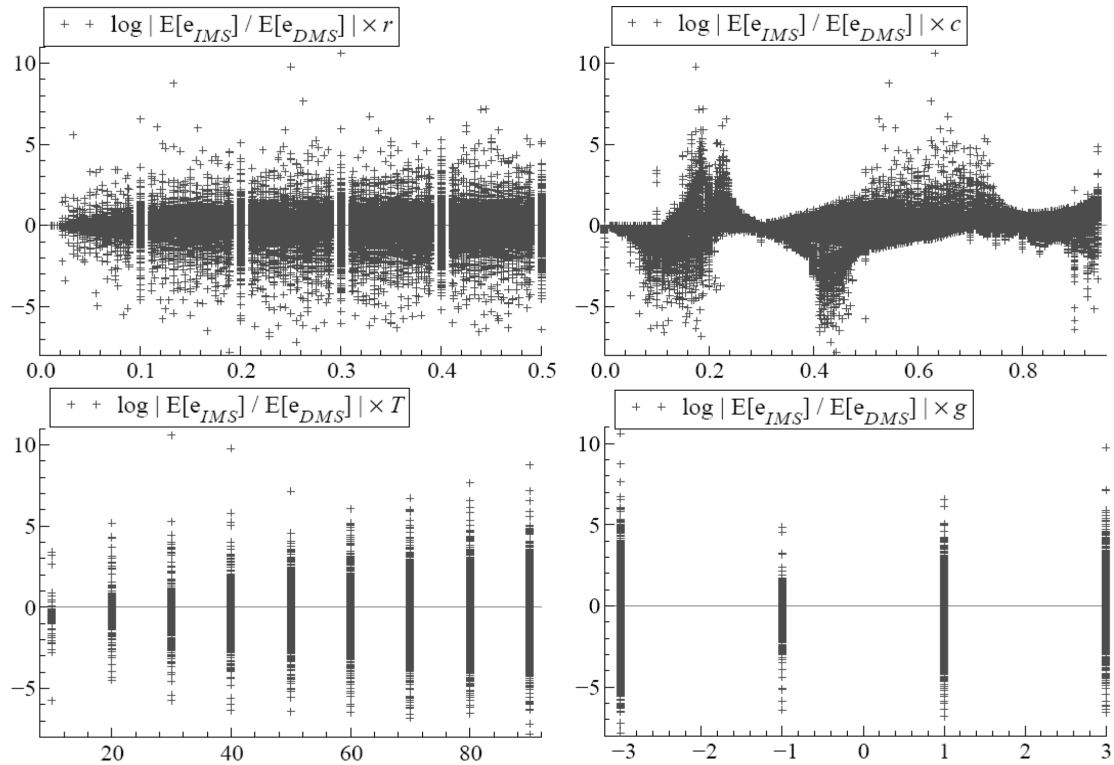

Figure 2: Logarithm of the absolute ratio of the IMS mean forecast error over that of DMS for a range of parameters.

By contrast, the effect of the break date is nonlinear: on panel b, the mean IMS error is smaller than that of DMS for recent breaks (low c). The relative performances reverse for c slightly above 0.2. For less recent breaks, the relative performances are more difficult to compare properly from this setting but DMS outperforms IMS in terms of mean forecast bias when $r < 1 - c$.

Now, figure 3 presents scatterplots of the log ratios of the absolute biases of IMS over that of IMSIC, and DMS over that of DMSIC For both IMS and DMS, intercept correction improves the average forecast error for recent breaks, but the gain is strictly positive until older breaks using DMSIC. As for the horizon, improvements that IC brings to IMS and DMS varies: for the iterated, panel c shows that the distribution of the log ratios is mostly positive, whereas that of DMS has an mean that decreases with r. Hence the benefit to DMS disappears, whereas that to IMS remains both when r and c become large. With respect to the sample size, no pattern emerges clearly for DMS: the amplitude is wider for larger samples but as previously, it may be a simulation artifact; by contrast the distribution moves from being symmetric around zero positive values for larger T. Intercept correction therefore benefits IMS more in larger sample but the number of observations does not matter to DMSIC.

Finally, for intercept correction, section 3.2 provided expression (8) in the known parameter case. In order to validate it, figure 4 records the Monte Carlo mean forecast error of DMSIC. The upper panel shows that the known parameter expression explains the observed biases when the latter are negative, but not when they are positive. To understand patterns better, the lower panel records the same mean biases, but ordered according to the Monte Carlo experiments: letting T, then c then r vary. We observe that the DMSIC error is largely negative for very recent shifts and that again is well explained by expression (8) and does not depend on the sample size. The bias tends to increase with c as long as $r < 1-c$ and this comes from estimation error. When $r \geq 1-c$, $E\left[\widetilde{e}_{DMSIC}\right]$ is much closer to zero and this is the situation where (14) cannot be derived since DMS cannot be estimated over the pre-break subsample only.

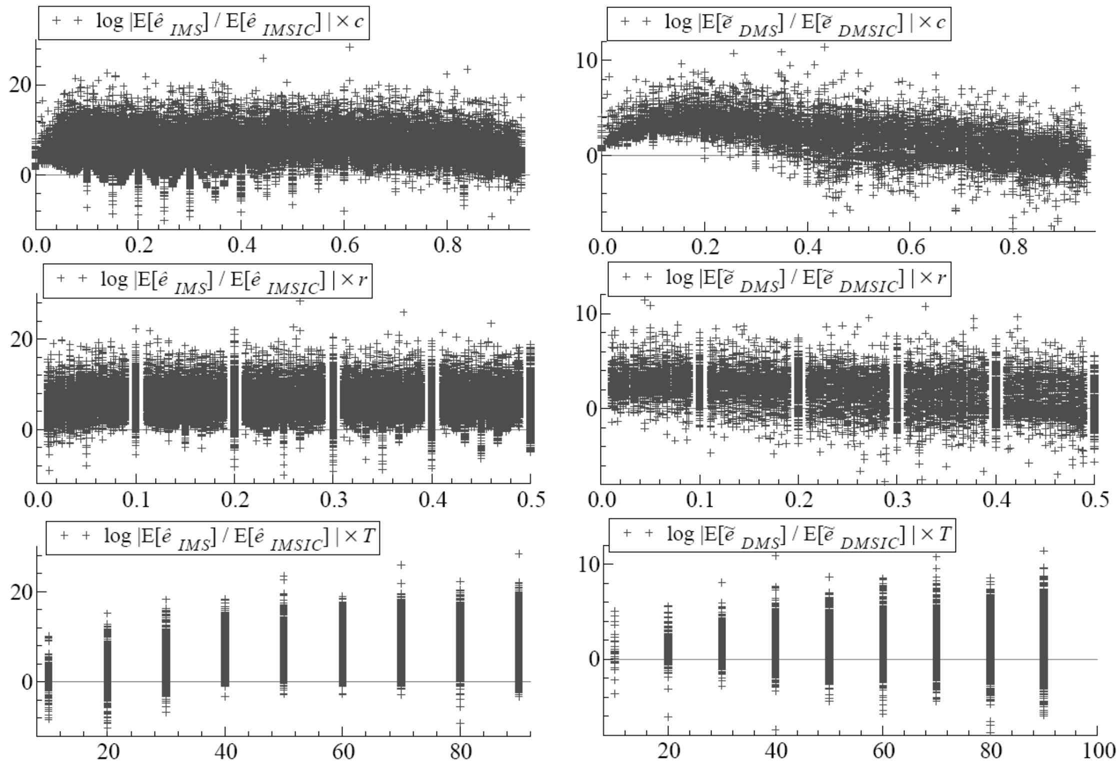

Figure 3: Monte Carlo log ratio of the IMS (and DMS) absolute mean forecast error over that of IMSIC (and DMSIC): scatterplot against the experiment parameters.

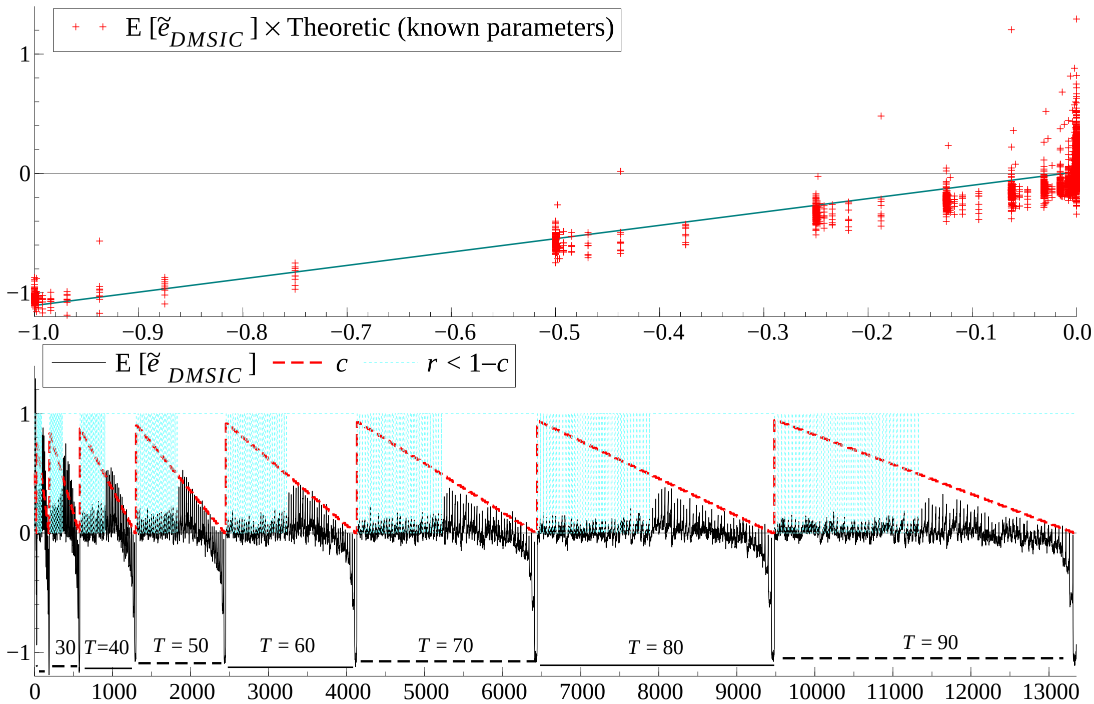

Figure 4: Monte Carlo DMSIC mean forecast error. The upper panel presents a scatterplot against the known-parameter theoretical expression (8). The lower panel presents the same but against the range of parameters used in the simulation. $r < 1 - c$ is an indicator variable that takes value 1 if the inequality holds.

7 Empirical Analysis

We assess the previous analysis by revisiting the forecasting exercise in Pesaran and Timmermann (2005). These authors evaluate the forecasting properties of various estimation schemes using the Stock and Watson (2004) dataset, which is quarterly and covers the period 1959-99. This consists in forecasting growth in industrial production and real GDP, the inflation rate and the short interest rates for six of the seven G7 economies (Canada, France, Germany, Japan, the UK and the US). To see the impact of a break on forecasting, we used rolling window estimation of AR(1) models over samples of 40 observations (the *long rolling window* in Pesaran and Timmermann, 2005). We also estimated the date of a unique break using the method and code proposed by Bai and Perron (2003): the resulting post-break window size averaged over the 121 subsamples is denoted by $\overline{k}$ and is reported in table 1 together with the ratios of RMSFE. Values of the RMSFE ratio below unity favor IMS against the alternative.

Overall the table shows that IMS is not preferred except for the forecasts of interest rates in Germany. This variable is specific inasmuch as it exhibits the lowest $\overline{k}$ by a large margin. Also IMS outperforms DMS at low horizon for French interest rates which exhibit the second lowest $\overline{k} = 12$. Hence, as expression (14) showed, it is recent breaks and low horizons that benefit IMS over DMS.

For all the other cases. The most precise technique out of the four is IMSIC which is most accurate in 40 instances, followed by DMSIC in 31, then DMS in 13, with IMS only beating the other in four cases. Hence some form of intercept correction is really to be advocated for. When looking at the details according to the horizon, we see that at $h = 2$, 15 times out of 22 IMSIC is best, only beaten in the remaining cases by DMSIC (four instances) and the other two in three cases. At horizons 5 and 10, rankings are reversed: DMSIC is best in respectively 13 and 9 instances, IMSIC in 6 and 8, and DMS also proves most accurate in 2 and 4 cases. It is at long horizons that DMS performs best, being the most accurate in 6 cases, with IMSIC in 11 cases but DMSIC only in 4. Since for all series and samples breaks occur, the conclusion from observing table 1 is that in the presence of breaks iterated multi-step forecasting performs very poorly and that it seems preferable to use techniques that are more robust to breaks: at very short horizons, IMSIC performs best, at intermediate horizons DMSIC is preferred and at long horizons all but IMSIC or DMS are equivalent, except in the presence of a very recent break.

To refine these recommendations, we report in table 2 the same ratios of RMSFE according to the value of k in the estimation sample. The results are then more strongly in favour of intercept correction, be it IMS or DMS as they perform best in 10 and 6 instances respectively, pure DMS in 4 and IMS never. Forecast accuracy follows the same patterns in relation to the horizon as in table 1. With respect to k, recent breaks favor IMSIC, ancient ones DMSIC, whereas for breaks that occur towards the end of sample, IMSIC benefits for low horizons and DMS catches up at large h.

8 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 breaks matter to the relative performances of multi-step forecasting techniques and not only if they occur towards the forecast origin. Indeed, as the post-shift sample size vary, so do the forecasting properties of DMS and IMS. We have seen that the rationale for using IMS no longer applies when the process undergoes a break, unless the latter happens very close to the forecast origin, the forecast horizon is small and the sample size is large enough. By contrast, robust methods such as intercept correction do perform well at all horizons and whenever the break occurs in the sample. We analyzed two techniques of intercept correction and showed (i) that direct multi-step intercept correction performs better when the horizon is not too large with respect to the sample size, (ii) that IMSIC is best at very low or large horizons and (iii) that DMS improves with the horizon. When it is desired to forecast at an horizon which is relatively large compared to the available sample (towards $h = T/2$), then intercept correction did not seem to provide such a robust superior forecast performance and standard DMS may prove sufficient.

		IMS over DMS				IMS over IMSIC				IMS over DMSIC			
	$\overline{k}$	$h=2$	$h=5$	$h=10$	$h=20$	$h=2$	$h=5$	$h=10$	$h=20$	$h=2$	$h=5$	$h=10$	$h=20$
Inflation													
Canada	18	1.30	2.42	1.08	1.39	2.48	3.30	3.00	5.08	1.87	3.25	2.28	1.91
France	21	1.34	1.54	1.74	1.36	4.81	2.57	3.20	7.79	1.98	3.43	3.16	3.29
Germany	17	1.10	4.20	0.97	2.76	2.81	2.89	5.20	1.52	4.00	4.37	8.38	1.87
Japan	22	1.39	1.35	1.14	1.53	1.84	4.83	7.53	2.22	1.80	2.63	3.13	3.01
UK	19	1.57	1.61	1.21	1.98	4.92	2.77	5.55	4.09	2.52	3.16	8.64	1.51
US	18	1.15	2.33	1.83	2.19	2.24	3.19	1.93	3.00	2.00	2.48	18.44	1.50
Industrial production growth													
Canada	19	1.47	1.74	1.28	1.57	1.88	1.14	5.69	1.62	1.73	1.38	6.62	1.31
France	20	1.44	1.42	1.12	7.61	1.64	1.25	3.06	1.73	1.20	1.68	1.76	3.97
Germany	21	1.08	1.18	1.25	1.43	9.34	1.79	2.10	1.25	1.88	1.26	2.72	1.16
Japan	20	1.07	2.19	2.47	1.42	3.34	3.03	0.75	7.70	2.94	1.39	1.00	2.53
UK	28	1.43	2.13	1.64	4.48	1.64	3.39	1.14	1.79	2.44	4.17	1.50	1.83
US	21	1.28	1.54	1.28	1.51	16.28	1.75	2.10	1.77	1.75	5.28	1.74	1.63
Real GDP growth													
Canada	20	1.30	1.33	3.60	1.62	3.24	1.45	29.09	2.98	1.84	3.34	3.33	5.13
France	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Germany	23	1.15	1.44	1.19	1.22	1.29	5.12	1.65	16.86	1.58	1.53	9.99	2.88
Japan	18	1.34	1.26	1.40	1.74	6.59	2.42	3.60	1.48	2.31	79.13	3.74	4.22
UK	21	1.27	1.61	1.20	1.39	2.36	1.44	20.57	1.34	2.01	4.42	3.37	1.39
US	22	1.96	1.04	1.79	1.37	1.42	1.52	2.79	5.43	1.73	1.73	3.78	2.54
Interest rate													
Canada	18	1.17	3.36	4.85	4.52	3.19	3.10	2.60	7.42	3.58	3.43	2.43	2.02
France	12	0.60	1.09	0.99	1.50	1.61	2.18	3.47	1.90	1.34	2.40	1.23	1.39
Germany	7	0.39	0.49	0.49	0.53	0.47	0.42	0.40	0.47	0.42	0.93	0.74	0.78
Japan	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
UK	22	1.88	3.44	24.25	3.67	1.97	2.50	3.92	2.45	1.80	6.68	3.18	2.04
US	19	2.67	2.80	2.08	2.55	2.51	2.28	2.55	1.97	4.76	1.18	5.10	1.26

Table 1: Out-of-sample forecast performance for the G7 economies: the table reports the average ratios of root-MSFE.

		IMS over DMS				IMS over IMSIC				IMS over DMSIC			
	$\overline{k}$	$h \in \{2\}$	$[3,5]$	$[6,10]$	$[11,20]$	$\{2\}$	$[3,5]$	$[6,10]$	$[11,20]$	$\{2\}$	$[3,5]$	$[6,10]$	$[11,20]$
$k \in [0,5]$	2.5	0.85	1.32	0.92	1.43	4.96	1.83	9.88	1.67	1.87	25.71	2.65	1.28
$(5,10]$	8.0	1.18	1.74	1.97	2.44	2.28	2.81	3.39	2.37	1.99	3.23	2.01	2.21
$(10,20]$	15.4	1.55	2.57	2.37	3.00	5.62	2.50	2.95	2.12	1.77	2.48	3.02	2.26
$(20,30]$	25.7	1.60	1.55	5.83	1.93	2.45	2.40	4.09	6.45	2.45	1.81	4.61	2.39
$(30,40]$	35.4	1.29	1.98	1.73	2.22	2.23	2.75	5.71	4.90	2.55	2.88	7.78	2.71

Table 2: Out-of-sample performance according the post-break window size: the table reports the average ratios of root-MSFE.

The analysis that we have performed here mostly focused on autoregressive forecasting in the presence of a unique break. In line with work by Pesaran and Timmermann, it appears that we should also compare our results with the cases of models with more parameters (more lags) and exogenous variables. Also, evaluating the performance via the forecast error first two moments may not be enough and it would be of interest to consider directional and asymmetric assessment of forecasting performances. Yet, this paper shows that robust forecasting methods are available when the data generating process is not constant.

References

**** (2009). ****. *International Journal of Forecasting*, **forthcoming**.

Bai, J., and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, **18**, 1–22.

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.

Bhansali, R. J., and Kokoszka, P. (2002). Computation of the forecast coefficients for multistep prediction of long-range dependent time series. *International Journal of Forecasting*, **18**, 181–206.

Brodsky, J., and Hurvich, C. M. (1999). Multi-step forecasting for long-memory processes. *Journal of Forecasting*, **18**, 59–75.

Chevillon, G. (2007). Direct multi-step estimation and forecasting. *Journal of Economic Surveys*, **21**, 746–85.

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

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

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

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

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

Clements, M. P., and Hendry, D. F. (2006). Forecasting with breaks. In Elliott, G., Granger, C., and Timmermann, A. (eds.), *Handbook of Economic Forecasting*, Vol. 1, Ch. 12. Amsterdam: North-Holland.

Doornik, J. A. (2006). *An Object-Oriented Matrix Language - Ox 4.* London: Timberlake Consultants Press.

Eklund, J., and Karlsson, S. (2005). Forecast combination and model averaging using predictive measures. *Econometric Reviews*, **26**, 329–.

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

Grubb, D., and Symons, J. (1987). Bias in regressions with a lagged dependent variable. *Econometric Theory*, *3*(3), 371–86.

Haywood, J., and Tunnicliffe-Wilson, G. (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. (2006). Robustifying forecasts from equilibrium-correction systems. *Journal of Econometrics*, **135**, 399–426.

Ing, C.-K. (2003). Multistep prediction in autoregressive processes. *Econometric Theory*, **19**, 254–279.

Ing, C.-K. (2004). Selecting optimal multistep predictors for autoregressive process of unknown order. *The Annals of Statistics*, **32**, 693–722.

Jorda, O., and Marcellino, M. (2007). Path forecast evaluation. mimeo, UC Davis.

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.

Kendall, M. G. (1954). Note on bias in the estimation of autocorrelation. *Biometrika*, **41**, 403–4.

Kiviet, J. F., and Phillips, G. D. (2005). Moment approximation for least-squares estimators in dynamic regression models with a unit root. *Econometrics Journal*, **8**, 115–42.

Kiviet, J. F., Phillips, G. D., and Schipp, B. (1999). Alternative bias approximations in first-order dynamic reduced form models. *Journal of Economic Dynamics and Control*, **23**, 909–28.

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

Maekawa, K. (1987). Finite sample properties of several predictors from an autoregressive model. *Econometric Theory*, *3*(3), 359–70.

Marcellino, M., Stock, J. H., and Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, **135**, 499–526.

Marriott, F. H. C., and Pope, J. A. (1954). Bias in the estimation of autocorrelations. *Biometrika*, **41**, 390–402.

Nankervis, J., and Savin, N. (1988). The exact moments of the least-squares estimator for the autoregressive model: Corrections and extensions. *Journal of Econometrics*, **37**, 381–8.

Pantula, S. G., and Fuller, W. A. (1985). Mean estimation bias in least squares estimation of autoregressive processes. *Journal of Econometrics*, **27**, 99–121.

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

Perron, P., and Qu, Z. (2007). An analytical evaluation of the log-periodogram estimate in the presence of level shifts. Working paper, Boston University.

Pesaran, M. H., and Timmermann, A. (2004). How costly is it to ignore breaks when forecasting the direction of a time series. *International Journal of Forecasting*, **20**, 411–25.

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

Proietti, T. (2008). Direct and iterated multistep AR methods for difference stationary processes. mimeo, University of Rome, Tor Vergata.

Schorfheide, F. (2005). VAR forecasting under local misspecification. *Journal of Econometrics*, **128**, 99–136.

Schumacher, C., and Breitung, J. (2008). Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting*, *24*(3), 386–98.

Stock, J. H., and Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, **23**, 405–30.

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

Tiao, G. C., and Xu, D. (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.

9 Appendix

9.1 Two-regime estimation

We compute the approximation to IMS estimators:

$$
\begin{aligned}
\widehat{\gamma}_{\{h\}} &= \left(\sum_{i=0}^{h-1}\widehat{\Pi}^{i}\right)\widehat{\gamma} = \left(\mathbf{I}-\widehat{\Pi}\right)^{-1}\left(\mathbf{I}-\widehat{\Pi}^{h}\right)\widehat{\gamma} \\
&= \left(\mathbf{I}-\widehat{\Pi}^{(\,)}-\boldsymbol{\lambda}\widehat{\boldsymbol{\Delta}}_{\Pi}\right)^{-1}\widehat{\gamma} - \left(\mathbf{I}-\boldsymbol{\lambda}_{\{h\}}\right)\left(\mathbf{I}-\widehat{\Pi}^{(\,)}-\boldsymbol{\lambda}\widehat{\boldsymbol{\Delta}}_{\Pi}\right)^{-1}\widehat{\Pi}^{(\,)h}\widehat{\gamma} \\
&\quad -\boldsymbol{\lambda}_{\{h\}}\left(\mathbf{I}-\widehat{\Pi}^{(*)}-\left(\mathbf{I}-\boldsymbol{\lambda}\right)\widehat{\boldsymbol{\Delta}}_{\Pi}\right)\widehat{\Pi}^{(*)h}\widehat{\gamma}
\end{aligned}
$$

and $\left(\mathbf{I}-\widehat{\Pi}^{(\,)}-\boldsymbol{\lambda}\widehat{\boldsymbol{\Delta}}_{\Pi}\right)^{-1} = \left(\mathbf{I}-\widehat{\Pi}^{(\,)}\right)^{-1} + O_p\left(\widehat{\boldsymbol{\Delta}}_{\Pi}\right)$ hence

$$
\begin{aligned}
\widehat{\gamma}_{\{h\}} &= \left(\mathbf{I}-\widehat{\Pi}^{(\,)}\right)^{-1}\left(\mathbf{I}-\widehat{\Pi}^{(\,)h}\right)\widehat{\gamma} - \boldsymbol{\lambda}_{\{h\}}\left[\widehat{\Pi}^{(*)\{h\}}-\widehat{\Pi}^{(\,)\{h\}}\right]\widehat{\gamma} + O_p\left(\widehat{\boldsymbol{\Delta}}_{\Pi}\right) \\
&= \left(\mathbf{I}+\boldsymbol{\lambda}_{\{h\}}\right)\widehat{\Pi}^{(\,)\{h\}}\widehat{\gamma} - \boldsymbol{\lambda}_{\{h\}}\widehat{\Pi}^{(*)\{h\}}\widehat{\gamma} + O_p\left(\widehat{\boldsymbol{\Delta}}_{\Pi}\right)
\end{aligned}
$$

We replace $\widehat{\gamma}$ with $\left(\mathbf{I}-\boldsymbol{\mu}\right)\widehat{\gamma}^{(\,)} + \boldsymbol{\mu}\widehat{\gamma}^{(*)}$ and rearrange

$$
\begin{aligned}
\widehat{\gamma}_{\{h\}} &= \left(\mathbf{I}+\boldsymbol{\lambda}_{\{h\}}\right)\widehat{\Pi}^{(\,)\{h\}}\left[\left(\mathbf{I}-\boldsymbol{\mu}\right)\widehat{\gamma}^{(\,)} + \boldsymbol{\mu}\widehat{\gamma}^{(*)}\right] \\
&\quad -\boldsymbol{\lambda}_{\{h\}}\widehat{\Pi}^{(*)\{h\}}\left[\left(\mathbf{I}-\boldsymbol{\mu}\right)\widehat{\gamma}^{(\,)} + \boldsymbol{\mu}\widehat{\gamma}^{(*)}\right] + O_p\left(\widehat{\boldsymbol{\Delta}}_{\Pi}\right) \\
&= \left(\mathbf{I}+\boldsymbol{\lambda}_{\{h\}}\right)\left(\mathbf{I}-\boldsymbol{\mu}\right)\widehat{\gamma}^{(\,)}_{\{h\}} - \boldsymbol{\lambda}_{\{h\}}\boldsymbol{\mu}\widehat{\gamma}^{(*)}_{\{h\}} \\
&\quad + \left(\mathbf{I}+\boldsymbol{\lambda}_{\{h\}}\right)\widehat{\Pi}^{(*)\{h\}}\boldsymbol{\mu}\widehat{\gamma}^{(*)} - \left(\mathbf{I}+\boldsymbol{\lambda}_{\{h\}}\right)\left[\widehat{\Pi}^{(*)\{h\}}-\widehat{\Pi}^{(\,)\{h\}}\right]\boldsymbol{\mu}\widehat{\gamma}^{(*)} \\
&\quad -\boldsymbol{\lambda}_{\{h\}}\left(\mathbf{I}-\boldsymbol{\mu}\right)\widehat{\Pi}^{(\,)\{h\}}\widehat{\gamma}^{(\,)} - \boldsymbol{\lambda}_{\{h\}}\left(\mathbf{I}-\boldsymbol{\mu}\right)\left[\widehat{\Pi}^{(*)\{h\}}-\widehat{\Pi}^{(\,)\{h\}}\right]\widehat{\gamma}^{()} \\
&\quad + O_p\left(\widehat{\boldsymbol{\Delta}}_{\Pi}\right) \\
&= \left(\mathbf{I}+\boldsymbol{\lambda}_{\{h\}}\right)\widehat{\gamma}^{(\,)}_{\{h\}} - \boldsymbol{\lambda}_{\{h\}}\widehat{\gamma}^{(*)}_{\{h\}} \\
&\quad - \left(\mathbf{I}+\boldsymbol{\lambda}_{\{h\}}\right)\left[\widehat{\Pi}^{(*)\{h\}}-\widehat{\Pi}^{(\,)\{h\}}\right]\boldsymbol{\mu}\widehat{\gamma}^{(*)} \\
&\quad -\boldsymbol{\lambda}_{\{h\}}\left(\mathbf{I}-\boldsymbol{\mu}\right)\left[\widehat{\Pi}^{(*)\{h\}}-\widehat{\Pi}^{(\,)\{h\}}\right]\widehat{\gamma}^{(\,)} \\
&\quad + O_p\left(\widehat{\boldsymbol{\Delta}}_{\Pi}\right)
\end{aligned}
$$

We use $\left[\widehat{\Pi}^{(*)\{h\}}-\widehat{\Pi}^{(\,)\{h\}}\right] = O_p\left(\widehat{\boldsymbol{\Delta}}_{\Pi}\right)$ then

$$
\widehat{\gamma}_{\{h\}} = \left(\mathbf{I}+\boldsymbol{\lambda}_{\{h\}}\right)\widehat{\gamma}^{(\,)}_{\{h\}} - \boldsymbol{\lambda}_{\{h\}}\widehat{\gamma}^{(*)}_{\{h\}} + O_p\left(\widehat{\boldsymbol{\Delta}}_{\Pi}\right)
$$

Now the DMS slope estimator is $\widetilde{\mathbf{\Pi}}_h = (\mathbf{I}_n - \boldsymbol{\lambda}_h)\widetilde{\mathbf{\Pi}}_h^{(\ )} + \boldsymbol{\lambda}_h \widetilde{\mathbf{\Pi}}^{(*)}$ and the multi-step intercept:

$$\begin{aligned}
\widetilde{\gamma}_h &= \left(1 - \frac{k}{T-h+1}\right)\left[\overline{\mathbf{x}}^{(\ )} - \widetilde{\mathbf{\Pi}}_h \overline{\mathbf{x}}_{-h}^{(\ )}\right] + \frac{k}{T-h+1}\left[\overline{\mathbf{x}}^{(*)} - \widetilde{\mathbf{\Pi}}_h \overline{\mathbf{x}}_{-h}^{(*)}\right] \\
&= \left(1 - \frac{k}{T-h+1}\right)\left[\overline{\mathbf{x}}^{(\ )} - \left(\widetilde{\mathbf{\Pi}}_h^{(\ )} + \boldsymbol{\lambda}_h \widetilde{\mathbf{\Delta}}_\Pi\right)\overline{\mathbf{x}}_{-h}^{(\ )}\right] \\
&\quad + \frac{k}{T-h+1}\left[\overline{\mathbf{x}}^{(*)} - \left(\widetilde{\mathbf{\Pi}}_h^{(*)} - (\mathbf{I}_n - \boldsymbol{\lambda}_h)\widetilde{\mathbf{\Delta}}_\Pi\right)\overline{\mathbf{x}}_{-h}^{(*)}\right] \\
&= \left(1 - \frac{k}{T-h+1}\right)\widetilde{\gamma}_h^{(\ )} + \frac{k}{T-h+1}\widetilde{\gamma}_h^{(*)} + O_p\left(\widetilde{\mathbf{\Delta}}_\Pi\right).
\end{aligned}$$

9.2 Forecasting

First, the IMS is

$$\begin{aligned}
\widehat{\mathbf{e}}_{T+h|T} &= \mathbf{\Pi}^{\{h\}}\boldsymbol{\gamma}^* - \widehat{\boldsymbol{\gamma}}_{\{h\}}^* + \left(\mathbf{\Pi}^h - \widehat{\mathbf{\Pi}}^h\right)\mathbf{x}_T^* + \sum_{i=0}^{h-1}\mathbf{\Pi}^i \boldsymbol{\epsilon}_{T+h-i} \\
&= \mathbf{\Pi}^{\{h\}}\boldsymbol{\gamma}^* - \left[\left(\mathbf{I} - \boldsymbol{\mu}_{\{h\}}\right)\widehat{\boldsymbol{\gamma}}_{\{h\}}^{(\ )} + \boldsymbol{\mu}_{\{h\}}\widehat{\boldsymbol{\gamma}}_{\{h\}}^{(*)}\right] \\
&\quad + \left(\mathbf{\Pi}^h - \left(\mathbf{I} - \boldsymbol{\lambda}_{\{h\}}\right)\widehat{\mathbf{\Pi}}^{(\ )h} - \boldsymbol{\lambda}_{\{h\}}\widehat{\mathbf{\Pi}}^{(*)h}\right)\mathbf{x}_T^* + \sum_{i=0}^{h-1}\mathbf{\Pi}^i \boldsymbol{\epsilon}_{T+h-i} + O_p\left(\widehat{\mathbf{\Delta}}_\Pi\right)
\end{aligned}$$

We let appear the forecast errors from the two submodels

$$\begin{aligned}
\widehat{\mathbf{e}}_{T+h|T} &= \left(\mathbf{I} - \boldsymbol{\mu}_{\{h\}}\right)\left[\boldsymbol{\gamma}_{\{h\}} - \widehat{\boldsymbol{\gamma}}_{\{h\}}^{(\ )}\right] + \boldsymbol{\mu}_{\{h\}}\left[\boldsymbol{\gamma}_{\{h\}}^* - \widehat{\boldsymbol{\gamma}}_{\{h\}}^{(*)}\right] + \left(\mathbf{I} - \boldsymbol{\mu}_{\{h\}}\right)\mathbf{\Pi}^{\{h\}}\nabla\boldsymbol{\gamma} \\
&\quad + \left(\mathbf{I} - \boldsymbol{\mu}_{\{h\}}\right)\left(\mathbf{\Pi}^h - \widehat{\mathbf{\Pi}}^{(\ )h}\right)\mathbf{x}_T - \boldsymbol{\mu}_{\{h\}}\left(\mathbf{\Pi}^h - \widehat{\mathbf{\Pi}}^{(*)h}\right)\mathbf{x}_T^* \\
&\quad + \left(\mathbf{I} + \boldsymbol{\mu}_{\{h\}}\right)\left(\mathbf{\Pi}^h - \widehat{\mathbf{\Pi}}^{(\ )h}\right)\left[\mathbf{x}_T^* - \mathbf{x}_T\right] \\
&\quad - 2\boldsymbol{\mu}_{\{h\}}\left[\left(\mathbf{\Pi}^h - \widehat{\mathbf{\Pi}}^{(*)h}\right)\mathbf{x}_T^* - \left(\mathbf{\Pi}^h - \widehat{\mathbf{\Pi}}^{(\ )h}\right)\mathbf{x}_T\right] \\
&\quad + \sum_{i=0}^{h-1}\mathbf{\Pi}^i \boldsymbol{\epsilon}_{T+h-i} + O_p\left(\widehat{\mathbf{\Delta}}_\Pi\right) \\
&= \left(\mathbf{I} - \boldsymbol{\mu}_{\{h\}}\right)\widehat{\mathbf{e}}_{T+h|T}^{(\ )} + \boldsymbol{\mu}_{\{h\}}\widehat{\mathbf{e}}_{T+h|T}^{(*)} \\
&\quad + \left(\mathbf{I} - \boldsymbol{\mu}_{\{h\}}\right)\mathbf{\Pi}^{\{h\}}\nabla\boldsymbol{\gamma} \\
&\quad + \left(\mathbf{I} + \boldsymbol{\mu}_{\{h\}}\right)\left(\mathbf{\Pi}^h - \widehat{\mathbf{\Pi}}^{(\ )h}\right)\left[\mathbf{x}_T^* - \mathbf{x}_T\right] \\
&\quad - 2\boldsymbol{\mu}_{\{h\}}\left[\left(\mathbf{\Pi}^h - \widehat{\mathbf{\Pi}}^{(*)h}\right)\mathbf{x}_T^* - \left(\mathbf{\Pi}^h - \widehat{\mathbf{\Pi}}^{(\ )h}\right)\mathbf{x}_T\right] + O_p\left(\widehat{\mathbf{\Delta}}_\Pi\right)
\end{aligned}$$

Now $\mathbf{x}_T^* - \mathbf{x}_T = \mathbf{\Pi}^{\{k+1\}}\nabla\boldsymbol{\gamma}$ hence, since $\widehat{\mathbf{\Pi}}^{(\ )h} - \widehat{\mathbf{\Pi}}^{(*)h} = O_p\left(\widehat{\mathbf{\Delta}}_\Pi\right)$

$$\begin{aligned}
\widehat{\mathbf{e}}_{T+h|T} &= \left(\mathbf{I} - \boldsymbol{\mu}_{\{h\}}\right)\widehat{\mathbf{e}}_{T+h|T}^{(\ )} + \boldsymbol{\mu}_{\{h\}}\widehat{\mathbf{e}}_{T+h|T}^{(*)} \\
&\quad + \left(\mathbf{I} - \boldsymbol{\mu}_{\{h\}}\right)\left[\mathbf{\Pi}^{\{h\}} + \left(\mathbf{\Pi}^h - \widehat{\mathbf{\Pi}}^{(\ )h}\right)\mathbf{\Pi}^{\{k+1\}}\right]\nabla\boldsymbol{\gamma} \\
&\quad + O_p\left(\widehat{\mathbf{\Delta}}_\Pi\right)
\end{aligned}$$

And for DMS

$$
\begin{aligned}
\widetilde{\mathbf{e}}_{T+h|T} &= \mathbf{\Pi}^{\{h\}}\boldsymbol{\gamma}^* - \widetilde{\boldsymbol{\gamma}}_h^* + \left(\mathbf{\Pi}^h - \widetilde{\mathbf{\Pi}}_h\right)\mathbf{x}_T^* + \sum_{i=0}^{h-1}\mathbf{\Pi}^i\boldsymbol{\epsilon}_{T+h-i} \\
&= \mathbf{\Pi}^{\{h\}}\boldsymbol{\gamma}^* - \left[(1-\mu_h)\,\widetilde{\boldsymbol{\gamma}}_h^{(\,)} + \mu_h\widetilde{\boldsymbol{\gamma}}_h^{(*)}\right] \\
&\quad + \left(\mathbf{\Pi}^h - (\mathbf{I}-\boldsymbol{\lambda}_h)\,\widetilde{\mathbf{\Pi}}_h^{(\,)} - \boldsymbol{\lambda}_h\widetilde{\mathbf{\Pi}}_h^{(*)}\right)\mathbf{x}_T^* + \sum_{i=0}^{h-1}\mathbf{\Pi}^i\boldsymbol{\epsilon}_{T+h-i} + O_p\left(\widehat{\mathbf{\Delta}}_\Pi\right) \\
&= (1-\mu_h)\left[\mathbf{\Pi}^{\{h\}}\boldsymbol{\gamma} - \widetilde{\boldsymbol{\gamma}}_h^{(\,)}\right] + \mu_h\left[\mathbf{\Pi}^{\{h\}}\boldsymbol{\gamma}^* - \widetilde{\boldsymbol{\gamma}}_h^{(*)}\right] + (1-\mu_h)\,\mathbf{\Pi}^{\{h\}}\nabla\boldsymbol{\gamma} \\
&\quad + (1-\mu_h)\left[\mathbf{\Pi}^h - \widetilde{\mathbf{\Pi}}_h^{(\,)}\right]\mathbf{x}_T + \mu_h\left[\mathbf{\Pi}^h - \widetilde{\mathbf{\Pi}}_h^{(*)}\right]\mathbf{x}_T^* + \sum_{i=0}^{h-1}\mathbf{\Pi}^i\boldsymbol{\epsilon}_{T+h-i} \\
&\quad + (\mu_h\mathbf{I} - \boldsymbol{\lambda}_h)\left[\mathbf{\Pi}^h - \widetilde{\mathbf{\Pi}}_h^{(\,)}\right]\mathbf{x}_T + (\boldsymbol{\lambda}_h - \mu_h\mathbf{I})\left[\mathbf{\Pi}^h - \widetilde{\mathbf{\Pi}}_h^{(*)}\right]\mathbf{x}_T^* \\
&\quad + (\mathbf{I}-\boldsymbol{\lambda}_h)\left[\mathbf{\Pi}^h - \widetilde{\mathbf{\Pi}}_h^{(\,)}\right]\left[\mathbf{x}_T^* - \mathbf{x}_T\right] + O_p\left(\widehat{\mathbf{\Delta}}_\Pi\right)
\end{aligned}
$$

i.e.

$$
\begin{aligned}
\widetilde{\mathbf{e}}_{T+h|T} &= (1-\mu_h)\,\widetilde{\mathbf{e}}_{T+h|T}^{(\,)} + \mu_h\widetilde{\mathbf{e}}_{T+h|T}^{(*)} \\
&\quad + (1-\mu_h)\,\mathbf{\Pi}^{\{h\}}\nabla\boldsymbol{\gamma} + (\mathbf{I}-\boldsymbol{\lambda}_h)\left[\mathbf{\Pi}^h - \widetilde{\mathbf{\Pi}}_h^{(\,)}\right]\left[\mathbf{\Pi}^{\{k+1\}}\nabla\boldsymbol{\gamma}\right] \\
&\quad + (\mu_h\mathbf{I} - \boldsymbol{\lambda}_h)\left[\mathbf{\Pi}^h - \widetilde{\mathbf{\Pi}}_h^{(\,)}\right]\mathbf{x}_T \\
&\quad + (\boldsymbol{\lambda}_h - \mu_h\mathbf{I})\left[\mathbf{\Pi}^h - \widetilde{\mathbf{\Pi}}_h^{(*)}\right]\left(\mathbf{x}_T + \mathbf{\Pi}^{\{k+1\}}\nabla\boldsymbol{\gamma}\right) + O_p\left(\widehat{\mathbf{\Delta}}_\Pi\right) \\
&= (1-\mu_h)\,\widetilde{\mathbf{e}}_{T+h|T}^{(\,)} + \mu_h\widetilde{\mathbf{e}}_{T+h|T}^{(*)} \\
&\quad + (1-\mu_h)\left[\mathbf{\Pi}^{\{h\}} + \left(\mathbf{\Pi}^h - \widetilde{\mathbf{\Pi}}_h^{(\,)}\right)\mathbf{\Pi}^{\{k+1\}}\right]\nabla\boldsymbol{\gamma} \\
&\quad + O_p\left(\widehat{\mathbf{\Delta}}_\Pi\right)
\end{aligned}
$$