Multistep Forecasting in the Presence of Location Shifts

August 7, 2014

Abstract

This paper studies the properties of iterated and direct multistep forecasting techniques in the presence of in-sample location shifts (breaks in the mean). It also considers the interaction of these techniques with multistep intercept corrections that are designed to exhibit robustness to the shifts. In a local-asymptotic parameterization for the probability of breaks, we provide analytical expressions for forecast biases and mean-square forecast errors. We also provide simulations which show that breaks provide a rationale for using other methods than iterated multistep. In particular, we study how the relative accuracy of the methods relates to the forecast horizon, the sample size and the timing of the shifts. We show that direct multistep forecasting provides forecasts that are relatively robust to breaks and that its benefits increase with the forecast horizon. In an empirical application, we revisit an oft-used dataset of G7 macroeconomic series and corroborate our theoretical results.
Keywords: Multistep forecasting, Location Shifts, Local Asymptotics, Intercept Correction.

JEL *Classification*: C32, C53, E37.

1 Introduction

When a forecaster 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 multistep, IMS) or *directly* modeling the relation between the end-of-sample observation and its hth successor in order to forecast the latter (direct multistep, DMS). The direct technique has a long pedigree but it was originally thought that it brought little benefit, until Weiss (1991) found asymptotic relevance in matching estimation and forecast efficiency criteria, in particular in the presence of model misspecification. From then, many authors have produced theoretical analyses. In particular Tiao and Xu (1993), Clements and Hendry (1996), Chevillon and Hendry (2005) and Proietti (2011) 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) and McElroy and Wildi (2013) focused on the frequency domain, Schorfheide (2005) allowed for asymptotically vanishing misspecification. More recently Findley, Pötscher, and Wei (2004), Ing (2003, 2004) and Ing, Lin and Yu (2009) derived results for very general settings, see Elliott and Timmermann (2008) for an exposition of the problem and 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 do not appear clearly in practice, see e.g. the references in Chevillon (2007) as well as the more recent evidence in Eklund and Karlsson (2005), Marcellino, Stock, and Watson (2006), Jordà and Marcellino (2010), Schumacher and Breitung (2008) and Pesaran, Pick, and Timmermann (2011). By contrast, Tiao and Tsay (1994) and Tsay (1993) do find significant benefits in using DMS. Yet, many authors who find evidence against DMS use 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—relative to other countries, but instabilities were still present, see the discussion in Stock and Watson (2004) and Clark and McCracken (2008). Several theoretical analyses—as in Peña (1994) for breaks, Bhansali (1997) whose framework of long memory can be seen in the light of Diebold and Inoue (2001) and Perron and Qu (2007) as relevant for regularly occurring shifts, 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 Chevillon (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 finds substantial evidence sustaining the use of DMS in multivariate models: it outperforms iterated methods, unless the latter are 'intercept corrected' as in Clements and Hendry (1998).

Hence, we propose in this paper to analyze the theoretical properties of DMS forecasting in the presence of breaks. Clements and Hendry (1999, 2006) 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 step dummies. This result has been confirmed by Pesaran and Timmermann (2005) and Pesaran et al. (2011). In view of this evidence, we focus in this paper on breaks that affect the unconditional expectation and we parameterize them as arising from a shift in the intercept of an autoregressive process.

The present paper mostly relates to Pesaran and Timmermann (2005), referred to as PT. The differences are that these authors consider one-step forecasts under general models, AR(p) data generating processes (DGP) and various types of breaks. Here we focus in the theory on an AR(1) DGP that undergoes intercept shifts and consider forecasting from estimated AR(p) models; we address the issue of multistep forecasts via iterated and direct methods. The location shifts that we consider are occasional and it seems reasonable to assume that there are few of these, if any, in a sample consisting of, say, less than a hundred observations. We hence resort to a local asymptotic parameterization proposed by Leipus and Viano (2003) and used also by Diebold and Inoue (2001) and Perron and Qu (2007): we specify the probability of a shift at every period as a function of the observable sample and analytically derive the distributions of estimators of forecasts that result. As a complement to PT, we are hence able to provide analytical results both on the forecast bias and the mean-square forecast error whereas PT only compute the latter via simulation. Also, we assess the properties of intercept correction since this method is designed specifically to achieve some robustness to this sort of break.

The conclusion of this paper is that there exists a rationale for direct multistep estimation and forecasting in the presence of breaks. In particular the benefits in using DMS are stronger at longer horizons when the break is relatively recent. It is also worth using DMS forecasting under estimated autoregressive models of moderate order. The empirical analysis shows that DMS performs at least as well as IMS in more than 50% of the cases considered. By contrast, our results do not find any systematic benefits from intercept correction at multistep horizons.

The plan of this paper is as follows: we first describe the data generating process and consider the properties of estimators. We then provide analytical results for the resulting forecasts. Section 4 presents simulation results. In Section 5, we proceed to an empirical application. Additional extensive empirical results are reported in a supplement available from the author's website. Throughout the paper we use the following notation: for any real numbers a and b, $a \wedge b$ and $a \vee b$ denote respectively the minimum and maximum of a and b; $\lfloor x \rfloor$ denotes the integer part of any real scalar x.

2 The Data Generating Process and Estimators

This section defines a univariate data generating process (DGP) which allows for stochastic location shifts. We then consider the distributions of least squares estimators.

2.1 A model of stochastic breaks

We consider for our analysis the suggestion of infrequent shocks by Balke and Fomby (1991) and frame it within a local-asymptotic parameterization which has also been used by Diebold and Inoue (2001) in the context of long memory, Perron and Qu (2007) who study the properties of the correlogram in the presence of shifts, and Chevillon (2014) for modeling weak deterministic trends. The data generating process is autoregressive of order one, AR(1), with a random intercept:

$$y_t = \tau_t + \rho y_{t-1} + \epsilon_t, \tag{1}$$

for $t = 1, ..., T + h$ and where the error process ϵ_t is white noise with variance σ_ϵ^2. Here τ_t is written as the partial sum of an *i.i.d.* process:

$$\tau_t = \sum_{i=1}^{t} q_i v_i, \tag{2}$$

where $\nu_t \sim iid\left(0, \sigma_v^2\right)$, $q_t \overset{iid}{\sim} Bernoulli\left(\pi_T\right)$, and q_i, v_j are independent for all i and j. The variable q_t which follows a Bernoulli distribution indicates whether a break occurs at time t, and ν_t represents the magnitude of the break at that instant. The probability that a break occurs is parameterized as π_T, a function of the sample size. This implies that the process under consideration is a triangular array as has been commonly assumed in the econometric literature in the context of Pitman drifts and near-unit root asymptotics (since Bokboski, 1983, Phillips, 1987, and Chan and Wei, 1987). In practice, we assume that there exists $\pi \in (0, T)$ such that

$$\pi_T = \frac{\pi}{T}, \tag{3}$$

so π is the expected number of breaks in a sample of T observations. Condition (3) ensures that τ_t satisfies a functional central limit theorem (FCLT). Leipus and Viano (2003) and Georgiev (2002) show that, for $r \in [0, 1]$

$$\tau_{\lfloor rT \rfloor} \Rightarrow J\left(r\right), \tag{4}$$

where[1] J is a compound Poisson process with jump intensity π defined as $J\left(r\right) = \sum_{i=1}^{N(r)} v_i$, with $N\left(r\right)$ a counting process which has on average as many jumps as $\tau_{\lfloor rT \rfloor}$.

An important modification to the previous FCLT, expression (4), is to work under the condition that at least one break occurs in a sample of T observations. Georgiev also proved the convergence of $\tau_{\lfloor rT \rfloor}$ conditional on knowing the timing and/or the number of breaks. He shows that it possible to replace expression (4) with a different compound Poisson process. In this paper, the event we consider is the following:

$$\mathcal{E}_{c_k} : \{q_i\} = 1_{\{i = \lfloor c_k T \rfloor\}},$$

for $c_k \in (0, 1)$ and where $1_{\{\cdot\}}$ denotes the indicator function. We denote the realization:

$$v_{\lfloor c_k T \rfloor} = \gamma.$$

[1] $\Rightarrow$ denotes weak convergence of the associated probabiility measure under the Skorokhod topology.

Working under event $\mathcal{E}_{c_k}$ allows to analyze the impact of the timing of the break. The variance of v_t then plays no role since $q_t v_t$ is identically zero except at $t = \lfloor c_k T \rfloor$. The corresponding distribution is

$$\tau_{[rT]} | \mathcal{E}_{c_k} \Rightarrow J_+ (r) .$$

Under the conditions above, Georgiev (2002, Corollary 5 and Example 1) proves the joint convergence:

$$\begin{bmatrix} T^{-1} \sum_{t=1}^{T} y_t \\ T^{-1} \sum_{t=1}^{T} y_t^2 \\ T^{-1} \sum_{t=1}^{T} y_{t-1} y_t \end{bmatrix} \Rightarrow \begin{bmatrix} \frac{1}{1-\rho} \int_0^1 J(r)\, dr \\ \frac{\sigma_\epsilon^2}{1-\rho^2} + \frac{1}{(1-\rho)^2} \int_0^1 J^2(r)\, dr \\ \frac{\rho\sigma_\epsilon^2}{1-\rho^2} + \frac{1}{(1-\rho)^2} \int_0^1 J^2(r)\, dr \end{bmatrix},$$

with corresponding distributions conditional on event $\mathcal{E}_{c_k}$, replacing J with J_+.

2.2 Estimators

We consider the so-called direct multi-step (DMS) estimators $(\widetilde{\tau}_h, \widetilde{\rho}_h)$ obtained by ordinary least-squares (OLS) projection of y_t on an intercept and its hth lag, y_{t-h}, for $h \geq 1$. When $h = 1$, the estimators reduce to the standard OLS of an AR(1), which we denote by $(\widehat{\tau}, \widehat{\rho})$. To derive the asymptotic distribution of the DMS estimators, we express the horizon h as a fraction of the sample size, i.e. letting

$$h = \lfloor c_h T \rfloor .$$

In the absence of breaks, the intercept τ_t is constantly zero, so the correct model is an AR(1) without drift and $(\widetilde{\tau}_h, \widetilde{\rho}_h) \xrightarrow{p} \left(0, \rho^h\right)$ (with distribution given in, e.g., Chevillon and Hendry, 2005) Under the DGP delineated above, by contrast, the asymptotic distribution of the estimators is given by the following proposition:

Proposition 1 *Assume the DGP is given by (1), (2) and (3), and that one or more shifts almost surely occur. The OLS estimators of the projection of y_t on $(1, y_{t-h})$ satisfy:*

· *if $h = 1$,*

$$\begin{bmatrix} \widehat{\tau} \\ \widehat{\rho} - \rho \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & \frac{\int_0^1 J(r)dr}{1-\rho} \\ \frac{\int_0^1 J(r)dr}{1-\rho} & \frac{\sigma_\epsilon^2}{1-\rho^2} + \frac{\int_0^1 J^2(r)dr}{(1-\rho)^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\int_0^1 J(r)dr}{1-\rho} \\ \frac{\int_0^1 J^2(r)dr}{(1-\rho)^2} \end{bmatrix};$$

· *if* $h = \lfloor c_h T \rfloor$,

$$\begin{bmatrix} \widetilde{\tau}_h \\ \widetilde{\rho}_h - \rho^h \end{bmatrix} \Rightarrow \begin{bmatrix} 1 - c_h & \frac{\int_0^{1-c_h} J(r)dr}{1-\rho} \\ \frac{\int_0^{1-c_h} J(r)dr}{1-\rho} & \frac{\sigma_\epsilon^2(1-c_h)}{1-\rho^2} + \frac{\int_0^{1-c_h} J^2(r)dr}{(1-\rho)^2} \end{bmatrix}^{1} \begin{bmatrix} \frac{\int_0^{1-c_h} J(r+c_h)dr}{1-\rho} \\ \frac{\int_0^{1-c_h} J(r)J(r+c_h)dr}{(1-\rho)^2} \end{bmatrix},$$

with corresponding distributions conditional on event $\mathcal{E}_{c_k}$, *replacing* J *with* J_+.

Conditionally on $\mathcal{E}_{c_k}$, the asymptotic distributions of the estimators are nonstochastic and we denote them as:[2]

$$(\widehat{\tau}, \widehat{\rho}) \underset{T\to\infty}{\overset{p}{\to}} (\widehat{\tau}_\infty, \widehat{\rho}_\infty), \quad \text{and} \quad (\widetilde{\tau}_h, \widetilde{\rho}_h) \underset{T\to\infty}{\overset{p}{\to}} (\widetilde{\tau}_{\infty,c_h}, \widetilde{\rho}_{\infty,c_h}).$$

We define $\mu_y = \gamma / (1-\rho) = \lim_{h\to\infty} \mathsf{E}[y_{T+h} | \mathcal{E}_{c_k}]$ the post-break long-run mean of y_t, and $\sigma_y^2 = \sigma_\epsilon^2 / \left(1 - \rho^2\right)$. The following corollary provides the limits of the estimators.

Corollary 2 *The asymptotic limits given in Proposition 1 conditional on event* $\mathcal{E}_{c_k}$ *write*

$$\begin{bmatrix} \widehat{\tau}_\infty \\ \widehat{\rho}_\infty \end{bmatrix} = \begin{bmatrix} 1 & c_k\mu_y \\ c_k\mu_y & \sigma_y^2 + c_k\mu_y^2 \end{bmatrix}^{-1} \begin{bmatrix} c_k\mu_y \\ \rho\sigma_y^2 + c_k\mu_y^2 \end{bmatrix},$$

and

$$\begin{bmatrix} \widetilde{\tau}_{\infty,c_h} \\ \widetilde{\rho}_{\infty,c_h} \end{bmatrix} = \begin{bmatrix} 1 - c_h & [(c_k - c_h) \vee 0]\,\mu_y \\ [(c_k - c_h) \vee 0]\,\mu_y & [(c_k - c_h) \vee 0]\,\mu_y^2 \end{bmatrix}^{-1} \begin{bmatrix} [((1-2c_h) \wedge (c_k - c_h)) \vee 0]\,\mu_y^2 \\ ((1-c_h) \wedge c_k)\,\mu_y \end{bmatrix}.$$

The expression from the previous corollary are difficult to interpret, so we next turn to the two limiting situations where the horizon is small relative to the sample size or where the break occurs toward the very end of the sample.

Corollary 3 *Under the assumptions of Proposition 1, and the notations of Corollary 2, the asymptotic distributions of the estimators satisfy, conditionally on* $\mathcal{E}_{c_k}$:

(i) As $c_k \to 0$, $(\widehat{\tau}_\infty, \widehat{\rho}_\infty) \sim (\gamma c_k, \rho)$,

and if $c_k < c_h$, $(\widetilde{\tau}_{\infty,c_h}, \widetilde{\rho}_{\infty,c_h}) \sim \left(\mu_y \frac{c_k}{1-c_h}, 0\right)$.

(ii) As $c_h \to 0$ *such that* $c_h < c_k$, $(\widetilde{\tau}_{\infty,c_h}, \widetilde{\rho}_{\infty,c_h}) \sim \left(\mu_y, \frac{\mu_y^2}{\sigma_y^2}\right) c_k$.

The corollary shows that when the break occurs toward the very end of the sample, it does not affect the one-step OLS estimators which are then consistent. By contrast the multistep estimators are affected. When the horizon is very small relative to the sample

[2] We assume a projection facility, see Marcet and Sargent (1989), such that the modeler changes the sample if the estimated parameter $\widehat{\rho} \geq 1$ to ensure that $|\widehat{\rho}| < 1$.

size, the multistep intercept estimator tends to the unconditional expectation of the sample mean of y_t, weighted by the number of observations used in the multistep estimation. The slope estimator is then nonzero but is a function of the ratio of the squared expectation of y_T over its variance.

3 Forecasts

We now consider the implications of previous results for the distributions of forecasts at horizon $h > 1$. We compare four forecasting techniques. Two are based on the AR(1), but with different estimation methods: the forecasts obtained from rolling forward using the estimators $(\widehat{\tau}, \widehat{\rho})$ are referred to as iterated multistep (IMS). Alternatively, Direct Multistep (DMS) forecasts are based on using $(\widetilde{\tau}_h, \widetilde{\rho}_h)$. In the light of Pettenuzzo and Timmermann (2011) who confirm Clements and Hendry (2006) and Giacomini and Rossi (2009) in emphasizing the importance of model instability over parameter estimation and model uncertainty for forecast breakdowns, it appears that multistep intercept correction could yield substantial forecast accuracy improvements at short horizon.

Two additional methods are examined: their aim is to put the forecast 'back on track' (i.e. reducing or suppressing the forecast bias) 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, 2005, for an analysis of its use in forecasting with econometric models and Clark and McCracken, 2008, for an extensive study of their empirical benefits and that of variants thereof. We refer to these methods as one-step and multistep intercept corrections depending which sample residuals are used. One-step intercept correction necessarily uses the one-step model on which IMS is based (and we refer to it as IC), whereas multistep can relate to either of IMS or DMS (hence denoted by IMSIC or DMSIC).

Throughout, we assume that the horizon is less than $T/2$ so $h < T - h$. We compute distributions conditionally on $\mathcal{E}_{c_k}$ and assume it implicit when we refer to the mean-square forecast errors (MSFE). As the results depend on the magnitude of the break vis-à-vis the standard deviation of the errors, we define the following ratio:

$$\lambda = \mu_y^2 / \sigma_y^2. \tag{5}$$

3.1 IMS and DMS Forecasts

We first consider the situation where the process, absent the location shifts, is correctly specified as an AR(1) (if $\rho \neq 0$).

The iterated multistep ahead forecasts are obtained from the OLS estimators: $\widehat{y}_{T+h|T} = \widehat{\rho}_{\{h\}}\widehat{\tau} + \widehat{\rho}^h y_T$, where $\widehat{\rho}_{\{h\}} = \sum_{j=0}^{h-1} \widehat{\rho}^j$. The resulting forecast errors $\widehat{e}_{T+h|T} = y_{T+h} - \widehat{y}_{T+h|T}$ are equal to:

$$\widehat{e}_{T+h|T} = \underbrace{\rho_{\{h\}}\tau_T - \widehat{\rho}_{\{h\}}\widehat{\tau}}_{\substack{\text{intercept} \\ \text{estimation}}} + \underbrace{\left(\rho^h - \widehat{\rho}^h\right) y_T}_{\substack{\text{slope} \\ \text{estimation}}} + \underbrace{\sum_{i=0}^{h-1} \rho^i \left(\tau_{T+h-i} - \tau_T\right)}_{\text{future breaks}} + \underbrace{\sum_{i=0}^{h-1} \rho^i \epsilon_{T+h-i}}_{\text{future errors}},$$

which we have decomposed into the effects of the estimation uncertainty surrounding multistep intercept and slope parameters, the impact of future breaks and of future errors. Since the shifts are independent over time and uncorrelated with the errors ϵ_t, the impact of future breaks is orthogonal to the other elements. Hence we assume for forecast comparison that there are no breaks over the forecast horizon, i.e. setting $\tau_{T+i} = \tau_T$ for all $i > 1$.

The DMS forecasts are generated as

$$\widetilde{y}_{T+h|T} = \widetilde{\tau}_h + \widetilde{\rho}_h y_T,$$

with corresponding forecast errors $\widetilde{e}_{T+h|T} = y_{T+h} - \widetilde{y}_{T+h|T}$. The latter differ from $\widehat{e}_{T+h|T}$ only via intercept and slope estimation errors (which may interact with the forecasting origin, see Ing, 2004). Conditional on event $\mathcal{E}_{c_k}$, the forecast origin satisfies:

$$y_T \underset{T\to\infty}{\overset{L}{\to}} \mathsf{N}\left(\frac{J_+(1)}{1-\rho}, \sigma_y^2\right).$$

We define the random variable with zero conditional mean:

$$Y = \lim_{T\to\infty} \left(y_T - \frac{J_+(1)}{1-\rho}\right). \tag{6}$$

so that we can write that conditional on $\mathcal{E}_{c_k}$,

$$y_T \Rightarrow \frac{J_+(1)}{1-\rho} + Y.$$

Conditional on $\mathcal{E}_{c_k}$, we saw previously that the estimators converge towards a nonstochastic limit. The following proposition formulates our results concerning the conditional moments of the forecast errors.

Proposition 4 *Under the assumptions of Proposition 1, the moments of the forecast errors, conditional on* $\mathcal{E}_{c_k}$*, satisfy as* $T \to \infty$*,*

$$\mathsf{E}\left(\widehat{e}_{T+h|T}|\mathcal{E}_{c_k}\right) \to \mu_y - \frac{\widehat{\tau}_\infty}{1-\widehat{\rho}_\infty}, \quad \mathsf{Var}\left(\widehat{e}_{T+h|T}|\mathcal{E}_{c_k}\right) \to \sigma_y^2;$$

and,

if $c_h < c_k : \mathsf{E}\left(\widetilde{e}_{T+h|T}|\mathcal{E}_{c_k}\right) \to \mu_y\left(1-\widetilde{\rho}_{\infty,c_h}\right) - \widetilde{\tau}_{\infty,c_h}, \quad \mathsf{Var}\left(\widetilde{e}_{T+h|T}|\mathcal{E}_{c_k}\right) \to \left(1+\widetilde{\rho}^2_{\infty,c_h}\right)\sigma_y^2,$

if $c_k \leq c_h : \mathsf{E}\left(\widetilde{e}_{T+h|T}|\mathcal{E}_{c_k}\right) \to \mu_y\left(0 \vee \left(1-\frac{c_k}{1-c_h}\right)\right), \quad \mathsf{Var}\left(\widetilde{e}_{T+h|T}|\mathcal{E}_{c_k}\right) \to \sigma_y^2.$

3.2 Intercept corrections

The modeler may suspect that a break has occurred. She could base her 'suspicion' either on her knowledge of breaks in other parts of the economy or, reasonably, on a break test. Difficulties arise if she is to use the result of her test in her forecasting model: as mentioned in Pesaran and Pick (2011), she needs an accurate estimate of the break date, break amplitude, and possibly the nature of the break (in the intercept, first-order autocorrelation...). Given that break tests such as Bai and Perron (1998) require trimming at the beginning and end of the sample, the uncertainty surrounding the break may be substantive in finite samples and the resulting forecasts of poor quality, see Elliott (2005) and Paye and Timmermann (2006), which we confirm in Section 5. Hence, as discussed in Clark and McCracken (2008), the modeler might prefer to resort to robust methods that do not need information about the breaks: she may for instance wish to intercept-correct the model to produce a more robust or less biased forecast, see Clements and Hendry (1998), Chapter 8. An alternative would be to use break techniques that do not require trimming, as for instance the Impulse Indicator Saturation (IIS) method of Santos, Hendry, and Johansen (2008) and Johansen and Nielsen (2009). Yet, timing the shift always results in uncertainty.

Hence, if the forecaster has no hypothesis about the date of the break, then she may resort to intercept correction (IC) based on the latest observation, i.e. using the correction $\delta_{IC} = y_T - \widehat{y}_{T|T-1}$. This yields the one-step IC forecast error:

$$\widehat{e}^{IC}_{T+1|T} = y_{T+1} - \widehat{y}^{IC}_{T+1|T} = y_{T+1} - \left(\widehat{y}_{T+1|T} + \delta_{IC}\right),$$

where $\delta_{IC} = \tau_T - \widehat{\tau} + (\rho - \widehat{\rho})\, y_{T-1} + \epsilon_T$. Rolling forward the correction leads to the IMSIC forecast: $\widehat{y}^{IMSIC}_{T+1|T} = \widehat{y}^{IC}_{T+1|T}$, and for $h \geq 2 : \widehat{y}^{IMSIC}_{T+h|T} = \widehat{\tau} + \widehat{\rho}\widehat{y}^{IMSIC}_{T+h-1|T}$ in which case

$$\delta_{IMSIC} = \widehat{y}^{IMSIC}_{T+1|T} - \widehat{y}_{T+1|T} = \widehat{\rho}^{h-1}\delta_{IC}. \tag{7}$$

A natural extension in the context of DMS forecasting is to generate the forecast $\widetilde{y}^{DMSIC}_{T+h|T} = \widetilde{y}_{T+h|T} + \delta_{DMSIC}$ using the correction:

$$\delta_{DMSIC} = y_T - \widetilde{y}_{T|T-h} = \sum_{j=0}^{h-1} \rho^j \tau_{T-j} - \widetilde{\tau}_h + \left(\rho^h - \widetilde{\rho}_h\right) y_{T-h} + \sum_{j=0}^{h-1} \rho^j \epsilon_{T-j}. \tag{8}$$

Absent the issue of parameter estimation (i.e. if the parameters were known), the point of intercept correction is to provide unbiased forecasts under location shifts. Yet, parameter estimation may induce biases as the following proposition shows.

Proposition 5 *Under the assumptions of Proposition 1, the moments of the forecast errors, conditional on $\mathcal{E}_{c_k}$, satisfy as $T \to \infty$:*

$$\mathsf{E}\left(\widehat{e}^{IMSIC}_{T+h|T} | \mathcal{E}_{c_k}\right) \to \mu_y - \frac{\widehat{\tau}_\infty}{1 - \widehat{\rho}_\infty}, \quad \textit{and} \quad \mathsf{V}\left(\widehat{e}^{IMSIC}_{T+h|T} | \mathcal{E}_{c_k}\right) \to \sigma_y^2.$$

and

$$\mathsf{E}\left(\widetilde{e}^{DMSIC}_{T+h|T} | \mathcal{E}_{c_k}\right) \to 0, \quad \textit{and} \quad \mathsf{Var}\left(\widetilde{e}^{DMSIC}_{T+h|T} | \mathcal{E}_{c_k}\right) \to \begin{cases} \left(1 + 2\widetilde{\rho}^2_{\infty,c_h}\right)\sigma_y^2, & \textit{if } c_h < c_k, \\ \sigma_y^2, & \textit{if } c_k \leq c_h. \end{cases}$$

The proposition shows that DMSIC is asymptotically unbiased and that the MSFE of IMSIC does not depend on the horizon c_h. Also, IMSIC controls the variance of the forecast error. So does DMSIC, but only at long horizons (relative to k).

3.3 Forecast accuracy comparisons

We now consider the implications of the previous propositions for the asymptotic conditional MSFEs which we denote $\mathsf{MSFE}_{\infty}^{a}$ where $a \in \{IMS, DMS, IMSIC, DMSIC\}$ denotes the technique. Propositions 4 and 5 show that $\mathsf{MSFE}_{\infty}^{IMS}$ and $\mathsf{MSFE}_{\infty}^{IMSIC}$ do not depend on the forecast horizon c_h and, crucially that, as $\gamma \to \infty$, all MSFEs diverge except $\mathsf{MSFE}_{\infty}^{DMSIC}$ which tends to

$$\mathsf{MSFE}_{\infty}^{DMSIC} \underset{\gamma\to\infty}{\to} \left[1 + 2\left(1 - \frac{c_h}{1-c_k}\right)^2\right]\sigma_y^2 < \infty. \tag{9}$$

Hence DMSIC constitutes a technique that is robust to large breaks. DMS also shares the same property, but only when $c_k \leq c_h$. Indeed as $\gamma \to \infty$ the bias of DMS becomes zero but its second moment is only finite when $\widetilde{\rho}_{\infty} \to 0$, i.e. if $c_k \leq c_h$, in which case:

$$\mathsf{MSFE}_{\infty}^{DMS} \underset{\gamma\to\infty,\ c_k\leq c_h}{\to} \sigma_y^2.$$

Now, considering medium-sized breaks, so μ_y^2 and σ_y^2 have similar magnitudes, the following corollary provides a comparison when either $c_k \to 0$ or $c_h \to 0$.

Corollary 6 *Under the assumptions of Proposition 1, the asymptotic conditional MSFEs satisfy,*

$$\mathsf{MSFE}_{\infty}^{IMS} = \mathsf{MSFE}_{\infty}^{IMSIC} = \left[1 + \lambda\left(1-c_k\right)^2\right]\sigma_y^2,$$

As $c_k \to 0$ *with* $c_h \geq c_k$:
$\mathsf{MSFE}_{\infty}^{DMS} \underset{c_k\to 0}{\sim} \left[1 + \left(1 - \frac{c_k}{1-c_h}\right)^2\lambda\right]\sigma_y^2,$ *and* $\mathsf{MSFE}_{\infty}^{DMSIC} \underset{c_k\to 0}{\sim} \sigma_y^2;$
as $c_h \to 0$ *with* $c_k > c_h$:
$\mathsf{MSFE}_{\infty}^{DMS} \underset{c_h\to 0}{\sim} \left(1 + \left(1 - \frac{1}{1+(1-c_k)c_k\lambda}\right)^2 + \left[1 - \frac{\lambda}{1+(1-c_k)c_k\lambda}\left(1-c_k\right)c_k\right]^2\left(1-c_k\right)^2\lambda\right)\sigma_y^2,$
$\mathsf{MSFE}_{\infty}^{DMSIC} \underset{c_h\to 0}{\sim} \left(1 + 2\frac{\lambda^2}{[1+(1-c_k)c_k\lambda]^2}c_k^2\right)\sigma_y^2.$

The corollary allows to compare the MSFEs under the two limits. First, for recent breaks, as $c_k \to 0$ it must hold that $c_k/\left(1-c_h\right)$ becomes smaller than unity; hence the following ranking

$$c_k \to 0 : \mathsf{MSFE}_{\infty}^{DMSIC} < \mathsf{MSFE}_{\infty}^{DMS} \leq \mathsf{MSFE}_{\infty}^{IMS} = \mathsf{MSFE}_{\infty}^{IMSIC}, \tag{10}$$

where equality between $\mathsf{MSFE}_{\infty}^{DMS}$ and $\mathsf{MSFE}_{\infty}^{IMS}$ holds in the limit for $c_k = 0$. This shows that direct multi-step techniques are to be preferred in the presence of recent breaks.

Now considering the horizon, the corollary shows that as $c_h \to 0$, the actual value of the horizon does not matter for the MSFEs. The bias-variance trade-off is apparently difficult to assess from the formulae since it depends on the relative magnitudes of μ_y^2 and σ_y^2. Indeed in terms of bias, DMSIC is most accurate since it is asymptotically unbiased, then DMS then IMS and IMSIC; the ordering is reversed in terms of variance. As $c_h \to 0$, if $(1 - c_k)\, c_k \lambda >> 1$, i.e. the break is large and does not occur towards the very beginning or end of the sample, then

$$\begin{aligned}
\mathsf{MSFE}_{\infty}^{DMS} &\underset{c_h \to 0}{\sim} \left[1 + \left(1 - \frac{1}{(1 - c_k)\, c_k \lambda}\right)^2\right] \sigma_y^2, \\
\mathsf{MSFE}_{\infty}^{DMSIC} &\underset{c_h \to 0}{\sim} \left(1 + 2\frac{1}{(1 - c_k)^2\, \lambda^2}\right) \sigma_y^2.
\end{aligned}$$

This implies the following ordering as $c_h \to 0$,

$$(1 - c_k)\, c_k \lambda >> 1 : \mathsf{MSFE}_{\infty}^{DMSIC} \underset{c_h \to 0}{\leq} \mathsf{MSFE}_{\infty}^{DMS} \underset{c_h \to 0}{\leq} \mathsf{MSFE}_{\infty}^{IMS} \underset{c_h \to 0}{=} \mathsf{MSFE}_{\infty}^{IMSIC}. \quad (11)$$

Now for breaks that are small, very recent or that occur at the beginning of the sample, so $(1 - c_k)\, c_k \lambda << 1$, then

$$\begin{aligned}
\mathsf{MSFE}_{\infty}^{DMS} &\underset{c_h \to 0}{\sim} \left[1 + \left[1 - (1 - c_k)\, c_k \lambda\right]^2 (1 - c_k)^2 \lambda\right] \sigma_y^2, \\
\mathsf{MSFE}_{\infty}^{DMSIC} &\underset{c_h \to 0}{\sim} \left(1 + 2\lambda^2 c_k^2\right) \sigma_y^2.
\end{aligned}$$

This implies that $\mathsf{MSFE}_{\infty}^{DMS}$ is smaller than $\mathsf{MSFE}_{\infty}^{IMS}$ but only marginally so since $\left[1 - (1 - c_k)\, c_k \lambda\right]^2 \approx 1$. And then DMSIC is only more accurate than the other techniques if $\lambda < \frac{1}{2}\left(\frac{1 - c_k}{c_k}\right)^2$. Hence, for small shifts, all methods are approximately equivalent, with a slight advantage to DMSIC and possibly DMS.

The analysis above shows that the possibility of location shifts provides a rationale for deviating from IMS and using models that reduce the forecast bias.

3.4 Higher order autoregressions

The AR(1) model above is restrictive. A natural extension consists in considering estimation and forecasting using a more general, possibly misspecified, AR(p), model. To present the results while controlling the additional complexity, we focus on two cases, keeping the DGP as before: (i) the estimated model is an AR(2); and (ii) the estimated model is an AR(p) with $p = \lfloor c_p T \rfloor$, $c_p < c_k$.

3.4.1 Estimated AR(2)

Consider, first, the case where the DGP is as before, but the following model is estimated by OLS:

$$y_t = \tau + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \epsilon_t,$$

where the DGP implies $(\rho_1, \rho_2) = (\rho, 0)$. Notice that $\widehat{\rho}_1 - \rho_1$ and $\widehat{\rho}_2 - \rho_2$ asymptotically coincide (an abuse of notation leads us to use $\widehat{\tau}$ for both models but the meaning should be clear each time, we also let $\int J = \int_0^1 J(r)\,dr$):

$$\begin{bmatrix} \widehat{\tau} \\ \widehat{\rho}_1 - \rho_1 \\ \widehat{\rho}_2 - \rho_2 \end{bmatrix} \Rightarrow \frac{1}{(1-\rho)\,\sigma_\epsilon^2 + 2\left[\int J^2 - \left(\int J\right)^2\right]} \begin{bmatrix} \sigma_\epsilon^2 \int_0^1 J \\ \int J^2 - \left(\int J\right)^2 \\ \int J^2 - \left(\int J\right)^2 \end{bmatrix}. \tag{12}$$

Contrast with the case of an estimated AR(1) :

$$\begin{bmatrix} \widehat{\tau} \\ \widehat{\rho} - \rho \end{bmatrix} \Rightarrow \frac{1}{(1-\rho)\,\sigma_\epsilon^2 + (1+\rho)\left[\int J^2 - \left(\int J\right)^2\right]} \begin{bmatrix} \sigma_\epsilon^2 \int J \\ (1+\rho)\left[\int J^2 - \left(\int J\right)^2\right] \end{bmatrix}.$$

The previous expression shows that the intercept bias is lower under an estimated AR(2) rather than in an AR(1) . This translates into the one-step forecast errors, which, conditional on event $\mathcal{E}_{c_k}$ and in the absence of shifts over the forecast horizon, satisfy:

$$AR(2) : \widehat{e}_{T+1|T}^{AR(2)} - \epsilon_{T+1} \Rightarrow \frac{\sigma_\epsilon^2 (1-\rho-c_k) - 2c_k (1-c_k)\,\gamma\,(\gamma\rho + Y)}{(1-\rho)\,\sigma_\epsilon^2 + 2c_k (1-c_k)\,\gamma^2}\gamma, \tag{13}$$

$$AR(1) : \widehat{e}_{T+1|T}^{AR(1)} - \epsilon_{T+1} \Rightarrow \frac{\sigma_\epsilon^2 (1-\rho-c_k) - (1+\rho)\,c_k (1-c_k)\,\gamma\,(\gamma\rho + Y)}{(1-\rho)\,\sigma_\epsilon^2 + (1+\rho)\,c_k (1-c_k)\,\gamma^2}\gamma.$$

The only difference between the two is the coefficient $(1+\rho)$ under correct AR(1) specification which becomes 2 for the AR(2). This entails a reduction in the forecast bias when $\gamma > 0$ and $c_k \to 0$ using an AR(2). The implications for IMS are that the estimated long-run mean, which under Proposition 4 governs forecast accuracy, is here $\widehat{\tau}/\left(1-\widehat{\rho}_1-\widehat{\rho}_2\right) = \widehat{\tau}/\left[1-\rho-\left(\widehat{\rho}_1-\rho_1\right)+\left(\widehat{\rho}_2-\rho_2\right)\right]$. Conditional on $\mathcal{E}_{c_k}$, this asymptotically implies:

$$\lim_{T\to\infty} \mathsf{E}\left(\widehat{e}_{T+h|T}^{AR(2)}|\mathcal{E}_{c_k}\right)/\mu_y < \lim_{T\to\infty} \mathsf{E}\left(\widehat{e}_{T+h|T}^{AR(1)}|\mathcal{E}_{c_k}\right)/\mu_y,$$

i.e. the AR(2) model performs better for IMS forecasting than the AR(1) since the forecast variance is not affected asymptotically. Following Proposition 5, IMSIC will perform similarly.

Results similar to expression (12) hold for DMS under model $y_t = \tau_h + \rho_{h,1}y_{t-h} + \rho_{h,2}y_{t-h-1} + w_{h,t}$, with $h+1<k$ in the sense that $\widetilde{\rho}_{h,1}-\rho_{h,1}$ and $\widetilde{\rho}_{h,2}-\rho_{h,2}$ coincide and $(1+\rho)$ is then replaced by a 2 in the forecast errors. For DMS the forecast bias is directly given by an equivalent of (13) and hence using more lags reduces the forecast bias, yet at a cost of higher variance (as in Proposition 4). For DMSIC, Proposition 5 shows that the cost of the higher variance is magnified under an estimated AR(2) with little gain in terms of forecast bias.

The conclusions from this subsection are that in the presence of unmodeled location shifts, it is profitable in terms of forecast accuracy to estimate a model which uses more lags since this yields more degrees of freedom to capture the dynamics induced by the breaks. The next subsection explores the question of whether there is a benefit in extending the model further and include a *large* number of lags.

3.4.2 Large order AR

We now extend the previous analysis to the case where the estimated model is an AR(p) where $p = \lfloor c_p T \rfloor$, $0 < c_p < c_k$. The model is estimated by OLS and the estimators are denoted $\left(\widehat{\tau}, \widehat{\rho}_1, ..., \widehat{\rho}_p\right)'$. The one-step forecast error is then given as (as before we assume no shifts over the forecast horizon):

$$\widehat{e}_{T+1|T}^{AR(p)} = -\left(\widehat{\tau}-\tau_T\right) - \sum_{j=1}^{p}\left(\widehat{\rho}_j-\rho_j\right)y_{T-j+1} + \epsilon_{T+1}.$$

For ease of analytical expressions, we consider two cases where $\gamma << \sigma_\epsilon^2$ or $\gamma >> \sigma_\epsilon^2$, i.e. $\lambda << 1$ or $\lambda >> 1$, as this affects the matrix of empirical second moments. Conditional on event $\mathcal{E}_{c_k}$, we show in the appendix that

$$\lambda >> 1 : \widehat{e}^{AR(p)}_{T+1|T} - \epsilon_{T+1} \Rightarrow -(1-\rho) Y; \tag{14a}$$

$$\lambda << 1 : T^{-1}\left(\widehat{e}^{AR(p)}_{T+1|T} - \epsilon_{T+1}\right) \Rightarrow -2\gamma - \frac{\gamma^2}{2\sigma^2_\epsilon}(2c_k - c_p)\, c_p \left[\gamma + Y(1-\rho)\right]. \tag{14b}$$

Similar results hold for DMS estimation replacing c_k with $c_k - c_h$ as long as $c_h + c_p < c_k$.

This results shows that the benefits from long lag regressions are crucially only present in the presence of large unmodeled location shifts. If the shifts are present but small, then forecast errors may diverge under an estimated AR$(\lfloor c_p T \rfloor)$. Overall a forecaster may wish to consider increasing the lag order by a small amount, but she should resist using long lags.

4 Monte Carlo

This section considers forecasting the AR(1) DGP with stochastic breaks presented in Section 2. The expected frequency of breaks is set to $\pi = 1$ per sample. We consider parameter combinations[3] of $(\gamma, \rho) \in (0, 4] \times [-.9, .9]$, and specify the errors as standard Gaussian white noise. Forecasts over horizons of $h = 1$, 5, and 10 periods are computed using an estimated AR(p) model with $p \in \{1, 2, 4\}$ over a sample of $T = 50$ observations. We perform 5,000 Monte Carlo replications. We consider imposing, or not, the presence of at least one shift in every simulated sample: i.e. potentially drawing from the unconditional distribution, in which case samples exhibit no breaks with probability $(1 - \pi/T)^T \to e^{-\pi}$ as $T \to \infty$, which is here equal to 0.36. Also we do not impose that no shifts occur over the forecast horizon. We discuss below the modifications that arise from changing these assumptions. Figures 1 to 5 report the log ratios of square-root MSFEs (RMSFE) for the forecasting techniques considered.

[3] The simulations in Bai and Perron (2003) show some power (about 60%) for their test applied to an AR(1) process for a change in the intercept from 1 to 2 at mid-sample and $T = 100$ (the slope is 0.5 and errors are standard normal). Also the "tiny" target size for Impulse Indicator Saturation (IIS, see Santos, Hendry and Johansen, 2008) embedded in Oxmetrics 7.0 is 0.001, which corresponds to detecting outliers of an absolute value greater than $\gamma_0 \approx 3.3$. Hence values of $\gamma = 4$ correspond to values of a shift that should be detectable even when happening at the very end of the sample.

Figure 1 presents results at horizon $h = 1$ so IMS and DMS coincide there. It shows that intercept correction only improves forecast accuracy when the model is an AR(4) and the shift magnitude γ is low. This corresponds to the case where expression (14b) showed that the non intercept corrected forecasts errors may be large, so intercept correction seems to perform its purpose of "anchoring" the forecasts. When extending the forecast horizons to $h = 5$ as in Figures 2 and 3, for estimated models that are respectively an AR(1) and an AR(2), we see that IMS performs best overall, except when the process y_t presents sufficient persistence (ρ close to 1 or -1). Contrasting the figures, we see that DMS performs even better relative to all methods when the estimated model is an AR(2) (unreported results show that this is the case also for an estimated AR(4)). Forecasting at longer horizons, at $h = 10$ in Figures 4 and 5, we see that DMS performs best for a wider range of parameter values. In particular, DMS performs best for large γ and $|\rho|$ (this is also the case when the estimated model is an AR(4)).

In Figures 2 and 4 the differences between IMS and IMSIC are not very strong when γ is not too close to zero (so that there is indeed a shift to correct for). By contrast, Figures 3 and 5 show a substantial loss in using IMSIC: this is due to the fact that the model is not correctly specified (absent the shift) since the DGP is an AR(1) but the estimated model is an AR(2) . An issue arises with intercept correction in multistep ahead forecasting: putting the forecast 'back on track' requires also a correct specification of the underlying dynamics. DMSIC is more robust in this context.

In unreported simulations, we consider changing the parameter values above. Increasing the sample size to $T = 100$ does not qualitatively change the results, nor does increasing the frequency of shifts to $\pi = 5$. We also allow ϵ_t to follow an unmodeled MA(1) with parameter $\theta \in (-1, 1)$, the results are essentially the same to the exception that $\theta << 0$ improves the relative accuracy of IMSIC at long horizons, but that $\theta >> 0$ tends to reduce it. We also consider imposing that no shifts occur over the forecast horizon, the relative performances are mostly unaffected, except for the case of forecasting at $h = 5$ using an estimated AR(1) in which case DMS no longer shows the very good performance witnessed above when γ is low and ρ close to unity, but DMS then performs better than IMS for parameters combinations such that $|\rho|$ is not too close to unity and γ not too close to zero.

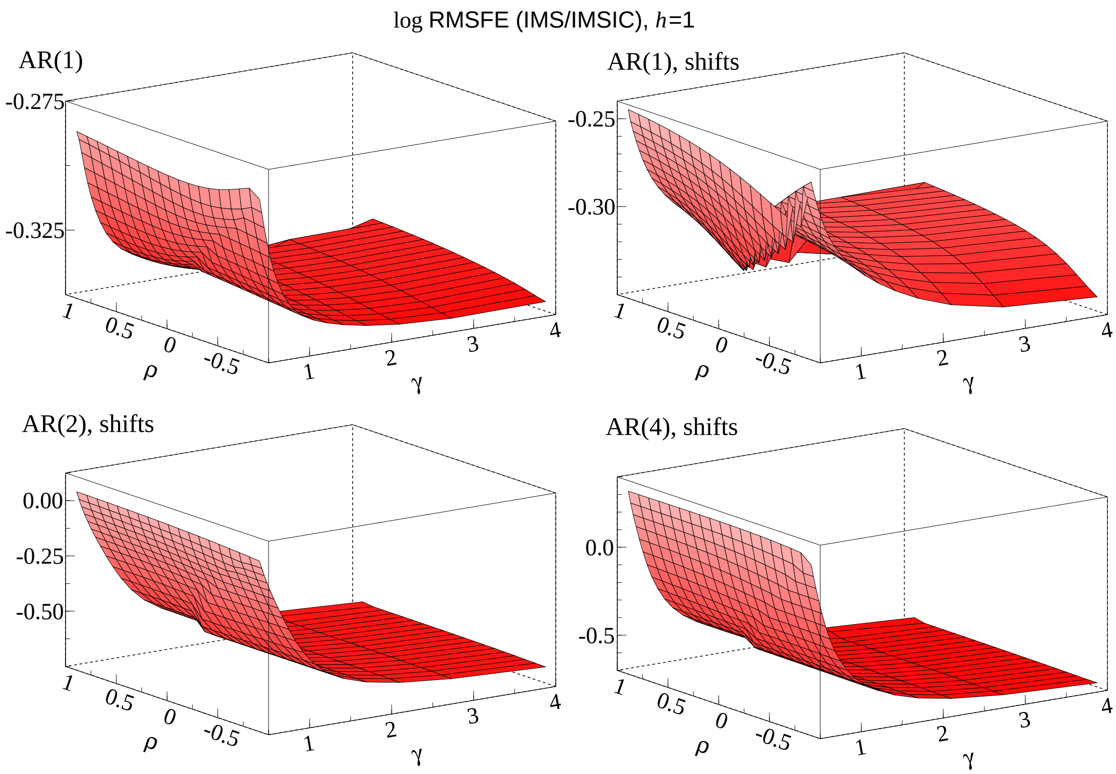

Figure 1: The figure presents the logarithms of the ratios of simulated root-mean-square forecast errors (RMSFE) of the IMS and IMSIC techniques. The sample size is $T = 50$, the forecast horizon $h = 1$. The DGP is an AR(1) with coefficient ρ and, at each period a location shift of magnitude γ occurs (independently from the past) with probability $p_T = 1/T$. The number of Monte Carlo replications is 5,000. The top/left panel records results when the model is an AR(1) and simulated data may not exhibit shifts in every period; the top right panel considers simulations when shifts occur in every sample; bottom panels report results from estimated AR(2) (left) and AR(4) (right) models respectively and shifts occur in every sample.

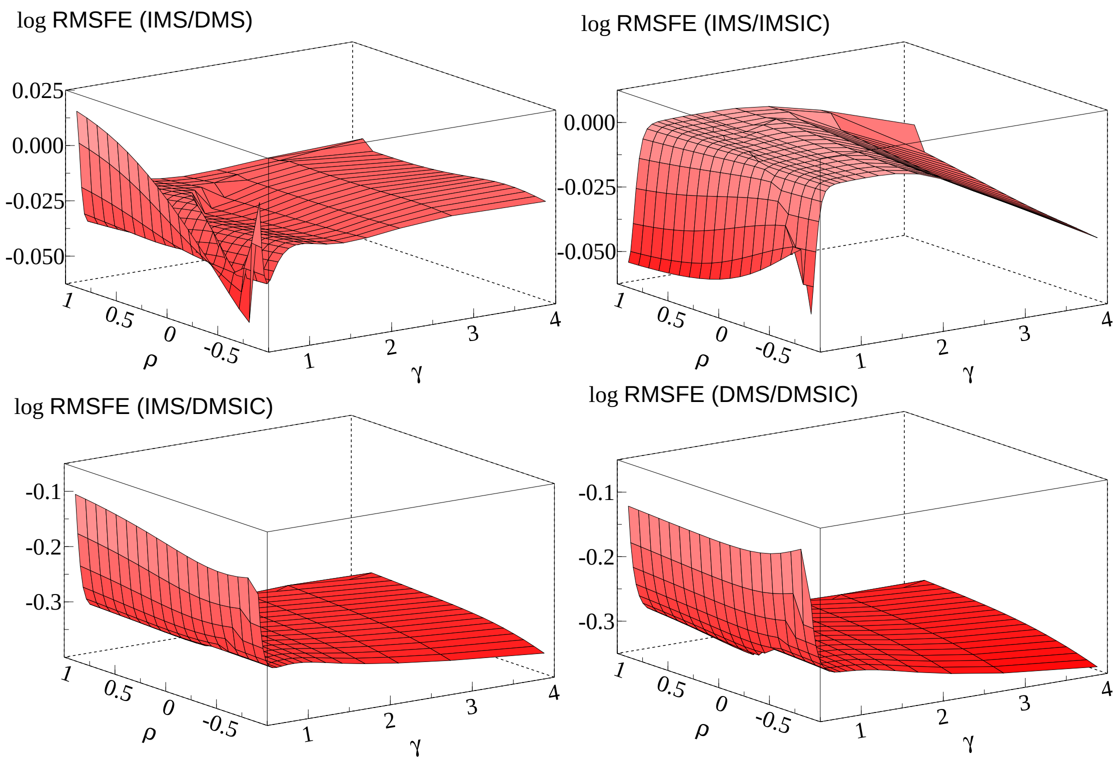

Figure 2: The figure presents the logarithms of the ratios of simulated root-mean-square forecast errors (RMSFE) of the IMS, DMS, IMSIC and DMSIC techniques. The sample size is $T = 50$, the forecast horizon $h = 5$. The DGP is an AR(1) with coefficient ρ and, at each period a location shift of magnitude γ occurs (independently from the past) with probability $p_T = 1/T$. The number of Monte Carlo replications is 5,000. The estimated model is an AR(1) and all simulated samples present at least one shift.

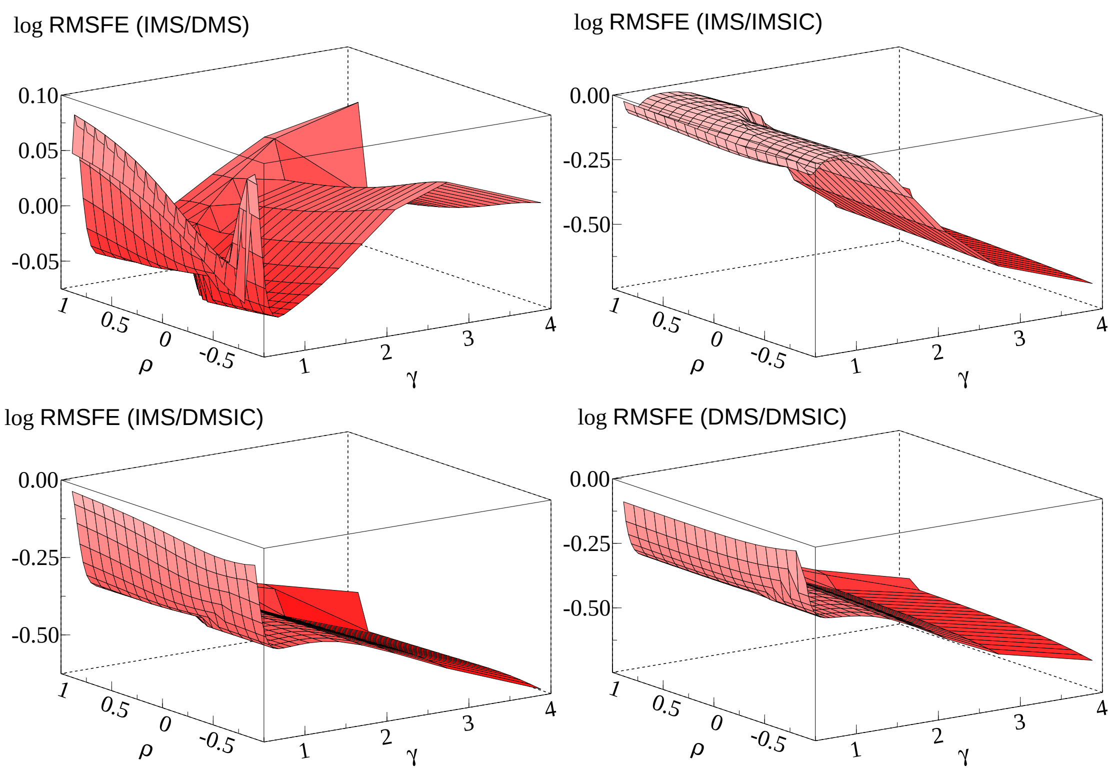

Figure 3: The figure presents the logarithms of the ratios of simulated root-mean-square forecast errors (RMSFE) of the IMS, DMS, IMSIC and DMSIC techniques. The sample size is $T = 50$, the forecast horizon $h = 5$. The DGP is an AR(1) with coefficient ρ and, at each period a location shift of magnitude γ occurs (independently from the past) with probability $p_T = 1/T$. The number of Monte Carlo replications is 5,000. The estimated model is an AR(2) and all simulated samples present at least one shift.

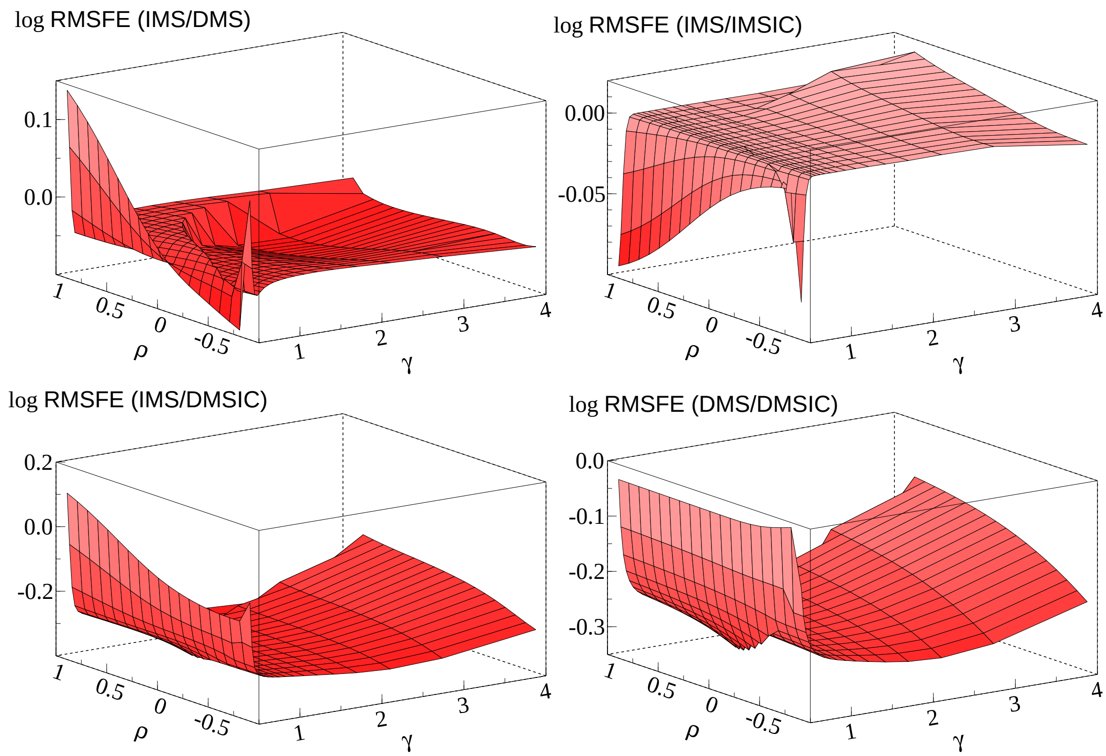

Figure 4: The figure presents the logarithms of the ratios of simulated root-mean-square forecast errors (RMSFE) of the IMS, DMS, IMSIC and DMSIC techniques. The sample size is $T = 50$, the forecast horizon $h = 10$. The DGP is an AR(1) with coefficient ρ and, at each period a location shift of magnitude γ occurs (independently from the past) with probability $p_T = 1/T$. The number of Monte Carlo replications is 5,000. The estimated model is an AR(1) and all simulated samples present at least one shift.

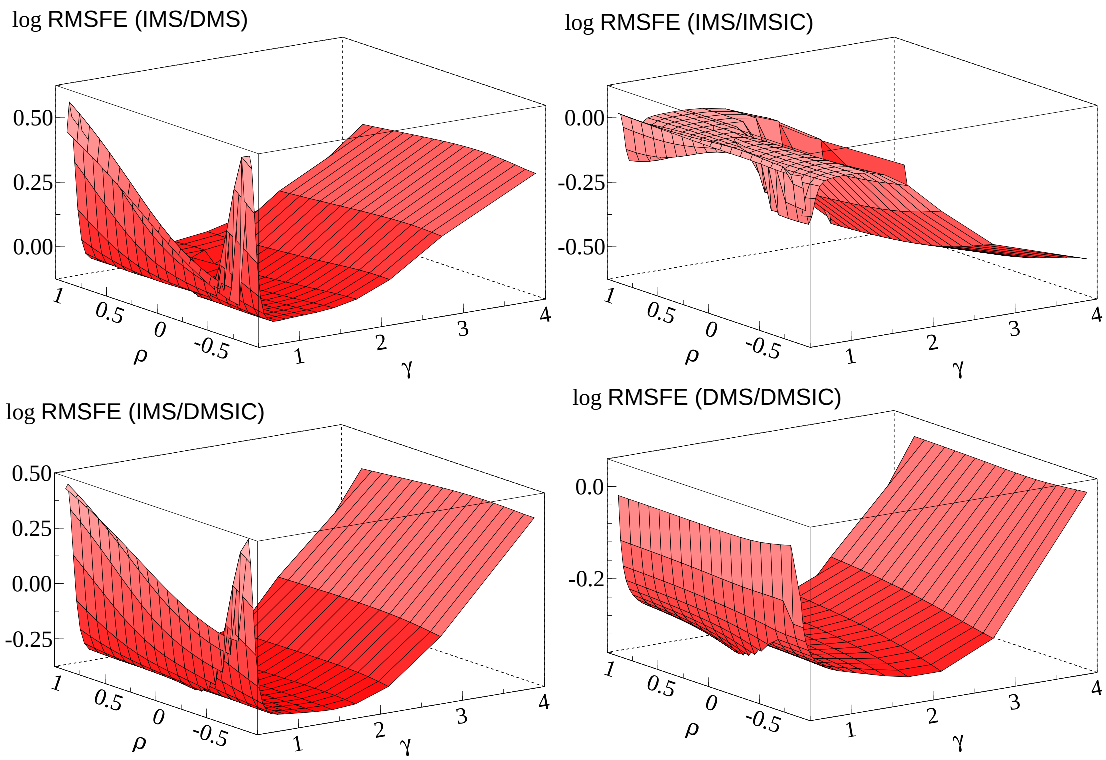

Figure 5: The figure presents the logarithms of the ratios of simulated root-mean-square forecast errors (RMSFE) of the IMS, DMS, IMSIC and DMSIC techniques. The sample size is $T = 50$, the forecast horizon $h = 10$. The DGP is an AR(1) with coefficient ρ and, at each period a location shift of magnitude γ occurs (independently from the past) with probability $p_T = 1/T$. The number of Monte Carlo replications is 5,000. The estimated model is an AR(2) and all simulated samples present at least one shift.

In addition to the simulations above, we considered fixing the break date $T-k$ to see its impact on relative forecast accuracy. We considered values of k ranging from 0 to 15 in order to study recent shifts. At horizon $h=1$, we find that imposing such recent breaks only has an impact on intercept correction – then a beneficial one – when the estimated model is an AR(4). For the other cases, the simulations were similar to those presented above. At longer horizons, the simulations show that a low k matters more. Figures 6 and 7 record the simulated log MSFE ratios of IMS over DMS. Each row corresponds to a different estimated AR(p) model, and each column to a different break date (no breaks occur over the forecast horizon). Figure 6 reports simulations at forecast horizon $h=5$ and Figure 7 at $h=10$. At horizon $h=5$, DMS using the AR(1) model is only beneficial when γ is large enough and $|\rho|$ not too close to 1. Here the shift must occur relatively close to the forecast origin for DMS to be useful, as in expression (10). By contrast, under the AR(2), DMS performs better when γ is larger (or when it is close to zero and ρ is close to unity). When the model is an AR(4), the benefits from DMS are lost. These remarks are in line with the analytical results of Section 3.4, where the increasing the lag order (but not too much) improves the relative performance of DMS. Similar results hold at longer horizons, such as $h=10$ in Figure 7, and the benefits from using a higher order AR(4) are then retained.

The conclusions from the simulations are that the relative performance of DMS improves with the horizon and the magnitude of recent shifts, and also when the model used for forecasting is an AR(p) of moderate order (relative to the forecast horizon): here the AR(2) is sufficient but the AR(4) often proves too large an order. The forms of intercept correction considered here do not bring substantial benefits over DMS.

5 Empirical Illustration

We assess the previous analysis by revisiting the forecasting exercise in PT. 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.[4] This consists

[4] For comparability, we did not extend the sample to include more recent data.

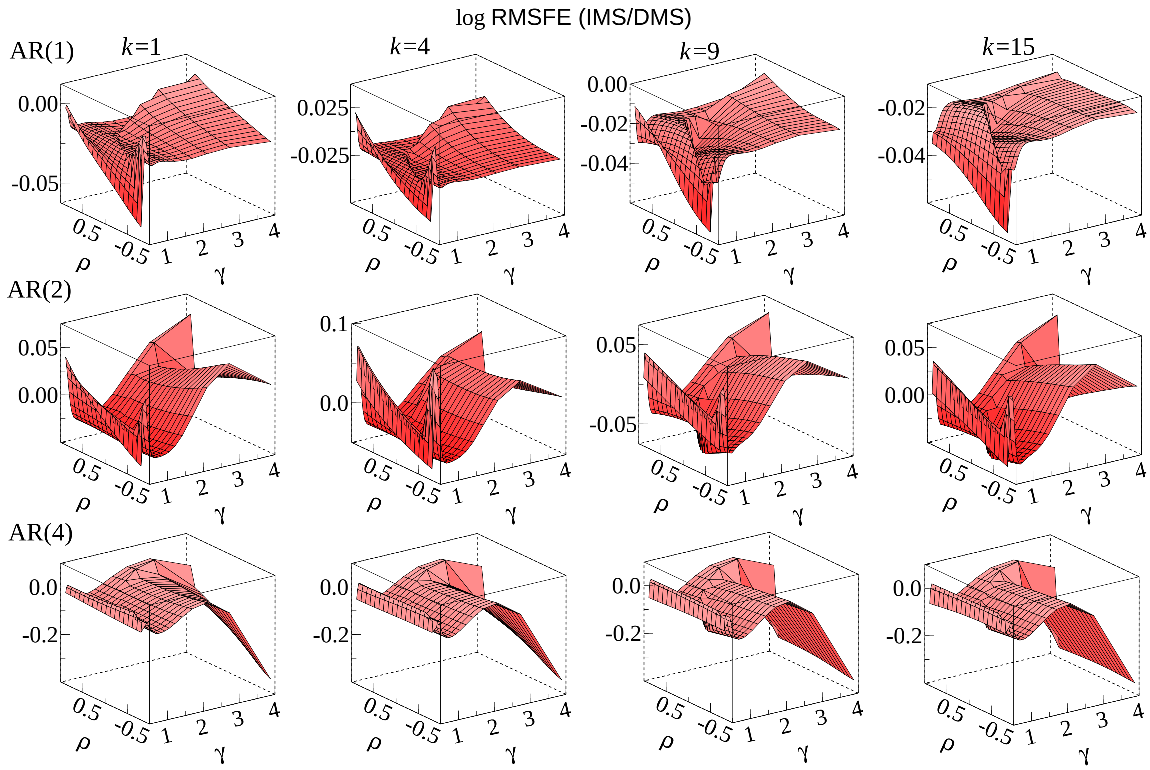

Figure 6: The figure presents the logarithms of the ratios of simulated root-mean-square forecast error (RMSFE) of the IMS and DMS techniques. The sample size is $T = 50$, the forecast horizon $h = 5$ The DGP is an AR(1) with coefficient ρ and, at date $T - k$, a location shift of magnitude γ. The number of Monte Carlo replications is 5,000. The estimated model is an AR(p) with $p = 1$ (first row), $p = 2$ (second row) or $p = 4$ (third row). Graphs in the first to fourth columns report measures when $k = 1$, 4, 9 and 15 respectively.

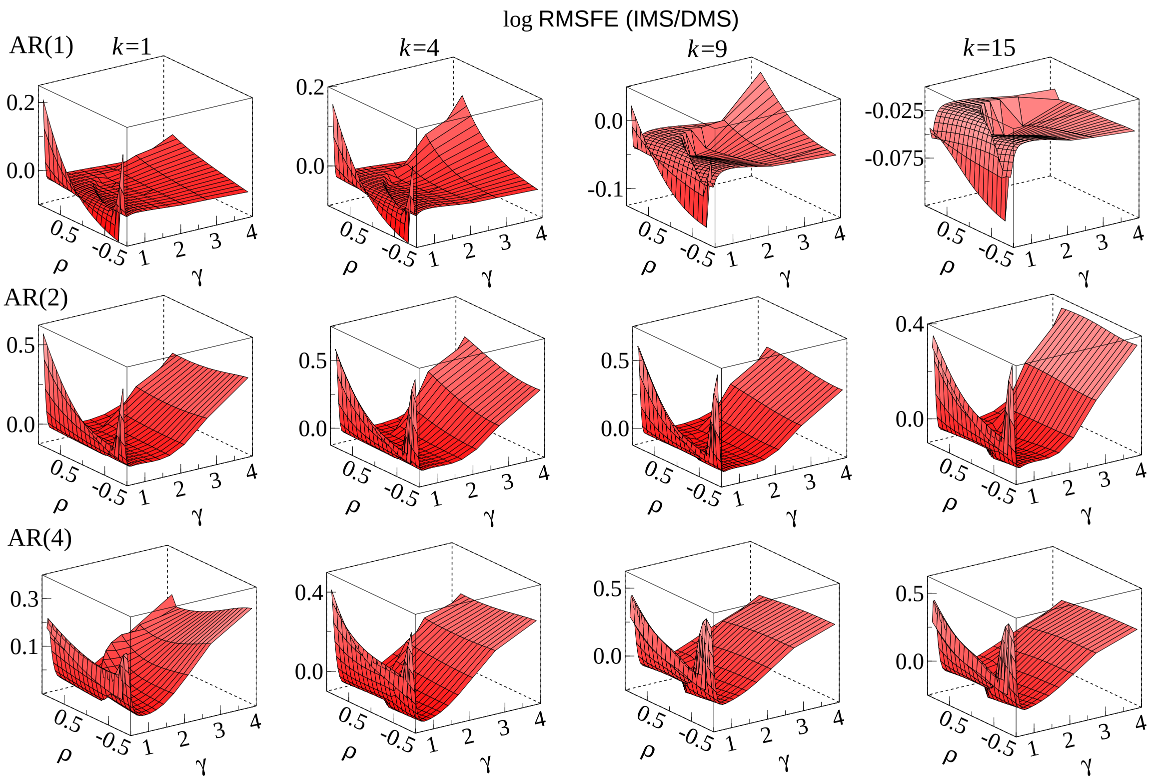

Figure 7: The figure presents the logarithms of the ratios of simulated root-mean-square forecast error (RMSFE) of the IMS and DMS techniques. The sample size is $T = 50$, the forecast horizon $h = 10$ The DGP is an AR(1) with coefficient ρ and, at date $T - k$, a location shift of magnitude γ. The number of Monte Carlo replications is 5,000. The estimated model is an AR(p) with $p = 1$ (first row), $p = 2$ (second row) or $p = 4$ (third row). Graphs in the first to fourth columns report measures when $k = 1$, 4, 9 and 15 respectively.

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). The dataset we use is relatively old and does not include the recent crisis. Our choice is deliberate as it allows for comparability of our results with earlier work.

To see the impact of a break on forecasting, we use *short* and *long* rolling windows of 25 and 40 observations, as well as an *expanding* window (starting with 40 observation) and carry estimation for AR(p) models for $p = 1, 2$, 4 as well as being selected by the Akaike Information Criterion (up to 4 lags as in Stock and Watson, 2004). We also estimate the date of a unique break using the method and code proposed by Bai and Perron (1998, 2003). The the test is that of a break in the conditional expectation, modeled as a shift in the intercept under the maintained assumption is that the model is an AR(1);[5] a trimming parameter of 15% excludes break dates at the beginning and end of the samples. Break significance is assessed at the 10% level and $\bar{b}$ denotes the percentage of significant breaks over the 102 subsamples;[6] the resulting post-break window size averaged over the subsamples with significant breaks is denoted by $\bar{k}$. In line with suggestions by *inter alia* Pesaran and Timmermann (2005) and Clark and McCracken (2009), we compute forecasts using the whole set of observations (in the subsample considered) or only those before (denoted with suffix pre) and after (post) the significant break date (i.e. using a two-step procedure). All methods are also assessed with multistep intercept correction (suffix IC). Since our interest lies in forecasting in the presence of breaks, and for comparability across the various methods, we only compute forecast errors when a break is significantly present. Overall, we find a substantial number of breaks in the rolling windows, much fewer in the expanding windows. In all situations, we report[7] both the ratios of out-of-sample (ex ante) empirical root MSFE (RMSFE), evaluated ex-post, as well as the median

[5] Breaks were only tested in the AR(1) since they refer to a change in the unconditional mean.

[6] A large $\bar{b}$ should advocate for a model with several in-sample breaks. We decide to leave this for future research and focus here on the timing of the most significant shift.

[7] We do not perform tests for equal or superior forecast accuracy as the properties of such tests are unknown when the forecasts are selected using in-sample pre-testing for breaks.

Extensive results are recorded in a supplement available from the author's website.

ratios of absolute forecast errors (mRAFE). Tables 1 to 3 present results ordered according either to the variable to forecast, the forecast horizon or per post-break sample size.

First, Table 1 records a summary of the performance comparisons. The table records, for the variables and horizons considered, the proportion of cases where IMS is the most accurate and the proportion of cases where alternatives perform at least as well as IMS. The latter enable to see directly whether there is a point in considering the techniques. The table does not report results for the estimated AR(2) forecasting model as these were comparable to the AR(1) . The table can be summarized as follows.

Regarding the horizon: the relative performance of IMS decreases with the forecast horizon for all models. It is at least as accurate as any other for more than 40% of the cases of rolling window of fixed size. It also very accurate when the model is an AR(p) whose order is chosen by the Akaike information criterion. Yet, overall (and also for fixed-order AR models) its performance is low at longer horizons. IMS performs very poorly when expanding windows of observations are used (Recursive Least Square, RLS); this corresponds to a proportion $\bar{b}$ of significant breaks that is low.[8] Now comparing other techniques, we see that DMS can perform at least as well as IMS at short horizons in over 50% of the cases considered. Its performance is especially strong in expanding windows — and also but to a lower extent when the model is an AR(p) with endogenously chosen order — and this carries to the overall proportion for all models considered. The other techniques do not perform as well (except for RLS). Yet, the variants of DMS (excluding itself) are on average better for fixed order AR and fixed window estimation.

Comparing the models across variables yields essentially similar analyses. A noticeable feature lies in the very different performances of IMS and DMS for industrial production and real GDP growth *versus* inflation and the interest rate. In Table 2, we see that the former two variables experience much fewer significant breaks than the latter two. This corroborates the analytical results derived in the previous sections.

[8] We refer the reader to the unreported supplementary web appendices for details.

5.1 Ordering per variable

Table 2 reports the outcome of the empirical analysis under the assumption of an AR(1) estimated over rolling windows of 25 observations. The results are contrasted: IMS fares relatively well as most ratios are below unity. Notable exceptions comprise DMS at short horizon which is more accurate 13 times out of 22 for $h = 2$, and 10 out of 22 for $h = 4$. Interest rates are interesting as they constitute the only type of data for which the IMS RMSFE is clearly increasing in the horizon. This is to the exception of Germany and the UK, for which the table shows that DMS retains an advantage at longer horizons. The German interest rate is especially worth looking at, for it exhibits much earlier breaks on average: unsurprisingly IMSpost is then most accurate overall. Using pre-break subsamples may also improve accuracy at intermediate horizons in the case of industrial production growth. Overall, intercept correction brings no benefit here. When mRAFE is used as a metric for accuracy comparisons, unreported tables also find DMS to be more accurate at $h = 2$, but less so at longer horizons. The main difference with the RMSFE is that according to the mRAFE, IMSIC proves better at forecasting inflation at horizons $h > 1$. The benefit of intercept correction is mainly in terms of absolute forecast error. This is exactly the purpose of intercept correction to reduce bias, at the cost of an increased variance. For intercept correction to achieve significant results on the RMSFE, the series must be experiencing frequent breaks.

Unreported tables show that when estimating AR(2) models, the mentioned results are still valid. Key differences include the lower accuracy of DMSpostIC at forecasting inflation and the interest rates (also DMSpreIC but to a lesser extent). These series differ from industrial production and real GDP growth which, when using expanding windows of observations, yield hardly any significant break; inflation and the interest rates exhibit by contrast a high occurrence of breaks. Increasing the autoregressive lag order, and hence reducing the autocorrelation of the residuals also has an impact on the relative performances of IMS and DMS, the former dominates as expected.

At long horizons, estimation uncertainty becomes a significant issue for DMS in small samples. As a consequence, IMS tends to present the lowest RMSFE when an AR(1) or an AR(p) , with p chosen by the AIC, are used.

Estimation over expanding windows enhances the benefits of DMS over IMS. As before, IMS performs best when the estimated model is an AR(4) . In particular, out of the 12 series that experience significant breaks, DMS forecasts better than IMS 5 times at $h = 2$, 11 times at $h = 5$, 7 times at $h = 10$ and 4 times at $h = 20$. The corresponding numbers of times DMS forecasts better using an AR(1) are 7, 10, 9 and 6 as the horizon grows from 2 to 20.

5.2 Ordering per post-break window

To view these results differently, we report the RMSFE for all series at once, but where the forecasts are categorized according the size of the post-break window of observations. This allows to examine the relationship between k and h. For conciseness, we only report one table: AR(1) and AR(4) models estimated over long windows (Table 3).

This table confirms the previous observations. In the AR(1) estimated over short windows DMS performs well for $h \leq 6$. This is also the case for IMSpost and DMSpost. Intercept correction only proves beneficial when estimation is carried over the pre-break subsample with k low and the horizon $h \leq k$. This remark also applies to other sample sizes and autoregressive lag orders to the exception of the AR(4) for $k \leq 7$ in which case IMS is outperformed when $h \geq k$. In this situation intercept correction also proves useful. When the model is an AR(4) , although the relative performance of IMS is decreasing in k (as long as $k \geq h$), this technique still dominates the others when the estimation window is long (for 25 observations, it can be dominated by, inter alia, DMS and DMSpostIC). Unreported tables exhibit similar results when considering expanding windows. The only difference lies in that intercept correction improves the forecasts when $h \geq k$, in particular for IMSpreIC and DMSpostIC.

5.3 Comparison with previous literature

The empirical analysis above confirms our theoretical results that long horizons, moderately recent breaks and small samples benefit DMS (and variations thereof) over IMS. As in Tiao and Xu (1993) and Chevillon and Hendry (2005), residual autocorrelation also favors DMS, although this technique can perform well at long horizons in the absence of residual autocorrelation: this could explain why Pesaran et al. (2011) recommend the use of the AIC over the BIC for selecting the lag order when forecasting at long horizons (since BIC selects fewer lags in general). Yet, the choice of lag order is not so important as the choice of forecasting technique and estimation window; this is in line with PT.

As analytically shown, whether $h < k$ also affects relative forecasting accuracy. When $h \geq k$, it appears that DMS should be preferred in a moderate order autoregressive model (AR(4) here). At shorter horizons, multistep intercept correction can reduce the forecast bias. That this improvement should come at the cost of an increase in the variance (indeed IMS performs better in terms of RMSFE than in mRAFE) does not necessarily imply that the overall MSFE is larger. Hence the role of the forecast bias must not be overlooked in the assessment of the forecast accuracy, this has been stressed by, inter alia, Hendry (2000), Clements and Hendry (2006), Pesaran and Timmermann (2004, 2005, 2009), and Giacomini and Rossi (2009).

We also show that the two-step procedures that rely solely on post-break data only perform well when breaks are very frequent (here for inflation and interest rates) and often in conjunction with intercept correction (in particular if k is low). Using only pre-breaks subsamples does not provide systematic benefits (except in IMSIC when k is very low). As noticed by PT, the benefits of increasing the number of observations are not strong in the presence of location shifts. Overall the results are hence in line with PT but extend them in the direction of multistep forecasting and with respect to the timing of the break. That pre-testing for breaks may not improve the accuracy of the forecasts has been discussed by many authors, see Elliott and Timmermann (2008).

The results also confirm and shed light on the empirical analysis of Marcellino et al. (2006), Proietti (2011) and Pesaran et al. (2011). The conclusions of these authors are that IMS is preferable when the horizon is long, the autoregressive order is large or when forecasting measures of real activity (this is also one of the conclusions of Clark and McCracken, 2008). This also arises here but we are able to qualify it further since we show that money and interest rates experience more frequent breaks. It could be argued that in fact these variables exhibit large moving average roots (see Schwert, 1987 and Stock and Watson, 2007) and that these yield spuriously significant breaks. Yet, the empirical relationship we observe between the horizon and the post-break sample size is in line with our theoretical predictions and so confirms that breaks indeed are at play. So the relative values of h and k are an important part of the forecasting performances and should not be neglected.

6 Conclusions

This paper shows that, when economic series undergo location shifts, there exists a rationale in using direct multistep estimation for forecasting. We have demonstrated that breaks matter to the relative performances of multistep 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 analyzed two techniques of intercept correction. We showed that direct multistep performs relatively better (*i*) at longer horizons, (*ii*) under recent breaks and (*iii*) when using an autoregressive model of moderate lag order. By contrast, intercept correction does not seem to provide systematic benefits in this context as it relies on correctly specifying the underlying dynamics.

The analysis that we have performed here mostly focused on autoregressive forecasting in the presence of a unique break. We could also compare our results with multivariate models where DMS has been shown by Chevillon (2009), Pesaran et al. (2011) and McElroy and McCracken (2012) to improve accuracy over IMS. Also, evaluating the performance using 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 as in Patton and Timmermann (2007).

References

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*, 47–78.

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

Balke, N. S. and T. B. Fomby (1991). Shifting trends, segmented trends, and infrequent permanent shocks. *Journal of Monetary Economics 28*(1), 61 – 85.

Bhansali, R. J. (1996). Asymptotically efficient autoregressive model selection for multi-step 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 P. Kokoszka (2002). Computation of the forecast coefficients for multistep prediction of long-range dependent time series. *International Journal of Forecasting 18*, 181–206.

Bobkoski, M. (1983). Hypothesis testing in nonstationary time series. Unpublished PhD thesis, Dept. of Statistics, University of Wisconsin, Madison.

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

Chan, N. H. and C. Z. Wei (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics 15*(3), 1050–1063.

Chevillon, G. (2007). Direct multi-step estimation and forecasting. *Journal of Economic Surveys 21*(4), 746–785.

Chevillon, G. (2009). Multi-step forecasting in emerging economies: An investigation of the South African GDP. *International Journal of Forecasting 25*, 602–28.

Chevillon, G. (2014). Robust cointegration testing in the presence of weak trends, with an application to the human origin of global warming. *Econometric Reviews* (forthcoming).

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

Clark, T. E. and M. W. McCracken (2008). Forecasting with small macroeconomic VARs in the presence of instabilities. *Frontiers of Economics and Globalization 3*, 93–147.

Clark, T. E. and M. W. McCracken (2009). Averaging forecasts from VARs with uncertain instabilities. *Journal of Applied Econometrics 50*, 363–95.

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 (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.

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 (2006). Forecasting with breaks. In G. Elliott, C. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1, Chapter 12. Amsterdam: North-Holland.

Diebold, F. X. and A. Inoue (2001). Long memory and regime switching. *Journal of Econometrics 105*(1), 131–159.

Eklund, J. and S. Karlsson (2005). Forecast combination and model averaging using predictive measures. *Econometric Reviews 26*, 329–363.

Elliott, G. (2005). Forecasting in the presence of a break. mimeo, UCSD.

Elliott, G. and A. Timmermann (2008). Economic forecasting. *Journal of Economic Literature 46*, 3–56.

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

Georgiev, I. (2002). Functional weak limit theory for rare outlying events. Working paper ECO No. 2002/22, European University Institute.

Giacomini, R. and B. Rossi (2009). Detecting and predicting forecast breakdowns. *Review of Economic Studies 76*, 669–705.

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. (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.

Ing, C.-K., J.-L. Lin, and S.-H. Yu (2009). Toward optimal multistep forecasts in non-stationary autoregressions. *Bernoulli 15*, 402–37.

Johansen, S. and B. Nielsen (2009). An analysis of the indicator saturation estimator as a robust regression estimator. In J. L. Castle and N. Shephard (Eds.), *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry.*, pp. 1–36. Oxford: Oxford University Press.

Jordà, O. and M. Marcellino (2010). Path forecast evaluation. *Journal of Applied Econometrics 25*(4), 635–662.

Leipus, R. and M.-C. Viano (2003). Long memory and stochastic trend. *Statistics and Probability Letters 61*, pp. 177–90.

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

Marcet, A. and T. J. Sargent (1989). Convergence of least-squares learning in environments with hidden state variables and private information. *Journal of Political Economy 97*(6), pp. 1306–1322.

McElroy, T. and M. W. McCracken (2012). Multi-step ahead forecasting of vector time series. Working Papers 2012-060, Federal Reserve Bank of St. Louis.

McElroy, T. and M. Wildi (2013). Multi-step-ahead estimation of time series models. *International Journal of Forecasting 29*(3), 378 – 394.

Patton, A. J. and A. Timmermann (2007). Properties of optimal forecasts under asymmetric loss and nonlinearity. *Journal of Econometrics 140*(2), 884–918.

Paye, B. and A. Timmermann (2006). Instability of return prediction models. *Journal of Empirical Finance 13*, 274–315.

Peña, D. (1994). Discussion: Second-generation time-series model, A comment. See Tiao and Tsay (1994), pp. 133–140.

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

Pesaran, H. and A. Pick (2011). Forecast combination across estimation windows. *Journal of Business and Economic Statistics 29*, 307–18.

Pesaran, H., A. Pick, and A. Timmermann (2011). Variable selection, estimation and inference for multi-period forecasting problems. *Journal of Econometrics 164*, 173–87.

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

Pettenuzzo, D. and A. Timmermann (2011). Predictability of stock returns and asset allocation under structural breaks. *Journal of Econometrics 164*(1), 60 – 78.

Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika 74*(3), 535–547.

Proietti, T. (2011). Direct and iterated multistep {AR} methods for difference stationary processes. *International Journal of Forecasting 27*(2), 266 – 280.

Santos, C., D. F. Hendry, and S. Johansen (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics 23*(2), 317–335.

Schorfheide, F. (2005). Var forecasting under misspecification. *Journal of Econometrics 128*(1), 99–136.

Schumacher, C. and J. Breitung (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.

Schwert, G. (1987). Effects of model specification on tests for unit roots in macroeconomic data. *Journal of Monetary Economics 20*, 73–103.

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

Stock, J. H. and M. W. Watson (2007). Why has US inflation become harder to forecast? *Journal of Money, Credit and Banking 39*(s1), 3–33.

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.

7 Appendix

Throughout the appendix, we denote references to corollary 5$(\cdot)$ of Georgiev (2002) as G-5$(\cdot)$.

7.1 Proof of Proposition 1

7.1.1 IMS

The OLS estimator of (τ, ρ) in the constant intercept model defined as

$$\begin{bmatrix} \widehat{\tau} \\ \widehat{\rho} \end{bmatrix} = \begin{bmatrix} T & \sum_{t=1}^{T} y_{t-1} \\ \sum_{t=1}^{T} y_{t-1} & \sum_{t=1}^{T} y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} y_t \\ \sum_{t=1}^{T} y_{t-1} y_t \end{bmatrix},$$

satisfy

$$\begin{bmatrix} \widehat{\tau} \\ \widehat{\rho} \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & \frac{1}{1-\rho} \int_0^1 J(r)\, dr \\ \frac{1}{1-\rho} \int_0^1 J(r)\, dr & \frac{\sigma_\epsilon^2}{1-\rho^2} + \frac{1}{(1-\rho)^2} \int_0^1 J^2(r)\, dr \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{1-\rho} \int_0^1 J(r)\, dr \\ \frac{\rho\sigma_\epsilon^2}{1-\rho^2} + \frac{1}{(1-\rho)^2} \int_0^1 J^2(r)\, dr \end{bmatrix}.$$

G-5(c) and G-5(f) imply that

$$\begin{aligned} T^{-1/2} \sum_{t=1}^{T} y_{t-1} \epsilon_t &\Rightarrow \frac{1}{1-\rho} \int_0^1 J(r)\, dW(r), \\ T^{-1} \sum_{t=1}^{T} y_t \tau_t &\Rightarrow \frac{1}{1-\rho} \int_0^1 J^2(r)\, dr, \end{aligned}$$

and the latter in particular implies also that $T^{-1} \sum_{t=1}^{T} y_{t-1} \tau_t \Rightarrow (1-\rho)^{-1} \int_0^1 J^2(r)\, dr$. We use these results together with

$$\begin{aligned} &\sqrt{T} \left(\begin{bmatrix} \widehat{\tau} \\ \widehat{\rho} - \rho \end{bmatrix} - \begin{bmatrix} 1 & T^{-1} \sum_{t=1}^{T} y_{t-1} \\ T^{-1} \sum_{t=1}^{T} y_{t-1} & T^{-1} \sum_{t=1}^{T} y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-1} \sum_{t=1}^{T} \tau_t \\ T^{-1} \sum_{t=1}^{T} y_{t-1} \tau_t \end{bmatrix} \right) \\ &= \begin{bmatrix} 1 & T^{-1} \sum_{t=1}^{T} y_{t-1} \\ T^{-1} \sum_{t=1}^{T} y_{t-1} & T^{-1} \sum_{t=1}^{T} y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-1/2} \sum_{t=1}^{T} \epsilon_t \\ T^{-1/2} \sum_{t=1}^{T} y_{t-1} \epsilon_t \end{bmatrix} \end{aligned}$$

to obtain the result.

7.1.2 DMS

Denoting $T_h = T - h + 1$ the effective sample, the DMS estimators are defined as

$$\begin{bmatrix} \widetilde{\tau}_h \\ \widetilde{\rho}_h \end{bmatrix} = \begin{bmatrix} T_h & \sum_{t=h}^{T} y_{t-h} \\ \sum_{t=h}^{T} y_{t-h} & \sum_{t=h}^{T} y_{t-h}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=h}^{T} y_t \\ \sum_{t=h}^{T} y_{t-h} y_t \end{bmatrix},$$

where $y_t = \rho^h y_{t-h} + \sum_{i=0}^{h-1} \rho^i (\epsilon_{t-i} + \tau_{t-i})$, so

$$\begin{bmatrix} \widetilde{\tau}_h \\ \widetilde{\rho}_h - \rho^h \end{bmatrix} = \begin{bmatrix} T_h & \sum_{t=h}^{T} y_{t-h} \\ \sum_{t=h}^{T} y_{t-h} & \sum_{t=h}^{T} y_{t-h}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=h}^{T} \sum_{i=0}^{h-1} \rho^i (\epsilon_{t-i} + \tau_{t-i}) \\ \sum_{t=h}^{T} y_{t-h} \left(\sum_{i=0}^{h-1} \rho^i (\epsilon_{t-i} + \tau_{t-i}) \right) \end{bmatrix}.$$

Assuming $h = \lfloor c_h T \rfloor$ implies, under the same conditions as in the one-step OLS estimators:

$$\begin{bmatrix} T^{-1} \sum_{t=h}^{T} y_{t-h} \\ T^{-1} \sum_{t=h}^{T} y_{t-h}^2 \end{bmatrix} \Rightarrow \begin{bmatrix} \frac{1}{1-\rho} \int_0^{1-c_h} J(r)\, dr \\ \frac{\sigma_\epsilon^2 (1-c_h)}{1-\rho^2} + \frac{1}{(1-\rho)^2} \int_0^{1-c_h} J^2(r)\, dr \end{bmatrix}.$$

We also need the limit of $T^{-1} \sum_{t=h}^{T} y_{t-h} y_t$. We express the sum as

$$\begin{aligned} T^{-1} \sum_{t=h}^{T} y_{t-h} y_t &= \rho^h T^{-1} \sum_{t=0}^{T-h} y_t^2 + T^{-1} \sum_{t=h}^{T} y_{t-h} \sum_{i=0}^{h-1} \rho^i (\epsilon_{t-i} + \tau_{t-i}) \\ &= \rho^h T^{-1} \sum_{t=0}^{T-h} y_t^2 + T^{-1} \sum_{t=h}^{T} y_{t-h} \sum_{i=0}^{h-1} \rho^i \epsilon_{t-i} + T^{-1} \sum_{t=0}^{T-h} y_t \sum_{i=0}^{h-1} \rho^i \tau_{t+h-i}, \end{aligned}$$

where we notice that $h = \lfloor c_h T \rfloor \to \infty$ implies that the term multiplied by ρ^h is asymptotically negligible. G-5(g) shows that

$$T^{-1} \sum_{t=h}^{T} y_{t-h} \sum_{i=0}^{h-1} \rho^i \epsilon_{t-i} = O_p\left(T^{-1/2}\right),$$

so we restrict our attention to

$$T^{-1} \sum_{t=0}^{T-h} y_t \sum_{i=0}^{h-1} \rho^i \tau_{t+h-i} = \frac{1-\rho^h}{1-\rho} T^{-1} \sum_{t=0}^{T-h} y_t \tau_t + T^{-1} \sum_{t=0}^{T-h} y_t \sum_{j=1}^{h} \frac{1-\rho^{h-j}}{1-\rho} q_{t+j} v_{t+j}.$$

Now G-5(b) implies that $\sum_{t=0}^{T-h} y_t \sum_{j=0}^{h-1} \rho^j q_{t+h-j} v_{t+h-j} = O_p(1)$. The previous expression rewrites hence as

$$\begin{aligned} T^{-1} \sum_{t=0}^{T-h} y_t \sum_{i=0}^{h-1} \rho^i \tau_{t+h-i} &= \frac{1}{1-\rho} T^{-1} \sum_{t=0}^{T-h} y_t \tau_t + T^{-1} \sum_{t=0}^{T-h} y_t \sum_{j=1}^{h} \frac{1}{1-\rho} q_{t+j} v_{t+j} + O_p\left(T^{-1}\right) \\ &= \frac{1}{1-\rho} T^{-1} \sum_{t=0}^{T-h} y_t \tau_t + \frac{1}{1-\rho} T^{-1} \sum_{t=0}^{T-h} y_t (\tau_{t+h} - \tau_t) + O_p\left(T^{-1}\right), \end{aligned}$$

i.e.

$$T^{-1}\sum_{t=0}^{T-h}y_t\sum_{i=0}^{h-1}\rho^i\tau_{t+h-i}=\frac{1}{1-\rho}T^{-1}\sum_{t=0}^{T-h}y_t\tau_{t+h}+O_p\left(T^{-1}\right),$$

where $T^{-1}\sum_{t=0}^{T-h}y_t\tau_{t+h}\Rightarrow\frac{1}{1-\rho}\int_0^{1-c}J(r)\,J(r+c)\,dr$. The result follows:

$$T^{-1}\sum_{t=h}^{T}y_{t-h}y_t\Rightarrow\frac{1}{(1-\rho)^2}\int_0^{1-c}J(r)\,J(r+c)\,dr.$$

G-5 similarly implies that

$$\left[\begin{array}{c}T^{-1}\sum_{t=h}^{T}\sum_{i=0}^{h-1}\rho^i\left(\epsilon_{t-i}+\tau_{t-i}\right)\\ T^{-1}\sum_{t=h}^{T}y_{t-h}\left(\sum_{i=0}^{h-1}\rho^i\left(\epsilon_{t-i}+\tau_{t-i}\right)\right)\end{array}\right]\Rightarrow\left[\begin{array}{c}\frac{1}{1-\rho}\int_0^{1-c}J(r+c)\,dr\\ \frac{1}{(1-\rho)^2}\int_0^{1-c}J(r)\,J(r+c)\,dr\end{array}\right].$$

We turn in the next subsection to the distributions conditional on event $\mathcal{E}_{c_k}$.

7.2 Conditional distributions in Corollaries 2 and 3

We assume a single break of magnitude γ occurs at time $\lfloor T(1-c_k)\rfloor$. This implies that $J_+(r)=\gamma\times 1_{\{r<1-c_k\}}$. We consider the corollaries together.

7.2.1 One-step estimators

The following holds under the stated assumptions:

$$\int_0^1 J_+(r)\,dr=c_k\gamma,\ \text{and}\ \int_0^1 J_+^2(r)\,dr=c_k\gamma^2,$$

so

$$\left[\begin{array}{c}\widehat{\tau}_\infty\\ \widehat{\rho}_\infty\end{array}\right]=\left[\begin{array}{cc}1&\frac{c_k\gamma}{1-\rho}\\ \frac{c_k\gamma}{1-\rho}&\frac{\sigma_\epsilon^2}{1-\rho^2}+\frac{c_k\gamma^2}{(1-\rho)^2}\end{array}\right]^{-1}\left[\begin{array}{c}\frac{c_k\gamma}{1-\rho}\\ \frac{\rho\sigma_\epsilon^2}{1-\rho^2}+\frac{c_k\gamma^2}{(1-\rho)^2}\end{array}\right],$$

which also rewrites as

$$\left[\begin{array}{c}\widehat{\tau}_\infty\\ \widehat{\rho}_\infty\end{array}\right]=\left[\begin{array}{c}\left(1-\frac{\frac{1+\rho}{1-\rho}c_k(1-c_k)\gamma^2}{\sigma_\epsilon^2+\frac{1+\rho}{1-\rho}c_k(1-c_k)\gamma^2}\right)\gamma c_k\\ 1-\frac{\sigma_\epsilon^2}{\sigma_\epsilon^2+\frac{1+\rho}{1-\rho}c_k(1-c_k)\gamma^2}(1-\rho)\end{array}\right].$$

This shows that as $c_k\to 0$, the asymptotic distributions satisfy:

$$\left[\begin{array}{c}\widehat{\tau}_\infty\\ \widehat{\rho}_\infty\end{array}\right]\underset{c_k\to 0}{\sim}\left[\begin{array}{c}\gamma c_k\\ \rho\end{array}\right].$$

7.2.2 Multistep estimators

We first compute the integrals, denoting $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. First if $1 - c_h \leq 1 - c_k$ (i.e. $c_k \leq c_h$)

$$\int_0^{1-c_h} J_+(r)\, dr = \int_0^{1-c_h} J_+^2(r)\, dr = 0; \quad \int_0^{1-c_h} J(r)\, J(r + c_h)\, dr = 0;$$
$$\int_{c_h}^{1} J_+(r)\, dr = \int_{c_h \vee (1-c_k)}^{1} J_+(r)\, dr = ((1 - c_h) \wedge c_k)\, \gamma;$$

then if $c_k \geq c_h$

$$\int_0^{1-c_h} J_+(r)\, dr = (c_k - c_h)\, \gamma; \quad \int_0^{1-c_h} J_+^2(r)\, dr = (c_k - c_h)\, \gamma^2;$$
$$\int_0^{1-c_h} J(r)\, J(r + c_h)\, dr = ((1 - 2c_h) \wedge (c_k - c_h))\, \gamma^2;$$
$$\int_{c_h}^{1} J_+(r)\, dr = ((1 - c_h) \wedge c_k)\, \gamma.$$

Hence

$$\int_0^{1-c_h} J_+(r)\, dr = [(c_k - c_h) \vee 0]\, \gamma; \quad \int_0^{1-c_h} J_+^2(r)\, dr = [(c_k - c_h) \vee 0]\, \gamma^2;$$
$$\int_0^{1-c_h} J(r)\, J(r + c_h)\, dr = [((1 - 2c_h) \wedge (c_k - c_h)) \vee 0]\, \gamma^2;$$
$$\int_{c_h}^{1} J_+(r)\, dr = ((1 - c_h) \wedge c_k)\, \gamma.$$

The previous expressions imply that the estimators hence, satisfy the following expressions. First, if $c_k \leq c_h$

$$\begin{bmatrix} \widetilde{\tau}_{\infty, c_h} \\ \widetilde{\rho}_{\infty, c_h} \end{bmatrix} = \begin{bmatrix} (1 - c_h)^{-1} & 0 \\ 0 & \left(\frac{\sigma_\epsilon^2 (1 - c_h)}{1 - \rho^2}\right)^{-1} \end{bmatrix} = \begin{bmatrix} \left(1 \wedge \frac{c_k}{1 - c_h}\right) \frac{\gamma}{1 - \rho} \\ 0 \end{bmatrix}$$
$$\underset{c_k \to 0}{\sim} \begin{bmatrix} \frac{\gamma}{1-\rho} \frac{c_k}{1 - c_h} \\ 0 \end{bmatrix}.$$

Now if $c_k > c_h$,

$$\begin{bmatrix} \widetilde{\tau}_{\infty,c_h} \\ \widetilde{\rho}_{\infty,c_h} \end{bmatrix} = \begin{bmatrix} 1-c_h & (c_k-c_h)\frac{\gamma}{1-\rho} \\ (c_k-c_h)\frac{\gamma}{1-\rho} & \frac{\sigma_\epsilon^2(1-c_h)}{1-\rho^2} + (c_k-c_h)\frac{\gamma^2}{(1-\rho)^2} \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} ((1-c_h)\wedge c_k)\frac{\gamma}{1-\rho} \\ ((1-2c_h)\wedge(c_k-c_h))\left(\frac{\gamma}{1-\rho}\right)^2 \end{bmatrix}$$

$$= \frac{1}{(1-c_h)(1-c_h)\frac{\sigma_\epsilon^2}{1-\rho^2} + (1-c_k)(c_k-c_h)\left(\frac{\gamma}{1-\rho}\right)^2}$$

$$\times \begin{bmatrix} (1-c_h)((1-c_h)\wedge c_k)\frac{\gamma}{1-\rho}\frac{\sigma_\epsilon^2}{1-\rho^2} + c_h(c_k-c_h)\left(\frac{\gamma}{1-\rho}\right)^3 \\ ((1-c_k)[(1-c_h)\wedge c_k] - c_h(1-c_h))\left(\frac{\gamma}{1-\rho}\right)^2 \end{bmatrix},$$

so as $c_h \to 0$

$$\begin{bmatrix} \widetilde{\tau}_{\infty,c_h} \\ \widetilde{\rho}_{\infty,c_h} \end{bmatrix} \underset{c_h\to 0}{\to} \frac{1}{\frac{\sigma_\epsilon^2}{1-\rho^2} + (1-c_k)c_k\left(\frac{\gamma}{1-\rho}\right)^2} \begin{bmatrix} \frac{\sigma_\epsilon^2}{1-\rho^2} \\ (1-c_k)\frac{\gamma}{1-\rho} \end{bmatrix} \frac{\gamma}{1-\rho}c_k.$$

Also, as $\gamma \to \infty$,

$$\begin{bmatrix} \widetilde{\tau}_{\infty,c_h} \\ \widetilde{\rho}_{\infty,c_h} \end{bmatrix} \underset{\gamma\to\infty}{\sim} \begin{bmatrix} \frac{c_h}{1-c_k}\left(\frac{\gamma}{1-\rho}\right) \\ \left[\left(\frac{1-2c_h}{c_k-c_h}\right)\wedge 1\right] - \frac{c_h}{1-c_k} \end{bmatrix},$$

so if $c_k - c_h \leq 1 - 2c_h$

$$\begin{bmatrix} \widetilde{\tau}_{\infty,c_h} \\ \widetilde{\rho}_{\infty,c_h} \end{bmatrix} \underset{\gamma\to\infty}{\sim} \begin{bmatrix} \frac{c_h}{1-c_k}\left(\frac{\gamma}{1-\rho}\right) \\ 1 - \frac{c_h}{1-c_k} \end{bmatrix}.$$

7.3 Proof of Proposition 4

We start with IMS:

$$\mathsf{E}\left(\widehat{e}_{T+h|T}|\mathcal{E}_{c_k}\right) = \mathsf{E}\left(\rho_{\{h\}}\tau_T + \rho^h y_T - \left(\widehat{\rho}_{\{h\}}\widehat{\tau} + \widehat{\rho}^h y_T\right)|\mathcal{E}_{c_k}\right)$$
$$\to \frac{\gamma}{1-\rho} - \frac{\widehat{\tau}_\infty}{1-\widehat{\rho}_\infty},$$

using Slutsky's formula since, conditional on $\mathcal{E}_{c_k}$, the estimators converge towards a non-stochastic limit. The variance is

$$\mathsf{Var}\left(\widehat{e}_{T+h|T}|\mathcal{E}_{c_k}\right) = \mathsf{Var}\left[\left(\rho^h - \widehat{\rho}^h\right)(y_T - \mathsf{E}(y_T|\mathcal{E}_{c_k}))|\mathcal{E}_{c_k}\right] + \mathsf{Var}\left(\sum_{i=0}^{h-1}\rho^i\epsilon_{T+h-i}|\mathcal{E}_{c_k}\right)$$
$$\to \sigma_y^2, \quad \text{as } T \to \infty.$$

Similarly, the DMS forecast error admits the conditional expectation

$$\begin{aligned}\mathsf{E}\left(\widetilde{e}_{T+h|T}|\mathcal{E}_{c_k}\right) &= \rho_{\{h\}}\tau_T + \rho^h\mathsf{E}\left(y_T|\mathcal{E}_{c_k}\right) - \left(\mathsf{E}\left(\widetilde{\tau}_h + \widetilde{\rho}_h y_T|\mathcal{E}_{c_k}\right)\right) \\ &\to \frac{\gamma}{1-\rho}\left(1-\widetilde{\rho}_{\infty,c_h}\right) - \widetilde{\tau}_{\infty,c_h},\end{aligned}$$

and conditional variance

$$\begin{aligned}\mathsf{Var}\left(\widetilde{e}_{T+h|T}|\mathcal{E}_{c_k}\right) &= \mathsf{Var}\left[\left(\rho^h - \widetilde{\rho}_h\right)\left(y_T - \mathsf{E}\left(y_T|E_{c_k}\right)\right)|E_{c_k}\right] + \mathsf{Var}\left(\sum_{i=0}^{h-1}\rho^i\epsilon_{T+h-i}|E_{c_k}\right) \\ &\to \left(1+\widetilde{\rho}^2_{\infty,c_h}\right)\sigma_y^2.\end{aligned}$$

7.4 Proof of Proposition 5

We now derive the distribution of intercept corrected forecasts. First

$$\delta_{IC} = y_T - \widehat{y}_{T|T-1} = \tau_T - \widehat{\tau} + \left(\rho - \widehat{\rho}\right)y_{T-1} + \epsilon_T,$$

and so the IMSIC forecast error is

$$\begin{aligned}\widehat{e}^{IMSIC}_{T+h|T} &= \rho_{\{h-1\}}\tau_T - \widehat{\rho}_{\{h-1\}}\widehat{\tau} + \rho^h\tau_T - \widehat{\rho}^h\widehat{\tau} + \left(\rho^{h-1} - \widehat{\rho}^{h-1}\right)\tau_T \\ &+ \left(\rho^{h+1} - \widehat{\rho}^{h+1}\right)y_{T-1} - \widehat{\rho}^{h-1}\left(\rho - \widehat{\rho}\right)y_{T-1} \\ &+ \sum_{j=1}^{h}\rho^{h-i}\epsilon_{T+i} + \left(\rho^h - \widehat{\rho}^h\right)\epsilon_T - \widehat{\rho}^{h-1}\epsilon_T,\end{aligned}$$

whose asymptotic conditional moments coincide with those of IMS:

$$\begin{aligned}&\mathsf{E}\left(\widehat{e}^{IMSIC}_{T+h|T}|\mathcal{E}_{c_k}\right) \to \frac{\gamma}{1-\rho} - \frac{1}{1-\widehat{\rho}_\infty}\widehat{\tau}_\infty; \\ &\mathsf{V}\left(\widehat{e}^{IMSIC}_{T+h|T}|\mathcal{E}_{c_k}\right) \to \sigma_y^2.\end{aligned}$$

Now, for DMS, the correction is $\delta_{DMSIC} = y_T - \widehat{y}_{T|T-h}$ which is equal to

$$\delta_{DMSIC} = \sum_{j=0}^{h-1}\rho^j\tau_{T-j} - \widetilde{\tau}_h + \left(\rho^h - \widetilde{\rho}_h\right)y_{T-h} + \sum_{j=0}^{h-1}\rho^j\epsilon_{T-j}.$$

The corresponding forecast error admits the following moments: first, the expectation $\mathsf{E}\left(\widetilde{e}^{DMSIC}_{T+h|T}|\mathcal{E}_{c_k}\right)$ can be shown to converge to zero both is $c_k \leq c_h$ and if $c_h \leq c_k$. As for the second moments, we use

$$\begin{aligned}
\mathsf{Cov}\left(y_T, \sum_{j=0}^{h-1} \rho^j \epsilon_{T-j} | \mathcal{E}_{c_k}\right) &\to \sigma_y^2; \\
\mathsf{Cov}\left(\left(y_T - \mathsf{E}\left(y_T | \mathcal{E}_{c_k}\right)\right), \delta_{DMSIC}\right) &\to \sigma_y^2; \\
\mathsf{Var}\left[\delta_{DMSIC} | \mathcal{E}_{c_k}\right] &\to \left(1 + \widetilde{\rho}^2_{\infty,c_h}\right) \sigma_y^2;
\end{aligned}$$

which imply that the variance satisfies

$$\begin{aligned}
\mathsf{Var}\left(\widetilde{e}^{DMSIC}_{T+h|T} | \mathcal{E}_{c_k}\right) &= \mathsf{Var}\left[\left(\rho^h - \widetilde{\rho}_h\right)\left(y_T - \mathsf{E}\left(y_T | \mathcal{E}_{c_k}\right)\right) - \delta_{DMSIC} | \mathcal{E}_{c_k}\right] + \mathsf{Var}\left(\sum_{i=0}^{h-1} \rho^i \epsilon_{T+h-i} | \mathcal{E}_{c_k}\right) \\
&\to \left(1 + 2\widetilde{\rho}^2_{\infty,c_h}\right) \sigma_y^2.
\end{aligned}$$

7.5 Proof of Corollary 6

We now compute the asymptotic conditional MSFEs:

$$\begin{aligned}
\mathsf{E}\left(\widehat{e}_{T+h|T} | \mathcal{E}_{c_k}\right)^2 + \mathsf{Var}\left(\widehat{e}_{T+h|T} | \mathcal{E}_{c_k}\right) &\to \mathsf{MSFE}^{IMS}_{\infty} = \left[\frac{\gamma}{1-\rho} - \frac{\widehat{\tau}_{\infty}}{1-\widehat{\rho}_{\infty}}\right]^2 + \sigma_y^2; \\
\mathsf{E}\left(\widetilde{e}_{T+h|T} | \mathcal{E}_{c_k}\right)^2 + \mathsf{Var}\left(\widetilde{e}_{T+h|T} | \mathcal{E}_{c_k}\right) &\to \mathsf{MSFE}^{DMS}_{\infty} \\
&\qquad = \left[\frac{\gamma}{1-\rho}\left(1 - \widetilde{\rho}_{\infty,c_h}\right) - \widetilde{\tau}_{\infty,c_h}\right]^2 + \left(1 + \widetilde{\rho}^2_{\infty,c_h}\right) \sigma_y^2; \\
\mathsf{E}\left(\widehat{e}^{IMSIC}_{T+h|T} | \mathcal{E}_{c_k}\right)^2 + \mathsf{Var}\left(\widehat{e}^{IMSIC}_{T+h|T} | \mathcal{E}_{c_k}\right) &\to \mathsf{MSFE}^{IMSIC}_{\infty} = \mathsf{MSFE}^{IMS}_{\infty}; \\
\mathsf{E}\left(\widetilde{e}^{DMSIC}_{T+h|T} | \mathcal{E}_{c_k}\right)^2 + \mathsf{Var}\left(\widetilde{e}^{DMSIC}_{T+h|T} | \mathcal{E}_{c_k}\right) &\to \mathsf{MSFE}^{DMS}_{\infty} = \left(1 + 2\widetilde{\rho}^2_{\infty,c_h}\right) \sigma_y^2.
\end{aligned}$$

Hence the multistep MSFEs:

$$\mathsf{MSFE}^{IMS}_{\infty} = \left(\frac{\gamma}{1-\rho}\right)^2 (1 - c_k)^2 + \sigma_y^2,$$

and

$$\mathsf{MSFE}_{\infty}^{DMS} \underset{c_k \to 0}{\sim} \left(\frac{\gamma}{1-\rho}\right)^2 \left(1 - \frac{c_k}{1-c_h}\right)^2 + \sigma_y^2;$$

$$\mathsf{MSFE}_{\infty}^{DMS} \underset{c_h \to 0}{\sim} \left(\frac{\gamma}{1-\rho}\right)^2 \left(1 - \frac{(1-c_k)\, c_k}{\frac{\sigma_\epsilon^2}{1-\rho^2} + (1-c_k)\, c_k \left(\frac{\gamma}{1-\rho}\right)^2} \left(\frac{\gamma}{1-\rho}\right)^2\right)^2 (1-c_k)^2$$
$$+ \left(1 + \left(1 - \frac{\frac{\sigma_\epsilon^2}{1-\rho^2}}{\frac{\sigma_\epsilon^2}{1-\rho^2} + (1-c_k)\, c_k \left(\frac{\gamma}{1-\rho}\right)^2}\right)^2\right) \sigma_y^2.$$

Finally, for $\mathsf{MSFE}_{\infty}^{DMSIC} = \left(1 + 2\widetilde{\rho}^2_{\infty, c_h}\right) \sigma_y^2$,

$$\mathsf{MSFE}_{\infty}^{DMSIC} \underset{c_k \to 0}{\sim} \sigma_y^2;$$

$$\mathsf{MSFE}_{\infty}^{DMSIC} \underset{c_h \to 0}{\sim} \left(1 + 2\left(\frac{\gamma}{1-\rho}\right)^4 \frac{c_k^2}{\left[\frac{\sigma_\epsilon^2}{1-\rho^2} + (1-c_k)\, c_k \left(\frac{\gamma}{1-\rho}\right)^2\right]^2}\right) \sigma_y^2.$$

Also if $c_k \leq 1 - c_h$, as $\gamma \to \infty$, $\mathsf{MSFE}_{\infty}^{DMSIC} \underset{\gamma \to \infty}{\sim} \left(1 + 2\left(1 - \frac{c_h}{1-c_k}\right)^2\right) \sigma_y^2$ remains finite contrary to the others.

7.6 Proof of Expression (14)

We consider the estimators

$$\begin{bmatrix} \widehat{\tau} \\ \widehat{\rho}_1 - \rho_1 \\ \vdots \\ \widehat{\rho}_p - \rho_p \end{bmatrix} = \begin{bmatrix} T-p+1 & \sum_{t=p}^T y_{t-1} & \cdots & \sum_{t=p}^T y_{t-p} \\ \sum_{t=p}^T y_{t-1} & \sum_{t=p}^T y_{t-1}^2 & \cdots & \sum_{t=p}^T y_{t-1} y_{t-p} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=p}^T y_{t-p} & \sum_{t=p}^T y_{t-1} y_{t-p} & & \sum_{t=p}^T y_{t-p}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=p}^T \tau_t + \epsilon_t \\ \sum_{t=p}^T y_{t-1} (\tau_t + \epsilon_t) \\ \vdots \\ \sum_{t=p}^T y_{t-p} (\tau_t + \epsilon_t) \end{bmatrix}.$$

Conditional on event $\mathcal{E}_{c_k}$, and assuming $p < k$ we know that the submatrix

$$\begin{bmatrix} \sum_{t=p}^{T} y_{t-1}^2 & \cdots & \sum_{t=p}^{T} y_{t-1}y_{t-p} \\ \vdots & \ddots & \vdots \\ \sum_{t=p}^{T} y_{t-1}y_{t-p} & \cdots & \sum_{t=p}^{T} y_{t-p}^2 \end{bmatrix} = \frac{\gamma^2}{(1-\rho)^2} \begin{bmatrix} k & k-1 & \cdots & k-p+1 \\ k-1 & k-1 & \cdots & k-p+1 \\ \vdots & \vdots & \ddots & \vdots \\ k-p+1 & k-p+1 & \cdots & k-p+1 \end{bmatrix}$$

$$+ \frac{T\sigma_\epsilon^2}{(1-\rho)^2} \begin{bmatrix} 1 & \cdots & \rho^{p-1} \\ \vdots & \ddots & \vdots \\ \rho^{p-1} & \cdots & 1 \end{bmatrix} + o\left(T\right).$$

Consider first the case where $\sigma_\epsilon^2 << \gamma^2$ so the first term on the right hand side above dominates. We then use the formula

$$\begin{bmatrix} a_1 & a_2 & a_3 & \cdots & a_{n-1} & a_n \\ a_2 & a_2 & a_3 & \cdots & a_{n-1} & a_n \\ a_3 & a_3 & a_3 & \ddots & a_{n-1} & a_n \\ \vdots & \vdots & \ddots & \ddots & a_{n-1} & a_n \\ a_{n-1} & a_{n-1} & a_{n-1} & a_{n-1} & a_{n-1} & a_n \\ a_n & a_n & a_n & a_n & a_n & a_n \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \frac{1}{a_1-a_2} & \frac{1}{a_2-a_1} & 0 & \cdots & 0 & 0 \\ \frac{1}{a_2-a_1} & \frac{a_1-a_3}{(a_1-a_2)(a_2-a_3)} & \frac{1}{a_3-a_2} & 0 & \cdots & 0 \\ 0 & \frac{1}{a_3-a_2} & \frac{a_2-a_4}{(a_2-a_3)(a_3-a_4)} & \ddots & \ddots & 0 \\ 0 & 0 & \frac{1}{a_4-a_3} & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \frac{a_{n-2}-a_n}{(a_{n-2}-a_{n-1})(a_{n-1}-a_n)} & \frac{1}{a_n-a_{n-1}} \\ 0 & 0 & 0 & 0 & \frac{1}{a_n-a_{n-1}} & \frac{a_{n-1}}{a_n}\frac{1}{a_n-a_{n-1}} \end{bmatrix},$$

i.e.

$$\begin{bmatrix} k & k-1 & \cdots & k-p+1 \\ k-1 & k-1 & \cdots & k-p+1 \\ \vdots & \vdots & \ddots & \vdots \\ k-p+1 & k-p+1 & \cdots & k-p+1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & -1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 2 & -1 \\ 0 & \cdots & 0 & -1 & \frac{k-p+2}{k-p+1} \end{bmatrix}.$$

Hence

$$\begin{bmatrix} \widehat{\tau} \\ \widehat{\rho}_1 - \rho_1 \\ \vdots \\ \widehat{\rho}_p - \rho_p \end{bmatrix} \to \begin{bmatrix} T-p+1 & \frac{\gamma}{1-\rho}\left[k \quad \cdots \quad k-p+1\right] \\ \frac{\gamma}{1-\rho}\begin{bmatrix} k \\ \vdots \\ k-p+1 \end{bmatrix} & \frac{\gamma^2}{(1-\rho)^2}\begin{bmatrix} k & k-1 & \cdots & k-p+1 \\ k-1 & k-1 & \cdots & k-p+1 \\ \vdots & \vdots & \ddots & \vdots \\ k-p+1 & k-p-1 & \cdots & k-p+1 \end{bmatrix} \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} (k+1)\,\gamma \\ k\frac{\gamma^2}{1-\rho} \\ \vdots \\ (k-p+1)\,\frac{\gamma^2}{1-\rho} \end{bmatrix}.$$

Now we use the properties of inverses of partitioned matrices:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} [M/D]^{-1} & -[M/D]^{-1}BD^{-1} \\ -D^{-1}C\,[M/D]^{-1} & D^{-1} + D^{-1}C\,[M/D]^{-1}BD^{-1} \end{bmatrix},$$

with $M/D = A - BD^{-1}C$. Using this formula for the matrix of OLS second empirical moments above, first letting

$$M/D = (T-p+1) - \left(\frac{\gamma}{1-\rho}\right)^2 \left[k \ \cdots \ k-p+1\right] D^{-1} \begin{bmatrix} k \\ \vdots \\ k-p+1 \end{bmatrix}$$

$$= (T-p+1) - \begin{bmatrix} k \\ \vdots \\ k-p+1 \end{bmatrix}' \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = T-p-k+1,$$

where we used

$$D^{-1}\begin{bmatrix} k \\ \vdots \\ k-p+1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Then

$$\begin{bmatrix} T-p+1 & \frac{\gamma}{1-\rho}\begin{bmatrix}k & \cdots & k-p+1\end{bmatrix} \\ \frac{\gamma}{1-\rho}\begin{bmatrix} k \\ \vdots \\ k-p+1 \end{bmatrix} & \frac{\gamma^2}{(1-\rho)^2}\begin{bmatrix} k & k-1 & \cdots & k-p+1 \\ k-1 & k-1 & \cdots & k-p+1 \\ \vdots & \vdots & \ddots & \vdots \\ k-p+1 & k-p-1 & \cdots & k-p+1 \end{bmatrix} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \frac{1}{T-p-k+1} & -\frac{(1-\rho)}{\gamma(T-p-k+1)}\begin{bmatrix}1 & 0 & \cdots & 0\end{bmatrix} \\ -\frac{(1-\rho)}{\gamma(T-p-k+1)}\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} & \frac{(1-\rho)^2}{\gamma^2}\begin{bmatrix} 1+\frac{1}{T-p-k+1} & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & -1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 2 & -1 \\ 0 & \cdots & 0 & -1 & \frac{k-p+2}{k-p+1} \end{bmatrix} \end{bmatrix}.$$

The limit of the OLS estimators is therefore

$$\begin{bmatrix} \widehat{\tau} \\ \widehat{\rho}_1 - \rho_1 \\ \vdots \\ \widehat{\rho}_p - \rho_p \end{bmatrix}$$

$$\xrightarrow{p} \begin{bmatrix} \frac{(k+1)\gamma}{T-p-k+1} - \frac{(1-\rho)}{\gamma(T-p-k+1)} k \frac{\gamma^2}{1-\rho} \\ -\frac{(1-\rho)}{\gamma(T-p-k+1)}(k+1)\gamma + \frac{(1-\rho)^2}{\gamma^2}\left(1+\frac{1}{T-p-k+1}\right) k \frac{\gamma^2}{1-\rho} - (k-1)\frac{(1-\rho)^2}{\gamma^2}\frac{\gamma^2}{1-\rho} \\ \frac{(1-\rho)^2}{\gamma^2}\left[-k\frac{\gamma^2}{1-\rho} + 2(k-1)\frac{\gamma^2}{1-\rho} - (k-2)\frac{\gamma^2}{1-\rho}\right] \\ \vdots \\ \frac{(1-\rho)^2}{\gamma^2}\left[-(k-p+3)\frac{\gamma^2}{1-\rho} + 2(k-p+2)\frac{\gamma^2}{1-\rho} - (k-p+1)\frac{\gamma^2}{1-\rho}\right] \\ \frac{(1-\rho)^2}{\gamma^2}\left[-(k-p+2)\frac{\gamma^2}{1-\rho} + \frac{k-p+2}{k-p+1}(k-p+1)\frac{\gamma^2}{1-\rho}\right] \end{bmatrix}$$

$$\times \begin{bmatrix} \frac{\gamma}{T-p-k+1} \\ (1-\rho)\frac{T-p-k}{T-p-k+1} \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}.$$

Hence

$$\begin{bmatrix} \widehat{\tau} - \tau_T \\ \widehat{\rho}_1 - \rho_1 \\ \vdots \\ \widehat{\rho}_p - \rho_p \end{bmatrix}' \begin{bmatrix} 1 \\ y_{T-1} \\ \vdots \\ y_{T-p} \end{bmatrix} = \frac{\gamma}{T-p-k+1} - \gamma + (1-\rho)\frac{T-p-k}{T-p-k+1} y_{T-1}$$

$$\to -\gamma + (1-\rho)\left[\frac{\gamma}{1-\rho} + Y\right] = (1-\rho)\, Y.$$

This proves the first result in expression (14a).

Now by contrast when $\sigma_\epsilon^2 >> \gamma$, the inverse of the empirical second moment matrix uses:

$$\begin{bmatrix} 1 & \cdots & \rho^{p-1} \\ \vdots & \ddots & \vdots \\ \rho^{p-1} & \cdots & 1 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{1-\rho^2} & -\frac{\rho}{1-\rho^2} & 0 & \cdots & 0 \\ -\frac{\rho}{1-\rho^2} & \frac{1+\rho^2}{1-\rho^2} & -\frac{\rho}{1-\rho^2} & \ddots & \vdots \\ 0 & -\frac{\rho}{1-\rho^2} & \ddots & & 0 \\ \vdots & \ddots & \ddots & \frac{1+\rho^2}{1-\rho^2} & -\frac{\rho}{1-\rho^2} \\ 0 & \cdots & 0 & -\frac{\rho}{1-\rho^2} & \frac{1}{1-\rho^2} \end{bmatrix},$$

so the matrix of interest is approximately

$$\begin{bmatrix} T-p+1 & \frac{\gamma}{1-\rho}\left[k \quad \cdots \quad k-p+1\right] \\ \frac{\gamma}{1-\rho}\begin{bmatrix} k \\ \vdots \\ k-p+1 \end{bmatrix} & T\frac{\sigma_\epsilon^2}{(1-\rho)^2}\begin{bmatrix} 1 & \cdots & \rho^{p-1} \\ \vdots & \ddots & \vdots \\ \rho^{p-1} & \cdots & 1 \end{bmatrix} \end{bmatrix}^{-1}.$$

We now use the formula for the partitioned matrix inverse, first letting:

$$M/D=(T-p+1)-\frac{\gamma^2}{(1-\rho^2)\,T\sigma_\epsilon^2}\begin{bmatrix} k \\ \vdots \\ k-p+1 \end{bmatrix}'\begin{bmatrix} (1-\rho)(k-1)+1 \\ \left(1-\rho^2\right)(k-1) \\ \vdots \\ \left(1-\rho^2\right)(k-p+2) \\ (1-\rho)(k-p+2)+1 \end{bmatrix}$$

$$=(T-p+1)-\frac{\gamma^2}{(1+\rho)\,T\sigma_\epsilon^2}\left[k(k-1)\left[1+(1+\rho)\frac{2k-1}{6}\right]\right.$$

$$\left.+(k-p+1)(k-p+2)\left[1-(1+\rho)\frac{2(k-p)+3}{6}\right]\right]$$

$$-\frac{\gamma^2}{(1-\rho^2)\,T\sigma_\epsilon^2}\left[2k-p+1\right],$$

i.e. when $k=\lfloor c_k T\rfloor$,

$$M/D\underset{T\to\infty}{\sim}(T-p+1)-\frac{\gamma^2}{3\sigma_\epsilon^2}\left\{\frac{k^3-(k-p)^3}{T}\right\}.$$

It follows that the inverse of empirical second moments is given by a 2×2 matrix whose bottom-right element is:

$$\frac{(1-\rho)^2}{T\sigma_\epsilon^2}\begin{bmatrix} \frac{1}{1-\rho^2} & -\frac{\rho}{1-\rho^2} & 0 & \cdots & 0 \\ -\frac{\rho}{1-\rho^2} & \frac{1+\rho^2}{1-\rho^2} & -\frac{\rho}{1-\rho^2} & \ddots & \vdots \\ 0 & -\frac{\rho}{1-\rho^2} & \ddots & & 0 \\ \vdots & \ddots & \ddots & \frac{1+\rho^2}{1-\rho^2} & -\frac{\rho}{1-\rho^2} \\ 0 & \cdots & 0 & -\frac{\rho}{1-\rho^2} & \frac{1}{1-\rho^2} \end{bmatrix}$$

$$+\frac{1}{T\left\{k^3-(k-p)^3\right\}}\frac{3\gamma^2}{(1+\rho)^2\,\sigma_\epsilon^2}\begin{bmatrix} (1-\rho)(k-1)+1 \\ \left(1-\rho^2\right)(k-1) \\ \vdots \\ \left(1-\rho^2\right)(k-p+2) \\ (1-\rho)(k-p+2)+1 \end{bmatrix}\begin{bmatrix} (1-\rho)(k-1)+1 \\ \left(1-\rho^2\right)(k-1) \\ \vdots \\ \left(1-\rho^2\right)(k-p+2) \\ (1-\rho)(k-p+2)+1 \end{bmatrix}'.$$

Hence the limit of the OLS is for large p (so $k^3-(k-p)^3$ is truly $O\left(k^3\right)$)

$$\left[\begin{array}{cc} O\left(k^{-3}\right) & O\left(Tk^{-2}\right) \\ O\left(Tk^{-2}\right) & D^{-1}+O\left(T^{-1}k^{-3}\right) \end{array}\right]\left[\begin{array}{c} (k+1)\,\gamma \\ k\frac{\gamma^2}{1-\rho} \\ \vdots \\ (k-p+1)\,\frac{\gamma^2}{1-\rho} \end{array}\right],$$

since, despite $\gamma << \sigma_\epsilon^2$, the vector

$$\left[\begin{array}{c} \sum_{t=p}^{T}\tau_t+\epsilon_t \\ \sum_{t=p}^{T}y_{t-1}\left(\tau_t+\epsilon_t\right) \\ \vdots \\ \sum_{t=p}^{T}y_{t-p}\left(\tau_t+\epsilon_t\right) \end{array}\right]=\left[\begin{array}{c} \sum_{t=p}^{T}\tau_t \\ \sum_{t=p}^{T}y_{t-1}\tau_t \\ \vdots \\ \sum_{t=p}^{T}y_{t-p}\tau_t \end{array}\right]+O_p\left(\sqrt{T}\right)$$

is dominated by the effect of the breaks. We see that

$$\begin{aligned}
&\frac{3T}{\gamma(1+\rho)\left\{k^3-(k-p)^3\right\}}\left[\begin{array}{c} (1-\rho)(k-1)+1 \\ \left(1-\rho^2\right)(k-1) \\ \vdots \\ \left(1-\rho^2\right)(k-p+2) \\ (1-\rho)(k-p+2)+1 \end{array}\right]'\left[\begin{array}{c} k\frac{\gamma^2}{1-\rho} \\ \vdots \\ (k-p+1)\,\frac{\gamma^2}{1-\rho} \end{array}\right] \\
&=\frac{3T\gamma}{(1+\rho)\left\{k^3-(k-p)^3\right\}}\left[\frac{\rho}{1-\rho}k-\rho k^2\right. \\
&+\left.\begin{array}{c} (1+\rho)\left[\frac{k(k+1)(2k+1)}{6}-\frac{(k-p)(k-p+1)(2k-2p+1)}{6}\right] \\ +\left[-\rho(k-p+1)^2+\frac{\rho}{1-\rho}(k-p+1)\right] \end{array}\right] \\
&=3T\gamma+O(T/k),
\end{aligned}$$

and

$$\frac{\gamma^2}{1-\rho} D^{-1} \begin{bmatrix} k \\ \vdots \\ k-p+1 \end{bmatrix} = \frac{\gamma^2}{1-\rho} \frac{(1-\rho)^2}{T\sigma_\epsilon^2} \frac{1}{1-\rho^2} \begin{bmatrix} (1-\rho)(k-1)+1 \\ \left(1-\rho^2\right)(k-1) \\ \vdots \\ \left(1-\rho^2\right)(k-p+2) \\ (1-\rho)(k-p+2)+1 \end{bmatrix}$$

$$= \frac{\gamma^2}{T(1+\rho)\sigma_\epsilon^2} \begin{bmatrix} (1-\rho)(k-1)+1 \\ \left(1-\rho^2\right)(k-1) \\ \vdots \\ \left(1-\rho^2\right)(k-p+2) \\ (1-\rho)(k-p+2)+1 \end{bmatrix},$$

so

$$\begin{bmatrix} \widehat{\tau} \\ \widehat{\rho}_1 - \rho_1 \\ \vdots \\ \widehat{\rho}_p - \rho_p \end{bmatrix} \sim \begin{bmatrix} 3T\gamma \\ \frac{\gamma^2}{\sigma_\epsilon^2} \begin{bmatrix} \frac{1-\rho}{1+\rho}\left(\frac{k-1}{T}\right)+1 \\ (1-\rho)\left(\frac{k-1}{T}\right) \\ \vdots \\ (1-\rho)\left(\frac{k-p+2}{T}\right) \\ \frac{1-\rho}{1+\rho}\left(\frac{k-p+2}{T}\right)+1 \end{bmatrix} \end{bmatrix}. \tag{15}$$

Now the forecast errors are obtained from

$$\begin{bmatrix} \widehat{\tau} - \tau_T \\ \widehat{\rho}_1 - \rho_1 \\ \vdots \\ \widehat{\rho}_p - \rho_p \end{bmatrix}' \begin{bmatrix} 1 \\ y_{T-1} \\ \vdots \\ y_{T-p} \end{bmatrix}$$

$$\underset{T\to\infty}{\sim} T\left(2\gamma + \frac{\gamma^2}{\sigma_\epsilon^2}\left[\gamma + Y(1-\rho)\right]\int_0^{c_p}(c_k-u)\,du\right)$$

$$= T\left(2\gamma + \frac{\gamma^2}{2\sigma_\epsilon^2}(2c_k-c_p)\,c_p\left[\gamma + Y(1-\rho)\right]\right),$$

and hence expression (14b). Finally, when considering the multistep estimators, as long as $h+p<k$, the asymptotics is the same, replacing k with $k-h$, c_k with c_k-c_h.

	Variables				**Horizons**				
	Infl.	*IP*	*GDP*	*IR*	*1*	*2*	*4*	*8*	*12*
					AR(1)				
IMS (*)	**.29**	**.28**	**.32**	**.33**	**.44**	**.33**	**.22**	**.28**	**.17**
DMS	.56	.59	.59	.63	1.0	.59	.55	.41	.39
IMSIC	.28	.23	.37	.29	.23	.30	.32	.29	.32
DMSIC	.30	.24	.36	.28	.23	.27	.30	.33	.33
all IMS	.24	.31	.39	.37	.33	.32	.33	.32	.31
all DMS	.31	.39	.42	.40	.44	.37	.38	.35	.36
all PRE	.17	.37	.37	.30	.24	.27	.34	.32	.32
all POST	.31	.34	.42	.47	.47	.38	.36	.35	.35
all IC	.28	.24	.37	.32	.28	.30	.31	.30	.31
					AR(4)				
IMS (*)	**.27**	**.28**	**.32**	**.29**	**.56**	**.28**	**.11**	**.17**	**.22**
DMS	.52	.50	.59	.59	1.0	.44	.55	.38	.36
IMSIC	.26	.23	.35	.23	.21	.26	.29	.26	.30
DMSIC	.33	.26	.35	.29	.21	.30	.35	.33	.33
all IMS	.20	.30	.37	.26	.26	.27	.30	.29	.28
all DMS	.35	.38	.42	.43	.38	.33	.44	.41	.41
all PRE	.23	.37	.37	.32	.23	.28	.38	.36	.35
all POST	.28	.33	.39	.37	.30	.29	.37	.37	.36
all IC	.34	.24	.34	.32	.26	.29	.33	.33	.34
					AR(p)				
IMS (*)	**.50**	**.49**	**.53**	**.51**	**.72**	**.61**	**.39**	**.39**	**.39**
DMS	.56	.61	.60	.65	1.0	.61	.62	.44	.35
IMSIC	.29	.23	.36	.49	.30	.35	.35	.35	.33
DMSIC	.31	.24	.35	.35	.30	.29	.32	.35	.29
all IMS	.18	.24	.34	.21	.22	.23	.25	.25	.24
all DMS	.22	.30	.38	.25	.35	.28	.31	.27	.23
all PRE	.10	.24	.33	.12	.17	.18	.21	.21	.20
all POST	.17	.25	.33	.15	.23	.22	.23	.22	.20
all IC	.21	.23	.34	.21	.24	.24	.26	.26	.23
					ALL MODELS				
IMS (*)	**.34**	**.38**	**.47**	**.25**	**.60**	**.39**	**.24**	**.25**	**.24**
DMS	.54	.58	.58	.61	1.0	.55	.56	.41	.36
IMSIC	.27	.23	.36	.31	.25	.29	.31	.30	.31
DMSIC	.31	.25	.35	.30	.25	.28	.33	.34	.31
all IMS	.22	.29	.37	.28	.28	.28	.29	.30	.29
all DMS	.31	.37	.41	.37	.40	.34	.39	.35	.34
all PRE	.18	.35	.37	.25	.22	.26	.32	.31	.30
all POST	.27	.32	.39	.34	.35	.31	.33	.33	.32
all IC	.30	.25	.35	.29	.27	.28	.30	.31	.31

	Variables				**Horizons**				
	Infl.	*IP*	*GDP*	*IR*	*1*	*2*	*4*	*8*	*12*
					$T = 25$				
IMS (*)	**.38**	**.39**	**.73**	**.31**	**.68**	**.64**	**.45**	**.42**	**.43**
DMS	.37	.43	.36	.51	.99	.43	.35	.18	.11
IMSIC	.14	.00	.00	.22	.05	.06	.11	.09	.14
DMSIC	.28	.03	.00	.25	.05	.07	.18	.24	.18
all IMS	.06	.07	.03	.17	.08	.07	.09	.08	.09
all DMS	.23	.20	.10	.32	.23	.16	.22	.22	.23
all PRE	.07	.16	.02	.19	.05	.07	.14	.14	.16
all POST	.15	.11	.08	.26	.13	.13	.14	.17	.18
all IC	.18	.03	.01	.23	.08	.07	.12	.13	.16
					$T = 40$				
IMS (*)	**.39**	**.43**	**.49**	**.31**	**.55**	**.49**	**.48**	**.50**	**.45**
DMS	.58	.38	.39	.58	1.0	.49	.41	.23	.30
IMSIC	.18	.03	.08	.40	.15	.18	.14	.16	.20
DMSIC	.18	.04	.06	.31	.15	.14	.11	.16	.16
all IMS	.11	.08	.08	.23	.15	.12	.12	.12	.12
all DMS	.26	.15	.14	.29	.29	.20	.20	.17	.19
all PRE	.11	.13	.08	.11	.06	.11	.14	.11	.12
all POST	.19	.09	.09	.29	.23	.14	.14	.15	.16
all IC	.17	.04	.05	.28	.13	.14	.11	.13	.15
					RLS				
IMS (*)	**.10**	**.04**	NA	**.07**	**.16**	**.07**	**.01**	**.01**	**.06**
DMS	.67	.93	NA	.75	1.0	.73	.93	.83	.67
IMSIC	.49	.67	NA	.32	.55	.64	.67	.64	.59
DMSIC	.48	.68	NA	.34	.55	.64	.68	.63	.60
all IMS	.47	.73	NA	.42	.60	.63	.68	.69	.64
all DMS	.43	.77	NA	.50	.67	.65	.74	.67	.61
all PRE	.35	.75	NA	.45	.55	.62	.68	.68	.63
all POST	.48	.75	NA	.46	.68	.65	.71	.67	.62
all IC	.53	.67	NA	.36	.59	.63	.68	.66	.61

(*): the lines denoted by IMS report the proportion of cases where IMS is at least as accurate as any other technique. All other entries report the proportion of cases where the corresponding forecasting technique is at least as accurate as IMS. The variables are abbreviated as follows: infl. is inflation, IP is Industrial Production Growth, GDP is real GDP growth and IR is the interest rate. The entries whose line is referred to starting with "all" denote the proportion over all the variants of the quoted technique, excluding the two standard IMS and DMS.

Table 1: Comparison of empirical out-of-sample relative forecasting performances across techniques, models, variables and horizons.

			$h=1$	2	5	10	20	1	2	5	10	20	1	2	5	10	20	1	2	5	10	20	1	2	5	10	20	1	2	5	10	20
Without Intercept Correction																																
	$\bar{k}$	$\bar{b}$	RMSFE IMS / σ					RMSFE IMS/DMS					RMSFE IMS/IMSpre					RMSFE IMS/DMSpre					RMSFE IMS/IMSpost					RMSFE IMS/DMSpost				
Inflation																																
Canada	19.4	0.87	0.54	0.69	0.90	1.02	1.17	1.00	1.03	0.96	0.93	0.79	0.52	0.57	0.59	0.60	0.63	0.52	0.52	0.48	0.55	0.56	1.16	1.20	1.20	1.18	1.18	1.16	1.04	0.76	0.75	0.61
France	20.3	0.73	0.47	0.64	0.88	1.04	1.10	1.00	1.03	1.05	1.01	0.85	0.47	0.48	0.55	0.60	0.63	0.47	0.39	0.48	0.49	0.49	1.03	1.02	1.03	1.06	0.93	1.03	0.79	0.85	0.67	0.52
Germany	22.3	0.57	0.97	1.06	1.04	1.14	1.04	1.00	1.00	1.03	0.87	0.82	0.74	0.77	0.79	0.90	0.86	0.74	0.70	0.68	0.68	0.58	1.14	1.14	1.06	0.97	0.86	1.14	1.02	0.70	0.62	0.56
Japan	19.0	0.32	0.84	1.29	0.79	0.71	0.88	1.00	0.97	1.00	0.61	0.77	0.48	0.69	0.55	0.51	0.59	0.48	0.69	0.69	0.76	1.06	0.60	0.64	0.22	0.08	0.04	0.60	0.63	0.54	0.61	1.03
UK	20.1	0.50	0.54	0.81	1.05	1.29	1.50	1.00	1.01	0.98	1.05	1.00	0.55	0.61	0.68	0.77	0.80	0.55	0.64	0.61	0.71	0.81	1.12	1.15	1.13	1.10	1.21	1.12	0.82	0.67	0.70	0.70
US	19.4	0.52	0.59	0.73	0.90	1.18	1.27	1.00	1.01	1.01	1.02	1.01	0.55	0.55	0.59	0.67	0.66	0.55	0.52	0.56	0.65	0.68	1.10	1.08	0.99	0.74	0.62	1.10	0.66	0.60	0.85	0.66
Industrial production growth																																
Canada	15.7	0.20	0.75	1.07	0.99	0.96	0.49	1.00	0.99	0.96	0.93	1.06	1.00	0.94	1.02	0.76	0.39	1.00	0.99	1.38	0.96	0.85	0.74	0.73	0.64	0.76	0.56	0.74	0.89	0.66	0.78	0.80
France	22.9	0.18	0.72	0.80	1.30	0.76	0.79	1.00	1.02	0.96	0.94	0.87	0.78	0.91	1.05	0.88	1.07	0.78	0.91	1.12	0.86	0.92	1.24	0.99	0.94	0.86	0.86	1.24	0.94	0.89	1.01	0.96
Germany	20.4	0.21	1.05	0.93	0.96	0.88	0.93	1.00	0.96	0.94	0.94	0.86	0.81	0.83	0.69	0.28	0.07	0.81	0.89	0.90	1.00	0.80	1.24	0.96	0.62	0.56	0.55	1.24	0.84	0.96	0.90	1.04
Japan	23.6	0.15	0.83	1.07	0.99	0.68	0.59	1.00	0.99	0.97	0.87	0.86	0.79	0.83	0.81	0.56	0.52	0.79	0.88	1.12	0.65	0.76	1.17	0.95	0.44	0.18	0.11	1.17	0.93	0.96	0.75	1.18
UK	18.9	0.09	0.66	0.47	0.50	0.67	0.44	1.00	1.04	1.30	0.91	0.88	0.68	1.15	0.74	1.29	1.18	0.68	0.97	1.22	1.12	1.09	1.24	0.56	0.58	0.65	0.47	1.24	0.44	0.89	0.84	0.85
US	17.4	0.15	0.88	0.74	0.44	0.56	0.40	1.00	1.01	0.96	0.75	1.13	1.09	1.09	0.65	1.01	0.66	1.09	1.14	0.73	0.79	0.76	1.00	0.78	0.50	0.50	0.48	1.00	0.91	0.83	0.62	0.91
Real GDP growth																																
Canada	21.5	0.16	0.81	0.93	0.65	0.93	0.76	1.00	1.03	0.93	0.87	0.89	0.78	0.83	0.82	0.77	0.60	0.78	0.96	0.97	0.91	0.74	0.96	0.65	0.41	0.58	0.51	0.96	0.68	0.50	0.77	0.61
Germany	20.3	0.44	0.83	0.78	0.80	0.91	0.72	1.00	0.97	0.90	0.98	1.01	0.81	0.86	0.84	0.88	0.82	0.81	0.81	0.83	0.98	0.93	1.30	0.97	0.87	0.87	0.77	1.30	0.94	0.94	0.96	0.89
Japan	20.7	0.60	0.63	0.78	0.54	0.68	0.75	1.00	1.08	1.12	0.94	0.99	0.54	0.81	0.65	0.71	0.75	0.54	0.84	0.64	0.73	0.75	1.33	1.20	1.17	1.09	1.01	1.33	1.11	0.83	1.02	0.91
UK	21.4	0.18	0.60	0.56	0.64	0.60	0.45	1.00	0.95	1.00	0.88	0.82	0.66	0.75	1.01	1.33	0.93	0.66	0.63	0.99	1.16	0.90	1.20	0.89	0.78	0.61	0.50	1.20	0.67	0.76	0.61	0.51
US	18.2	0.20	1.34	1.36	1.13	1.03	1.09	1.00	0.98	0.90	0.93	1.02	0.89	0.88	0.83	0.88	0.83	0.89	0.98	0.82	1.00	0.96	1.11	0.96	0.80	0.83	0.89	1.11	0.88	0.85	0.85	1.20
Interest rate																																
Canada	19.4	0.56	0.68	0.88	1.13	1.16	1.28	1.00	1.04	0.99	1.02	0.97	0.57	0.64	0.73	0.73	0.78	0.57	0.68	0.61	0.66	0.70	1.06	1.07	1.11	1.15	1.09	1.06	1.06	1.06	0.97	0.83
France	22.4	0.76	0.57	0.77	0.86	1.06	1.13	1.00	0.99	1.25	0.96	0.97	0.44	0.50	0.52	0.60	0.62	0.44	0.45	0.49	0.51	0.63	1.24	1.40	1.54	1.64	1.44	1.24	1.00	0.74	0.72	0.74
Germany	12.5	0.27	0.55	1.67	0.67	0.78	0.98	1.00	0.98	0.89	0.78	1.18	0.41	0.87	0.75	0.69	0.61	0.41	0.71	0.40	0.49	0.48	0.98	1.13	0.76	1.06	1.00	0.98	0.71	0.68	0.59	0.85
UK	19.8	0.56	0.62	0.89	1.23	1.04	1.11	1.00	1.14	1.06	1.34	1.27	0.45	0.59	0.67	0.62	0.63	0.45	0.62	0.58	0.76	0.75	1.02	1.19	1.22	1.52	1.52	1.02	1.34	1.17	1.43	1.07
US	19.8	0.74	0.91	1.13	1.14	1.40	1.45	1.00	1.01	1.08	0.91	0.94	0.61	0.67	0.67	0.80	0.85	0.61	0.53	0.71	0.75	0.90	1.09	1.13	1.13	1.07	1.03	1.09	1.06	1.00	0.92	1.01
With Intercept Correction																																
	$\bar{k}$	$\bar{b}$	RMSFE IMS/IMSIC					RMSFE IMS/DMSIC					RMSFE IMS/IMSpreIC					RMSFE IMS/DMSpreIC					RMSFE IMS/IMSpostIC					RMSFE IMS/DMSpostIC				
Inflation																																
Canada	19.4	0.87	0.72	0.82	0.94	1.00	1.00	0.72	0.80	0.89	0.99	1.05	0.54	0.70	0.70	0.61	0.56	0.54	0.81	1.01	1.05	1.07	0.80	0.85	0.73	0.58	0.60	0.80	0.98	1.05	1.12	1.07
France	20.3	0.73	0.74	0.92	0.97	1.01	1.03	0.74	0.91	0.93	0.99	0.97	0.50	0.79	0.90	0.67	0.58	0.50	0.67	0.95	1.07	1.05	0.81	0.74	0.60	0.56	0.56	0.81	1.03	1.18	1.20	1.19
Germany	22.3	0.57	0.68	0.74	0.83	0.75	0.72	0.68	0.72	0.74	0.74	0.75	0.65	0.79	0.76	0.65	0.54	0.65	0.76	0.83	0.81	0.82	0.77	0.77	0.67	0.52	0.42	0.77	0.83	0.85	0.81	0.75
Japan	19.0	0.32	0.53	0.85	0.49	0.48	0.57	0.53	0.75	0.51	0.94	0.42	0.92	0.87	0.39	0.33	0.37	0.92	1.04	0.78	0.69	0.40	0.47	0.40	0.15	0.06	0.04	0.47	0.97	0.56	0.50	0.58
UK	20.1	0.50	0.85	1.00	1.38	1.04	1.13	0.85	0.98	1.33	1.10	1.02	0.75	0.91	0.90	0.76	0.74	0.75	1.05	1.32	1.04	1.15	0.88	1.03	0.82	0.76	0.95	0.88	1.06	1.28	1.13	1.17
US	19.4	0.52	0.66	0.69	0.77	0.69	0.76	0.66	0.69	0.77	0.89	0.98	0.42	0.56	0.52	0.52	0.50	0.42	0.61	0.67	0.92	1.02	0.74	0.59	0.41	0.38	0.38	0.74	0.89	0.93	0.98	0.96
Industrial production growth																																
Canada	15.7	0.20	0.86	0.71	0.66	0.63	0.39	0.86	0.67	0.72	0.65	0.44	0.66	0.63	0.53	0.49	0.26	0.66	0.69	0.67	0.63	0.40	0.89	0.58	0.56	0.60	0.32	0.89	0.70	0.67	0.55	0.36
France	22.9	0.18	0.93	0.91	0.95	0.64	0.88	0.93	0.85	0.91	0.60	0.97	0.95	0.85	0.90	0.63	0.83	0.95	0.93	0.98	0.62	0.97	1.00	0.88	0.92	0.56	0.72	1.00	0.86	0.91	0.66	0.88
Germany	20.4	0.21	1.16	0.90	0.62	0.62	0.69	1.16	0.88	0.58	0.70	0.79	1.25	0.73	0.51	0.26	0.07	1.25	0.92	0.58	0.71	0.79	1.16	0.74	0.43	0.39	0.40	1.16	0.85	0.63	0.62	0.72
Japan	23.6	0.15	0.78	0.95	0.56	0.38	0.45	0.78	0.95	0.71	0.58	0.35	0.66	0.80	0.55	0.36	0.39	0.66	0.74	0.67	0.63	0.50	0.79	0.74	0.33	0.15	0.10	0.79	1.05	0.70	0.43	0.47
UK	18.9	0.09	0.87	0.45	0.38	0.67	0.36	0.87	0.44	0.45	0.56	0.36	0.93	0.42	0.34	0.65	0.34	0.93	0.45	0.46	0.56	0.38	1.04	0.38	0.30	0.47	0.27	1.04	0.38	0.38	0.67	0.36
US	17.4	0.15	0.73	0.72	0.38	0.41	0.38	0.73	0.82	0.42	0.42	0.38	0.63	0.57	0.34	0.41	0.36	0.63	0.78	0.42	0.43	0.41	0.79	0.53	0.25	0.29	0.23	0.79	0.73	0.39	0.41	0.38
Real GDP growth																																
Canada	21.5	0.16	0.84	0.56	0.42	0.54	0.45	0.84	0.57	0.44	0.54	0.42	0.82	0.51	0.36	0.46	0.36	0.82	0.60	0.43	0.56	0.42	0.86	0.51	0.34	0.44	0.38	0.86	0.56	0.41	0.56	0.46
Germany	20.3	0.44	1.07	0.77	0.74	0.73	0.59	1.07	0.71	0.58	0.69	0.62	1.10	0.76	0.68	0.67	0.53	1.10	0.73	0.61	0.72	0.60	1.06	0.72	0.67	0.64	0.52	1.06	0.74	0.68	0.73	0.62
Japan	20.7	0.60	0.92	1.06	0.87	0.86	0.86	0.92	0.98	0.81	0.95	0.82	0.98	0.98	0.57	0.50	0.52	0.98	1.04	0.87	0.95	0.87	0.93	0.98	0.74	0.64	0.65	0.93	1.06	0.88	0.87	0.86
UK	21.4	0.18	1.33	0.86	0.85	0.58	0.42	1.33	0.83	0.82	0.62	0.45	1.23	0.81	0.87	0.60	0.39	1.23	0.95	0.88	0.65	0.40	1.39	0.88	0.79	0.55	0.40	1.39	0.67	0.79	0.58	0.42
US	18.2	0.20	0.71	0.80	0.64	0.61	0.69	0.71	0.72	0.56	0.57	0.66	0.82	0.68	0.55	0.48	0.50	0.82	0.73	0.54	0.57	0.62	0.72	0.69	0.53	0.51	0.49	0.72	0.74	0.61	0.57	0.68
Interest rate																																
Canada	19.4	0.56	0.73	0.88	0.88	0.97	1.01	0.73	0.86	0.85	0.87	0.94	0.65	0.92	0.96	0.99	1.03	0.65	0.89	1.06	0.98	1.03	0.85	0.94	0.74	0.68	0.66	0.85	1.01	0.98	1.00	1.02
France	22.4	0.76	0.92	1.02	1.06	1.41	1.30	0.92	1.02	0.95	1.18	1.19	0.86	1.16	1.17	1.36	0.89	0.86	1.10	1.32	1.14	1.28	1.07	1.04	0.81	0.60	0.54	1.07	1.35	1.22	1.44	1.33
Germany	12.5	0.27	0.78	1.12	0.55	0.67	0.68	0.78	1.08	0.52	0.61	0.73	0.76	1.12	0.53	0.57	0.68	0.76	1.15	0.72	0.69	0.69	0.84	1.21	0.33	0.46	0.44	0.84	1.11	0.58	0.76	0.67
UK	19.8	0.56	0.89	1.26	1.54	1.49	2.13	0.89	1.16	1.43	1.30	1.62	0.73	1.18	1.35	1.40	1.36	0.73	1.12	1.40	1.38	1.83	1.03	1.30	1.39	1.41	1.21	1.03	1.24	1.23	1.54	2.31
US	19.8	0.74	0.76	0.82	1.01	0.97	0.97	0.76	0.81	0.83	0.88	0.81	0.79	0.95	1.01	0.92	0.86	0.79	1.00	1.00	0.98	0.97	0.87	0.91	0.91	0.80	0.68	0.87	0.99	1.02	1.01	0.98

Table 2: Out-of-sample performance ordered by variable: the table reports the ratios of root-MSFE for AR(1) models estimated over windows of 40 observations.

AR(1) model

Without Intercept Correction

$k \in$	$\bar{k}$	RMSFE IMS / σ				RMSFE IMS/DMS				RMSFE IMS/IMSpre				RMSFE IMS/DMSpre				RMSFE IMS/IMSpost				RMSFE IMS/DMSpost			
	$h \in$	{1}	{2}	[3, 6]	[7, 12]	{1}	{2}	[3, 6]	[7, 12]	{1}	{2}	[3, 6]	[7, 12]	{1}	{2}	[3, 6]	[7, 12]	{1}	{2}	[3, 6]	[7, 12]	{1}	{2}	[3, 6]	[7, 12]
[4, 7]	6.4	1.03	1.39	1.07	1.07	1.00	1.00	0.92	0.88	0.79	0.89	0.92	0.95	0.79	0.88	0.90	0.86	1.10	0.93	0.43	0.14	1.10	0.74	0.64	0.71
[8, 11]	9.4	0.82	0.87	1.03	0.99	1.00	1.01	1.05	0.80	0.73	0.78	0.81	0.85	0.73	0.74	0.78	0.74	1.11	0.97	0.93	0.83	1.11	0.93	0.65	0.65
[12, 15]	13.5	0.59	0.79	0.87	1.14	1.00	1.05	1.09	0.93	0.62	0.67	0.68	0.79	0.62	0.63	0.64	0.69	1.07	1.10	1.08	0.99	1.07	1.04	0.85	0.69
[16, 19]	17.6	0.58	0.73	0.90	1.11	1.00	1.03	1.06	1.06	0.58	0.60	0.63	0.69	0.58	0.54	0.58	0.62	1.09	1.08	1.07	1.19	1.09	1.05	0.98	0.68
[20, 29]	24.6	0.65	0.78	0.96	1.17	1.00	1.02	1.05	1.06	0.53	0.57	0.61	0.69	0.53	0.50	0.56	0.65	1.02	1.04	1.05	1.04	1.02	1.02	1.04	0.87
[30, 39]	32.5	0.76	0.87	1.00	0.94	1.00	0.98	0.97	0.99	0.53	0.59	0.60	0.27	0.53	0.53	0.63	0.67	1.03	1.04	0.99	1.01	1.03	0.98	0.94	1.00

With Intercept Correction

		RMSFE IMS/IMSIC				RMSFE IMS/DMSIC				RMSFE IMS/IMSpreIC				RMSFE IMS/DMSpreIC				RMSFE IMS/IMSpostIC				RMSFE IMS/DMSpostIC			
[4, 7]	6.4	0.97	1.00	0.67	0.68	0.97	0.97	0.66	0.69	1.01	1.03	0.73	0.75	1.01	1.06	0.70	0.74	0.91	0.59	0.28	0.13	0.91	1.04	0.75	0.75
[8, 11]	9.4	0.82	0.75	0.87	0.77	0.82	0.71	0.84	0.83	0.75	0.73	0.91	0.87	0.75	0.72	0.89	0.94	0.89	0.80	0.48	0.38	0.89	0.82	0.94	0.88
[12, 15]	13.5	0.80	0.92	0.98	0.98	0.80	0.89	0.82	0.90	0.61	0.89	0.98	0.94	0.61	0.90	0.95	0.99	0.90	1.01	0.72	0.57	0.90	0.98	0.96	1.00
[16, 19]	17.6	0.70	0.80	0.96	1.04	0.70	0.79	0.84	0.91	0.57	0.78	0.96	0.81	0.57	0.80	0.90	1.02	0.80	0.89	1.01	0.65	0.80	0.89	0.95	1.06
[20, 29]	24.6	0.67	0.78	0.91	0.93	0.67	0.75	0.85	0.88	0.63	0.84	0.79	0.64	0.63	0.79	0.89	0.94	0.74	0.85	0.95	0.83	0.74	0.88	0.96	0.95
[30, 39]	32.5	0.68	0.76	0.94	0.86	0.68	0.74	0.90	0.88	0.68	0.61	0.50	0.25	0.68	0.77	0.96	0.87	0.75	0.82	0.98	0.87	0.75	0.82	0.93	0.89

AR(4)

Without Intercept Correction

$k \in$	$\bar{k}$	RMSFE IMS / σ				RMSFE IMS/DMS				RMSFE IMS/IMSpre				RMSFE IMS/DMSpre				RMSFE IMS/IMSpost				RMSFE IMS/DMSpost			
	$h \in$	{1}	{2}	[3, 6]	[7, 12]	{1}	{2}	[3, 6]	[7, 12]	{1}	{2}	[3, 6]	[7, 12]	{1}	{2}	[3, 6]	[7, 12]	{1}	{2}	[3, 6]	[7, 12]	{1}	{2}	[3, 6]	[7, 12]
[4, 7]	6.4	1.03	1.43	1.89	5.35	1.00	0.99	1.33	3.21	0.81	0.94	1.59	4.71	0.81	0.92	1.47	3.73	0.56	0.66	1.12	3.56	0.56	0.66	1.12	3.56
[8, 11]	9.4	0.76	0.80	0.93	0.94	1.00	0.94	0.77	0.52	0.65	0.66	0.69	0.81	0.65	0.64	0.60	0.61	0.42	0.48	0.03	0.00	0.42	0.49	0.54	0.62
[12, 15]	13.5	0.58	0.76	0.81	1.13	1.00	0.98	0.93	0.71	0.55	0.59	0.58	0.78	0.55	0.57	0.50	0.61	0.80	0.87	0.65	0.19	0.80	0.74	0.58	0.69
[16, 19]	17.6	0.54	0.68	0.87	1.06	1.00	0.97	0.97	0.91	0.40	0.41	0.52	0.66	0.40	0.37	0.43	0.57	0.91	0.89	0.93	1.09	0.91	0.86	0.70	0.60
[20, 29]	24.6	0.66	0.76	0.95	1.15	1.00	1.01	1.01	1.00	0.32	0.38	0.43	0.11	0.32	0.32	0.48	0.63	0.90	0.93	0.99	1.01	0.90	0.93	0.91	0.72
[30, 39]	32.5	0.76	0.85	1.01	0.92	1.00	0.99	0.97	0.97	0.09	0.01	0.00	0.00	0.09	0.57	0.65	0.66	0.95	0.93	0.97	0.96	0.95	0.91	0.93	0.90

With Intercept Correction

		RMSFE IMS/IMSIC				RMSFE IMS/DMSIC				RMSFE IMS/IMSpreIC				RMSFE IMS/DMSpreIC				RMSFE IMS/IMSpostIC				RMSFE IMS/DMSpostIC			
[4, 7]	6.4	1.06	1.01	0.77	0.94	1.06	1.06	1.21	2.89	1.00	1.08	1.24	3.64	1.00	1.11	1.30	3.66	1.24	1.09	1.32	3.74	1.24	1.09	1.32	3.74
[8, 11]	9.4	0.84	0.72	0.73	0.64	0.84	0.69	0.72	0.56	0.70	0.65	0.78	0.77	0.70	0.63	0.75	0.79	0.49	0.34	0.03	0.00	0.49	0.72	0.86	0.83
[12, 15]	13.5	0.82	0.87	0.83	0.85	0.82	0.85	0.78	0.78	0.57	0.78	0.81	0.79	0.57	0.73	0.69	0.83	0.79	0.87	0.40	0.18	0.79	0.76	0.85	1.00
[16, 19]	17.6	0.77	0.77	0.89	0.95	0.77	0.77	0.84	0.84	0.50	0.62	0.79	0.65	0.50	0.52	0.56	0.87	0.87	0.86	0.67	0.58	0.87	0.83	0.83	1.02
[20, 29]	24.6	0.72	0.75	0.84	0.89	0.72	0.77	0.84	0.86	0.44	0.52	0.44	0.11	0.44	0.40	0.71	0.93	0.76	0.83	0.91	0.72	0.76	0.84	0.86	0.92
[30, 39]	32.5	0.73	0.80	0.94	0.84	0.73	0.81	0.94	0.87	0.09	0.01	0.00	0.00	0.09	0.84	0.99	0.85	0.75	0.81	0.93	0.81	0.75	0.78	0.89	0.82

Table 3: Out-of-sample performance ordered by post-break window size k: the table reports the ratios of square-root Mean Square Forecast Errors (the latter weighted by the variances of the variables) for AR(1) (above) and AR(4) (below) models estimated over windows of 40 observations.