DISCUSSION PAPER SERIES

No. 4401

SMALL SAMPLE PROPERTIES OF FORECASTS FROM AUTOREGRESSIVE MODELS UNDER STRUCTURAL BREAKS

M Hashem Pesaran and
Allan G Timmermann

INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP4401.asp**

ISSN 0265-8003

SMALL SAMPLE PROPERTIES OF FORECASTS FROM AUTOREGRESSIVE MODELS UNDER STRUCTURAL BREAKS

M Hashem Pesaran, University of Cambridge
Allan G Timmermann, University of California, San Diego and CEPR

Discussion Paper No. 4401
June 2004

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 4401

June 2004

ABSTRACT

Small Sample Properties of Forecasts From Autoregressive Models Under Structural Breaks*

This Paper develops a theoretical framework for the analysis of small sample properties of forecasts from general autoregressive models under structural breaks. Finite-sample results for the mean-squared forecast error of one-step-ahead forecasts are derived, both conditionally and unconditionally, and numerical results for different types of break specifications are presented. It is established that forecast errors are unconditionally unbiased even in the presence of breaks in the autoregressive coefficients and/or error variances so long as the unconditional mean of the process remains unchanged. Insights from the theoretical analysis are demonstrated in Monte Carlo simulations and on a range of macroeconomic time series from G7 countries. The results are used to draw practical recommendations for the choice of estimation window when forecasting from autoregressive models subject to breaks.

JEL Classification: C22 and C53
Keywords: autoregression, MSFE, rolling window estimator, small sample properties of forecasts and structural breaks

M Hashem Pesaran
Faculty of Economics and Politics
University of Cambridge
Austin Robinson Building
Sidgwick Avenue
Cambridge
CB3 9DE
Tel: (44 1223) 335216
Fax: (44 1223) 335471
Email: hashem.pesaran@econ.cam.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=100830

Allan G Timmermann
Department of Economics
University of California, San Diego
9500 Gilman Drive
La Jolla CA 92093-0508
USA
Tel: (1 858) 534 4860
Fax: (1 858) 534 7040
Email: atimmerm@weber.ucsd.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=116464

*We are grateful to the editor, four referees, and seminar participants at Cass Business School (London) for helpful comments on an earlier version of the Paper. We would also like to thank Mutita Akusuwan for excellent research assistance.

Submitted 01 April 2004

1. **Introduction**

Autoregressive models are used extensively in forecasting throughout economics and finance and have proved so successful and difficult to outperform in practice that they are frequently used as benchmarks in forecast competitions. Due in large part to their relatively parsimonious form, autoregressive models are frequently found to produce smaller forecast errors than those associated with models that allow for more complicated nonlinear dynamics or additional predictor variables, c.f. Stock and Watson (1999) and Giacomini (2002).

Despite their empirical success, there is now mounting evidence that the parameters of autoregressive (AR) models fitted to many economic time series are unstable and subject to structural breaks. For example, Stock and Watson (1996) undertake a systematic study of a wide variety of economic time series and find that the majority of these are subject to structural breaks. Alogoskoufis and Smith (1991) and Garcia and Perron (1996) are other examples of studies that document instability related to the autoregressive terms in forecasting models. Clements and Hendry (1998) view structural instability as a key determinant of forecasting performance.

This suggests a need to study the behaviour of the parameter estimates of AR models as well as their forecasting performance when these models undergo breaks. Despite the interest in econometric models subject to structural breaks, little is known about the small sample properties of AR models that undergo discrete changes. In view of the widespread use of AR models in forecasting, this is clearly an important area to investigate. The presence of breaks makes the focus on small sample properties more relevant: even if the combined pre- and post-break sample is very large, the occurrence of a structural break means that the post-break sample will often be quite small so that asymptotic approximations may not be nearly as accurate as is normally the case.

A key question that arises in the presence of breaks is how much data to use to estimate the parameters of the forecasting model that minimizes a loss function such as root mean squared forecast error (RMSFE). We show that the RMSFE-minimizing estimation window crucially depends on the size of the break as well as its direction (i.e., does the break lead to higher or lower persistence) and which parameters it affects (i.e., the mean, variance or autoregressive slope parameters). In some situations the optimal estimation window trades off an increased bias introduced by using pre-break data against a reduction in forecast error variance resulting from using a longer window of the data. However, in other situations the small sample bias in the autoregressive coefficients may in fact be reduced after introducing pre-break data if the size of the break is small or even when the break is large provided that it is in the right direction (e.g., when persistence declines).

In the presence of parameter instability it is common to use a rolling window estimator that makes use of a fixed number of the most recent data points, although the size of the rolling window is based on pragmatic considerations rather than an empirical analysis of the underlying time series process. Another possibility would be to test for breaks in the parameters and/or error variances and only use data after the most recent break, assuming a break is in fact detected. Alternatively, if no statistically significant break is found, an expanding window estimator could be used. Our theoretical analysis allows us to better understand when each of these procedures is likely to work well and why it is generally best to use pre-break data when forecasting using autoregressive models. First, breaks in the autoregressive parameters need not introduce bias in the forecasts (at least unconditionally). This tends to happen when an autoregressive coefficient declines after a break or the break only occurs in the intercept or variance parameter. Including pre-break data in such cases will tend to lead to a decline in RMSFE due to both a smaller squared bias and a reduction in the variance of the parameter estimate. Furthermore, in practice, there is likely to be a considerable error in detecting and estimating the point of the break of the autoregressive model. This leads to a worse performance of a post-break estimation procedure but also makes determination of the length of a rolling window more difficult.

Several practical recommendations emerge from our analysis regarding the choice of estimation window when forecasting from autoregressive models. First, for the macroeconomic data examined here, in general it appears to be difficult in practice to outperform expanding or long rolling window estimation methods. Unlike the case with exogenous regressors, forecasts from autoregressive models can be seriously biased even if only post-break observations are used. Including pre-break data in estimation of autoregressive models can simultaneously reduce the bias and the variance of the forecast errors. In most applications where breaks are not too large, expanding window methods or rolling window procedures with relatively large window sizes are likely to perform well. This conclusion may not of course carry over to longer data sets, e.g. high frequency financial data with thousands of observations, where estimation uncertainty can be reduced more effectively than with the relatively short macroeconomic data considered here.

The main contributions of this paper are as follows. First, we present a new procedure for computing the exact small sample properties of the parameters of AR models of arbitrary order, thus extending the existing literature that has focused on the AR(1) model. Our approach allows for fixed or random starting points and considers stationary AR models as well as models with unit root dynamics. We allow for the possibility of the AR model to switch from a unit root process to a stationary one and *vice versa*. Such regime switches could be particularly relevant to the analysis of inflation in a number of OECD countries since the first oil price shock in early 1970's. In addition to considering properties such as bias in the parameters, we also consider the RMSFE in finite samples. Second, we extend existing results on exact small sample properties of AR models to allow for a break in the underlying data generating process. We establish that one-step ahead forecast errors from AR models are unconditionally unbiased even in the presence of breaks in the autoregressive coefficients and in the error variances so long as the unconditional mean of the process remains unchanged. Our results also apply to models with unit roots. This extends Fuller (1996)'s result obtained for AR models with fixed parameters, and generalizes a related finding due to Clements and Hendry (1999, pp.39-42). Third, we present extensive numerical results quantifying the effect of the sizes of the pre-break and post-break data windows on parameter bias and RMSFE. Fourth, we undertake an empirical analysis for a range of macroeconomic time series from the G7 countries that compares the forecasting performance of expanding window, rolling window and post-break estimators. This analysis which allows for multiple breaks at unknown times confirms that, at least for macroeconomic time series such as those considered here, it is generally best to use pre-break data in estimation of the forecasting model.

The outline of the paper is as follows. Section 2 provides a brief overview of the small sample properties of the first-order autoregressive model that has been extensively studied in the extant literature. Theoretical results allowing us to characterize the small sample distribution of the parameters and forecast errors of autoregressive models are introduced in Section 3. Section 4 presents numerical results for AR models subject to breaks and Section 5 presents empirical results for a range of macroeconomic time series. Section 6 concludes with a summary and a discussion of possible extensions to our work.

2. Small Sample Properties of Forecasts from Autoregressive Models

A large literature has studied small sample properties of estimates of the parameters of autoregressive models. The majority of studies has concentrated on deriving either exact or approximate small sample results for the distribution of $\hat{\alpha}_T$ and $\hat{\beta}_T$, the Ordinary Least Squares (OLS) estimators of α and β, in the first-order autoregressive (AR(1)) model

$$y_t = \alpha + \beta y_{t-1} + \sigma\varepsilon_t,\ t = 1, 2, ..., T, \quad \varepsilon_t \sim iid(0, 1). \tag{1}$$

Analysis of the small sample bias of $\hat{\beta}_T$ dates back to at least Bartlett (1946). Early studies focus on the stationary AR(1) model without an intercept ($\alpha = 0$, $|\beta| < 1$) but have been extended to higher order models with intercepts (Sawa (1978)) and exogenous regressors (Grubb and Symons (1987), Kiviet and Phillips (1993, 2003a)). Assuming stationarity ($|\beta| < 1$), $\hat{\beta}_T$ has been shown to have an asymptotic normal distribution and its finite-sample distribution has been studied by Phillips (1977) and Evans and Savin (1981). The case with a unit root, $\beta = 1$, has been studied by, inter alia, Banerjee, Dolado, Hendry and Smith (1986), Phillips (1987), Stock (1987), Abadir (1993) and Kiviet and Phillips (2003b).

To a forecaster, the bias in $\hat{\alpha}_T$ and $\hat{\beta}_T$ is of direct interest only to the extent that it might adversely influence the forecasting performance. Ullah (2003) provides an extensive discussion and survey of the properties of forecasts from the AR(1) model. Box and Jenkins (1970) characterized the asymptotic mean squared forecast error (MSFE) for a stationary first-order autoregressive process considering both the single-period and multi-period horizon. Assuming a stationary process, Copas (1966) used Monte Carlo methods to study the MSFE of least-squares and maximum likelihood estimators under Gaussian innovations.

In practice, the conditional forecast error is of more interest than the unconditional error since the data needed to compute conditional forecasts is always available. A comprehensive asymptotic analysis for the stationary AR(p) model is provided in Fuller and Hasza (1981) and Fuller (1996). Using Theorem 8.5.3 in

Fuller (1996) it is easily seen that, conditional on y_T,

$$\begin{aligned} MSFE(\hat{y}_{T+1}\,|y_T) &= E\left[(y_{T+1}-\hat{y}_{T+1})^2\,|y_T\right] \\ &= \sigma^2(1+\frac{1}{T})+\frac{1-\beta^2}{T}\left(y_T-\frac{\alpha}{1-\beta}\right)^2+O(T^{-3/2}). \end{aligned}$$

This yields the more familiar unconditional result

$$MSFE(\hat{y}_{T+1}) = E\left(y_{T+1}-\hat{y}_{T+1}\right)^2 = \sigma^2(1+\frac{2}{T})+O(T^{-3/2}).$$

Generalizations to AR(p) and multi-step forecasts are also provided in Fuller (1996, pp. 443-449), where it is established that the forecast error, $y_{T+1}-\hat{y}_{T+1}$, is unbiased in small samples assuming ε_t has a symmetric distribution and $E\left(|\hat{y}_{T+1}|\right) < \infty$. This is particularly noteworthy considering the often large small sample bias associated with estimates of the autoregressive parameters.

3. AR(p) Model in the Presence of Structural Breaks

In parallel with the work on the small sample properties of estimates of autoregressive models, important progress has been made in testing for and estimating both the time and the size of breakpoints, as witnessed by the recent work of Andrews (1993), Andrews and Ploberger (1996), Bai and Perron (1998, 2003), Banerjee, Lumsdaine and Stock (1992), Chu, Stinchcombe and White (1996), Chong (2001), Elliott and Muller (2002), Hansen (1992), Inclan and Tiao (1994) and Ploberger, Kramer and Kontrus (1989).

Building on this work we consider the small sample problem of estimation and forecasting with AR(p) models in the presence of structural breaks. For this purpose, we consider the following AR(p) model defined over the period $t = 1, 2, ..., T$; and assumed to have been subject to a single structural break at time T_1 :

$$y_t = \begin{cases} \alpha_1+\beta_{11}y_{t-1}+\beta_{12}y_{t-2}+...+\beta_{1p}y_{t-p}+\sigma_1\varepsilon_t, & \text{, for } t \le T_1, \\ \alpha_2+\beta_{21}y_{t-1}+\beta_{22}y_{t-2}+...+\beta_{2p}y_{t-p}+\sigma_2\varepsilon_t, & \text{, for } t > T_1, \end{cases} \quad . \tag{2}$$

As before $\varepsilon_t \sim iid(0,1)$ for all t. For the analysis of the unit root case it is also convenient to consider the following parameterization of the intercept terms, α_i:

$$\alpha_i = \mu_i(1-\beta_i^*), \quad i = 1, 2, \tag{3}$$

where $\beta_i^* = \sum_{j=1}^{p} \beta_{ij}, = \boldsymbol{\tau}_p' \boldsymbol{\beta}_i$, $\boldsymbol{\beta}_i = (\beta_{i1}, \beta_{i2}, ..., \beta_{ip})'$ and $\boldsymbol{\tau}_p$ is a $p \times 1$ unit vector. Note that $-(1 - \beta_i^*)$ also represents the coefficient of y_{t-1} in the error correction representation of (2).

This specification is quite general and allows for intercept and slope shifts, as well as a change in error variances immediately after $t = T_1$. It is also possible for the y_t process to contain a unit root (or be integrated of order 1) in one or both of the regimes. The integration property of y_t under the two regimes is governed by whether $\beta_i^* = 1$ or $\beta_i^* < 1$. More specifically, we shall assume that the roots of

$$\sum_{j=1}^{p} \lambda^j \beta_{ij} - 1 = 0, \text{ for } i = 1, 2, \tag{4}$$

lie on or outside the unit circle.[1] As μ_i is allowed to vary freely, the intercepts $\alpha_i = \mu_i(1 - \beta_i^*)$ are unrestricted when the underlying AR processes are stationary. However, to avoid the possibility of generating linear trends in the y_t process, the intercepts are restricted ($\alpha_i = 0$) in the presence of unit roots. In the stationary case μ_i represents the unconditional mean of y_t in regime i. In the unit root case μ_i is not identified and we have $E(\Delta y_t) = 0$.

Analysis of forecast errors from AR models subject to structural change have been recently addressed by Clements and Hendry (1998,1999). However, these authors either abstract from the problem of parameter uncertainty, or only allow for it assuming that the parameters of the AR model remain unchanged during the estimation period. Consider first the analysis provided in Clements and Hendry (1998, pp.168-171), where it is assumed that parameters are known and the break takes place immediately prior to the forecasting period. In this case the one-step ahead forecast error is given by

$$y_{T+1} - \tilde{y}_{T+1} = \mu_2(1 - \beta_2^*) - \mu_1(1 - \beta_1^*) + \mathbf{x}_T' (\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1) + \sigma_2 \varepsilon_{T+1},$$

where $\mathbf{x}_T = (y_T, y_{T-1}, ..., y_{T-p+1})'$, $(\mu_1, \boldsymbol{\beta}_1)$ are the parameters prior to the forecast period, and $(\mu_2, \boldsymbol{\beta}_2)$ are the parameters during the forecast period, here $T + 1$. Following Clement and Hendry and noting that $\beta_i^* = \boldsymbol{\tau}_p' \boldsymbol{\beta}_i$, it is easily verified that

$$y_{T+1} - \tilde{y}_{T+1} = (\mu_2 - \mu_1)(1 - \beta_2^*) + (\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1)' (\mathbf{x}_T - \mu_1 \boldsymbol{\tau}_p) + \sigma_2 \varepsilon_{T+1},$$

[1] Our analysis can also allow for the possibility of y_t being integrated of order two in one or both of the two regimes. But in this paper we shall only consider the unit root case explicitly.

and

$$E\left(y_{T+1}-\tilde{y}_{T+1}\right)=\left(\mu_{2}-\mu_{1}\right)\left(1-\beta_{2}^{*}\right)+\left(\boldsymbol{\beta}_{2}-\boldsymbol{\beta}_{1}\right)' E\left(\mathbf{x}_{T}-\mu_{1}\boldsymbol{\tau}_{p}\right).$$

In the case where y_t is stationary we have $E\left(\mathbf{x}_{T}-\mu_{1}\boldsymbol{\tau}_{p}\right)=0$, and

$$E\left(y_{T+1}-\tilde{y}_{T+1}\right)=\left(\mu_{2}-\mu_{1}\right)\left(1-\beta_{2}^{*}\right),$$

which does not depend on the size of the break in the slope coefficients, $\boldsymbol{\beta}_2-\boldsymbol{\beta}_1$, and will be zero when $\mu_2=\mu_1$. This is an interesting theoretical result but its relevance is limited in practice where estimates of $(\mu_1, \boldsymbol{\beta}_1)$ based on past observations need to be used. One of the contributions of this paper might be viewed as identifying the circumstances under which the above result is likely to hold in the presence of estimation uncertainty.

In a related contribution Clements and Hendry (1999, pp. 39-42) consider the effect of estimation uncertainty on the forecast error decomposition using a first-order vector autoregressive model, and conclude estimation uncertainty to be relatively unimportant. However, their analysis assumes that the estimation is carried out immediately prior to the break, based on a correctly specified model which is not subject to any breaks. The assumption that parameters have been stable prior to forecasting is clearly restrictive, and it is therefore important that a more general framework is considered where the effect of estimation uncertainty can be analysed even in the presence of multiple breaks in the parameters (slope coefficients as well as error variances) over the estimation period. In this paper we provide such a framework in the case of AR(p) models subject to a single break point over the estimation period. But, it should become clear that the analysis readily extends to two or more break points.[2]

In particular, our interest in this paper lies in the point (or probability) forecast of y_{T+1} conditional on $\Omega_T=\{y_1, y_2, ..., y_T\}$ in the context of the break point specification (2). In the case where the post-break window size, $v_2=T-T_1$ is sufficiently large ($v_2\rightarrow\infty$), the structural break is relatively unimportant and the forecast of y_{T+1} can be based exclusively on the post-break observations. However, when v_2 is small it might be worthwhile to base the forecasting model on pre-break

[2]Explicitly allowing for breaks and parameter uncertainty prior to forecasting also raises the issue of the choice of observation window discussed in related papers in Pesaran and Timmermann (2002, 2003).

as well as post-break observations. The number of pre-break observations, which we denote by v_1, then becomes a choice parameter. In what follows we assume T_1 is known but consider forecasting y_{T+1} using the past $T-m+p+1$ observations, $m-p$ being the starting point of the estimation window,

$$\mathbf{y}_T(m-p) = (y_{m-p}, y_{m-p+1}, ..., y_{T_1}, y_{T_1+1} ..., y_T)', \tag{5}$$

with the p observations $y_{m-p}, y_{m-p+1}, ..., y_{m-1}$ treated as given initial values.[3] The length of the pre-break window is then given by $v_1 = T_1 - m + 1$, and the number of time periods used in estimation is therefore $v = v_1 + v_2 = T - m + 1$. To simplify the notations we shall consider values of $v_1 \geq p$, or $m \leq T_1 - p - 1$.

The point forecast of y_{T+1} conditional on $\mathbf{y}_T(m-p)$ is given by

$$\hat{y}_{T+1}(m) = \hat{\alpha}_T(m) + \mathbf{x}_T'\hat{\boldsymbol{\beta}}_T(m),$$

where $\mathbf{x}_T = (y_T, y_{T-1}, ..., y_{T-p+1})'$, $\hat{\boldsymbol{\beta}}_T(m) = \left(\hat{\beta}_{1T}(m), \hat{\beta}_{2T}(m), ..., \hat{\beta}_{pT}(m)\right)'$, $\boldsymbol{\tau}_v$ is a $v \times 1$ vector of ones, $\mathbf{M}_\tau = \mathbf{I}_\nu - \boldsymbol{\tau}_v(\boldsymbol{\tau}_v'\boldsymbol{\tau}_v)^{-1}\boldsymbol{\tau}_v'$, and

$$\mathbf{X}_T(m) = (\mathbf{y}_{T-1}(m-1), \mathbf{y}_{T-2}(m-2), ..., \mathbf{y}_{T-p}(m-p)),$$

so that

$$\hat{\boldsymbol{\beta}}_T(m) = \left[\mathbf{X}_T'(m)\mathbf{M}_\tau\mathbf{X}_T(m)\right]^{-1}\mathbf{X}_T'(m)\mathbf{M}_\tau\mathbf{y}_T(m), \tag{6}$$

$$\hat{\alpha}_T(m) = \frac{\boldsymbol{\tau}_v'\mathbf{y}_T(m) - \boldsymbol{\tau}_v'\mathbf{X}_T(m)\hat{\boldsymbol{\beta}}_T(m)}{v}, \tag{7}$$

The one-step ahead forecast error is

$$e_{T+1}(m) = y_{T+1} - \hat{y}_{T+1}(m) = \sigma_2\varepsilon_{T+1} - \xi_T(m), \tag{8}$$

where

$$\xi_T(m) = [\hat{\alpha}_T(m) - \alpha_2] + \mathbf{x}_T'\left(\hat{\boldsymbol{\beta}}_T(m) - \boldsymbol{\beta}_2\right). \tag{9}$$

$\boldsymbol{\beta}_2 = (\beta_{21}, \beta_{22}, ..., \beta_{2p})'$ and $\alpha_2 = \mu_2\left(1 - \boldsymbol{\tau}_p'\boldsymbol{\beta}_2\right)$. We consider both the unconditional and conditional mean squared forecast error given by $E_\varepsilon\left(e_{T+1}^2(m)\right)$ and $E_\varepsilon\left(e_{T+1}^2(m)\,|\Omega_T\right)$, respectively, where the expectations operator $E_\varepsilon(\cdot)$ is defined with respect to the distribution of the innovations ε_t. To the see how the MSFE

[3]Throughout the paper we shall use the notation $\mathbf{y}_T(k) = (y_k, ..., y_T)'$.

depends on the starting point of the estimation window, m, note that ε_{T+1} and $\xi_T(m)$ are independently distributed and

$$E_\varepsilon\left(e^2_{T+1}(m)\,|\Omega_T\right) = \sigma_2^2 + E_\varepsilon\left(\xi_T^2(m)\,|\Omega_T\right). \tag{10}$$

To carry out the necessary computations, an explicit expression for $\xi_T(m)$ in terms of the $\varepsilon_t' s$ is required. This is complicated and depends on the state of the process just before the first observation is used for estimation.

For a given choice of $m > p$ and a finite sample size T, the joint distribution of $\hat{\boldsymbol{\beta}}_T(m)$ and $\hat{\alpha}_T(m)$ depends on the distribution of the initial values

$$\mathbf{y}_{m-1}(m-p)=\left(y_{m-p}, y_{m-p+1}, ..., y_{m-1}\right)'. \tag{11}$$

We distinguish between the two important cases where the pre-break process is stationary and when it contains a unit root.

3.0.1. Pre-Break Process is Stationary

In the case where the pre-break regime is stationary and has been in operation for sufficiently long time, the distribution of $\mathbf{y}_{m-1}(m-p)$ does not depend on m and is given by

$$\mathbf{y}_{m-1}(m-p) \sim N(\mu_1 \boldsymbol{\tau}_p, \sigma_1^2 \mathbf{V}_p), \tag{12}$$

where $\mathbf{V}_p$ is defined in terms of the pre-break parameters. For example, for $p = 1$, $\mathbf{V}_1 = 1/(1-\beta_{11}^2)$, and for $p = 2$

$$\mathbf{V}_2 = \frac{1}{(1+\beta_{12})\left[(1-\beta_{12})^2 - \beta_{11}^2\right]} \begin{pmatrix} 1-\beta_{12} & \beta_{11} \\ \beta_{11} & 1-\beta_{12} \end{pmatrix}.$$

3.0.2. Pre-Break Process is I(1)

If the pre-break process contains a unit root, the covariance of $\mathbf{y}_{m-1}(m-p)$ is no longer given by $\sigma_1^2 \mathbf{V}_p$ and in general depends on m. Under a pre-break unit root, $\beta_1^* = 1$ and the pre-break process is given by

$$\Delta y_t = \sum_{j=1}^{p-1} \delta_{1j} \Delta y_{t-j} + \sigma_1 \varepsilon_t, \text{ for } t \leq T_1, \tag{13}$$

where $\delta_{1j} = -\sum_{\ell=j+1}^{p} \beta_{1\ell}$. The distribution of initial values can now be specified in terms of the stationary distribution of the first differences, $(\Delta y_2, \Delta y_3, ..., \Delta y_p)$, and an assumption concerning the first observation in the sample, y_1. In what follows we assume that y_1 is given by

$$y_1 = \mu_1 + \omega\varepsilon_1, \tag{14}$$

where ω will be treated as a free parameter, and $\varepsilon_1 \sim N(0,1)$. Using (13) and (14) it is now possible to derive the distribution of the initial values, $\mathbf{y}_{m-1}(m-p) = (y_{m-p}, y_{m-p+1}, ..., y_{m-1})'$, noting that

$$y_{m-i} = y_1 + \Delta y_2 + ... + \Delta y_{m-i}, \text{ for } i = 1, 2, ..., p.$$

In the AR(1) case we have

$$\Delta y_t = \sigma_1\varepsilon_t, \text{ for } t = 2, 3, ..., T_1,$$

and in conjunction with (14) we have

$$\begin{aligned} y_{m-1} &= y_1 + \Delta y_2 + ... + \Delta y_{m-1} \\ &= \mu_1 + \omega\varepsilon_1 + \sigma_1 \left(\varepsilon_2 + \varepsilon_3 + ... + \varepsilon_{m-1}\right), \end{aligned}$$

and hence $y_{m-1} \sim N\left(\mu_1, \mathbf{V}_{1,m}\right)$, where

$$\mathbf{V}_{1,m} = \omega^2 + (m-2)\sigma_1^2. \tag{15}$$

For the AR(2) specification we have $\mathbf{y}_{m-1}(m-2) = (y_{m-2}, y_{m-1})' \sim N\left(\mu_1\boldsymbol{\tau}_2, \mathbf{V}_{2,m}\right)$, where $\mathbf{V}_{2,m}$ is derived in Appendix A.

3.1. *OLS Estimates*

Using (12) and (2) for $t = m, m+1, ..., T$, in matrix notations we have

$$\mathbf{B}\ \mathbf{y}_T(m-p) = \mathbf{d} + \mathbf{D}\ \boldsymbol{\varepsilon}, \tag{16}$$

where

$$\mathbf{D} = \sigma_1 \begin{pmatrix} \boldsymbol{\psi}_p & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\nu_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (\sigma_2/\sigma_1)\,\mathbf{I}_{\nu_2} \end{pmatrix}, \ \mathbf{d} = \mu_1 \begin{pmatrix} \boldsymbol{\tau}_p \\ (1-\beta_1^*)\boldsymbol{\tau}_{v_1} \\ (\mu_2/\mu_1)\,(1-\beta_2^*)\boldsymbol{\tau}_{v_2} \end{pmatrix}, \tag{17}$$

$$\mathbf{B} = \begin{pmatrix} \mathbf{I}_p & \mathbf{0} & \mathbf{0} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{32} & \mathbf{B}_{33} \end{pmatrix}. \tag{18}$$

The sub-matrices, $\mathbf{B}_{ij}$, depend only on the slope coefficients, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ and are defined in Appendix B. $\mathbf{I}_{\nu_1}$ and $\mathbf{I}_{\nu_2}$ are identity matrices of order ν_1 and ν_2, respectively and $\boldsymbol{\varepsilon} = (\varepsilon_{m-p}, \varepsilon_{m-p+1}, ..., \varepsilon_T)' \sim N(\mathbf{0}, \mathbf{I}_{\nu+p})$.

The form of $\boldsymbol{\psi}_p$ depends on whether the pre-break process is stationary or contains a unit root. Under the former $\boldsymbol{\psi}_p$ is a lower triangular Cholesky factor of $\mathbf{V}_p$, namely $\mathbf{V}_p = \boldsymbol{\psi}_p\boldsymbol{\psi}_p'$, where $\mathbf{V}_p$ is the covariance matrix of $\mathbf{y}_{m-1}(m-p)$. Appropriate expressions for $\mathbf{V}_p$ in the case of $p = 1$ and 2 are already provided in Section 3.0.1. When the pre-break process has a unit root, $\boldsymbol{\psi}_p$ is given by the lower triangular Cholesky factor of $\mathbf{V}_{p,m}$, which is given by (15) above for $p = 1$ and in Appendix A by (38) for $p = 2$.

Using (40) derived in Appendix B, in general we have

$$\mathbf{y}_{T-i}(m-i) = \mathbf{G}_i(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon}), \text{ for } i = 0, 1, ..., p, \tag{19}$$

where $\mathbf{G}_i$ are $v \times (v+p)$ selection matrices defined by $\mathbf{G}_i = (\mathbf{0}_{v\times p-i} \vdots \mathbf{I}_\nu \vdots \mathbf{0}_{v\times i})$, $\mathbf{H} = \mathbf{B}^{-1}\mathbf{D}$, and $\mathbf{c} = \mathbf{B}^{-1}\mathbf{d}$. In particular,

$$\mathbf{y}_T(m) = \mathbf{G}_0(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon}),$$

and

$$\mathbf{X}_T(m) = \left[\mathbf{G}_1(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon}), \mathbf{G}_2(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon}), ..., \mathbf{G}_p(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon})\right].$$

Therefore, in general the (i, j) element of the product moment matrix, $\mathbf{X}_T'(m)\mathbf{M}_\tau\mathbf{X}_T(m)$, is given by $(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon})'\mathbf{G}_i'\mathbf{M}_\tau\mathbf{G}_j(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon})$, for $i, j = 1, 2, ..., p$, and the j^{th} element of $\mathbf{X}_T'(m)\mathbf{M}_\tau\mathbf{y}_T(\mathbf{m})$ is given by $(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon})'\mathbf{G}_j'\mathbf{M}_\tau\mathbf{G}_0(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon})$, for $j = 1, 2, ..., p$. Hence, $\hat{\boldsymbol{\beta}}_T(m) = \left(\hat{\beta}_{1T}(m), \hat{\beta}_{2T}(m), ..., \hat{\beta}_{pT}(m)\right)'$, is a non-linear function of the quadratic forms $(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon})'\mathbf{G}_i'\mathbf{M}_\tau\mathbf{G}_j(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon})$, for $i = 1, 2, ...p$, and $j = 0, 1, ..., p$, with known matrices $\mathbf{H}$, $\mathbf{G}_i$, $\mathbf{c}$, and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}_{\nu+p})$. Similarly, using (7) we have

$$\hat{\alpha}_T(m) = v^{-1}\boldsymbol{\tau}_v'\mathbf{G}_0(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon}) - v^{-1}\boldsymbol{\tau}_v'\sum_{i=1}^{p}\mathbf{G}_i(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon})\hat{\beta}_{iT}(m). \tag{20}$$

In the AR(1) case these results simplify to

$$\hat{\beta}_T(m) = \frac{(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon})'\mathbf{G}_1'\mathbf{M}_\tau\mathbf{G}_0(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon})}{(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon})'\mathbf{G}_1'\mathbf{M}_\tau\mathbf{G}_1(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon})}, \tag{21}$$

and

$$\hat{\alpha}_T(m) = v^{-1}\boldsymbol{\tau}_v'\mathbf{G}_0(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon}) - v^{-1}\boldsymbol{\tau}_\mathbf{v}'\mathbf{G}_1(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon})\hat{\beta}_T(m). \tag{22}$$

Using the above results in (6) it is now easily seen that in general $\hat{\boldsymbol{\beta}}_T(m)$ depends on the ratios, μ_1/σ_1, σ_1/σ_2 and μ_1/μ_2 (or μ_2/μ_1), whilst $\hat{\alpha}_T(m)$ depends on all the four coefficients, μ_1, μ_2, σ_1, and σ_2, individually. Two cases of special interest arise when there is no mean shift in the model, and when the post-break process contains a unit root. In both cases, as shown in Appendix B, $\mathbf{G}_i\mathbf{c} = \kappa\boldsymbol{\tau}_v$ where $\kappa = \mu$ when there is no mean shift (i.e. $\mu_1 = \mu_2 = \mu$), and $\kappa = \mu_1$ if there is a mean shift but $\beta_2^* = 1$. Under either of these two special cases we have $\mathbf{M}_\tau\mathbf{G}_i\mathbf{c} = \mathbf{0}$, for all i, and $\hat{\boldsymbol{\beta}}_T(m)$ will be a function of the quadratic terms, $\boldsymbol{\varepsilon}'\mathbf{H}'\mathbf{G}_i'\mathbf{M}_\tau\mathbf{G}_j\mathbf{H}\boldsymbol{\varepsilon}$, which depend only on the ratio of the error variances, σ_1/σ_2. These results also establish the following proposition:

Proposition 1 *Under $\mu_1 = \mu_2$ or if $\beta_1^* < 1$ and $\beta_2^* = 1$, $\hat{\boldsymbol{\beta}}_T(m)$ defined by (6) does not depend on the scale of the error variances (σ_1^2, σ_2^2) or the unconditional means, μ_1, μ_2, and is an even function of $\boldsymbol{\varepsilon}$.*

This proposition plays a key role in the analysis of prediction errors below. It is also worth noting that $\hat{\boldsymbol{\beta}}_T(m)$ will continue to be an even function of the errors in the more general case where the slope coefficients and/or the error variances are subject to multiple breaks, so long as the mean of the process remains unchanged. This proposition does not, however, extend to the OLS estimate of the intercept, $\hat{\alpha}_T(m)$. To see this, using (20) and noting that under $\mu_1 = \mu_2$, or if $\beta_2^* = 1$, $\mathbf{G}_i\mathbf{c} = \mu_1\boldsymbol{\tau}_v$ we have

$$\hat{\alpha}_T(m) = \mu_1\left(1 - \hat{\beta}_T^*(m)\right) + \left(\frac{\boldsymbol{\tau}_v'\mathbf{G}_0\mathbf{H}\boldsymbol{\varepsilon}}{v}\right) - \sum_{i=1}^{p}\left(\frac{\boldsymbol{\tau}_v'\mathbf{G}_i\mathbf{H}\boldsymbol{\varepsilon}}{v}\right)\hat{\beta}_{iT}(m), \qquad (23)$$

where $\hat{\beta}_T^*(m) = \sum_{i=1}^p \hat{\beta}_{iT}(m) = \boldsymbol{\tau}_p'\hat{\boldsymbol{\beta}}_T(m)$. It is clear that in this case $\hat{\alpha}_T(m)$ is an odd function of $\boldsymbol{\varepsilon}$, and depends on σ_1, σ_2 and μ_1 individually.

3.2. *Forecast Error Decomposition*

Using (20) and (9) in (8), and recalling that $\alpha_2 = \mu_2\left(1 - \boldsymbol{\tau}_p'\boldsymbol{\beta}_2\right)$, then after some algebra the forecast error, $e_{T+1}(m)$, can be decomposed as

$$e_{T+1}(m) = \sigma_2\varepsilon_{T+1} - X_{1T}(m) - X_{2T}(m) - X_{3T}(m), \qquad (24)$$

where

$$X_{1T}(m) = \left(\frac{\boldsymbol{\tau}_v'\mathbf{G}_0\mathbf{c}}{v} - \mu_2\right) - \sum_{i=1}^{p}\left(\frac{\boldsymbol{\tau}_v'\mathbf{G}_i\mathbf{c}}{v} - \mu_2\right)\hat{\beta}_{iT}(m), \qquad (25)$$

$$X_{2T}(m) = \frac{\boldsymbol{\tau}_v'\mathbf{G}_0\mathbf{H}\boldsymbol{\varepsilon}}{v} - \sum_{i=1}^{p}\left(\frac{\boldsymbol{\tau}_v'\mathbf{G}_i\mathbf{H}\boldsymbol{\varepsilon}}{v}\right)\hat{\beta}_{iT}(m), \tag{26}$$

and

$$X_{3T}(m) = (\mathbf{x}_T - \mu_2\boldsymbol{\tau}_p)'\left(\hat{\boldsymbol{\beta}}_T(m) - \boldsymbol{\beta}_2\right). \tag{27}$$

The first term in this decomposition refers to future uncertainty which is independently distributed of the other terms. The second term, $X_{1T}(m)$, is due to the mean shift and disappears under $\mu_1 = \mu_2 = \mu$. Recall that in this case $v^{-1}\boldsymbol{\tau}_v'\mathbf{G}_i\mathbf{c} = \mu$, for all i.[4] The third term, $X_{2T}(m)$, captures the uncertainty associated with the unconditional mean of the process and reduces to zero if $\mu_1 = \mu_2 = 0$. The last term represents the slope uncertainty and depends on whether the analysis is carried out unconditionally, or conditionally on $\mathbf{x}_T = (y_T, y_{T-1}, ..., y_{T-p+1})'$, in which case the extent of the bias will generally depend on the size of the gap $\mathbf{x}_T - \mu_2\boldsymbol{\tau}_p$.

3.3. *Unconditional MSFE*

To obtain the unconditional form of $e_{T+1}(m)$, we first note that $\mathbf{x}_T$ can be written as $\mathbf{S}_p\mathbf{y}_T(m)$, where $\mathbf{S}_p = (\mathbf{0}_{p\times(v-p)} \vdots \mathbf{J}_p)$, and $\mathbf{J}_p$ is the $p \times p$ matrix

$$\mathbf{J}_p = \begin{pmatrix} 0 & 0 & \cdots & & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 1 & \cdots & 0 & 0 \\ 1 & & \cdots & 0 & 0 \end{pmatrix}.$$

Therefore, using (19) we have

$$\mathbf{x}_T - \mu_2\boldsymbol{\tau}_p = (\mathbf{S}_p\mathbf{G}_0\mathbf{c} - \mu_2\boldsymbol{\tau}_{\mathbf{p}}) + \mathbf{S}_p\mathbf{G}_0\mathbf{H}\boldsymbol{\varepsilon},$$

and $X_{3T}(m)$, defined by (27), decomposes further as

$$X_{3T}(m) = (\mathbf{S}_p\mathbf{G}_0\mathbf{c} - \mu_2\boldsymbol{\tau}_{\mathbf{p}})'\left(\hat{\boldsymbol{\beta}}_T(m) - \boldsymbol{\beta}_2\right) + (\mathbf{S}_p\mathbf{G}_0\mathbf{H}\boldsymbol{\varepsilon})'\left(\hat{\boldsymbol{\beta}}_T(m) - \boldsymbol{\beta}_2\right).$$

[4] See the last section of Appendix B. Note also that $X_{1T}(m)$ does not disappear if $\beta_2^* = \boldsymbol{\tau}_p'\boldsymbol{\beta}_2 = 1$, so long as $\mu_1 \neq \mu_2$. However, under $\beta_2^* = 1$, it simplifies to

$$X_{1T}(m) = (\mu_1 - \mu_2)\left(1 - \boldsymbol{\tau}_p'\hat{\boldsymbol{\beta}}_T(m)\right).$$

However, under $\mu_1 = \mu_2 = \mu$ the first term of $X_{3T}(m)$ vanishes and we have[5]

$$e_{T+1}(m) = \sigma_2 \varepsilon_{T+1} - X_{2T}(m) - \left(\hat{\boldsymbol{\beta}}_T(m) - \boldsymbol{\beta}_2\right)' \mathbf{S}_p \mathbf{G}_0 \mathbf{H} \boldsymbol{\varepsilon}. \tag{28}$$

Also under $\mu_1 = \mu_2 = \mu$, $e_{T+1}(m)$, and hence $E_\varepsilon\left[e^2_{T+1}(m)\right]$, do not depend on the unconditional mean of the autoregressive process.

The computation of $E_\varepsilon\left[e^2_{T+1}(m)\right]$ can be carried out via stochastic simulations. We have

$$\hat{E}_R\left[e^2_{T+1}(m)\right] = \sigma_2^2 + \frac{1}{R}\sum_{r=1}^{R}\left[X^{(r)}_{1T}(m) + X^{(r)}_{2T}(m) + X^{(r)}_{3T}(m)\right]^2,$$

where the terms $X^{(r)}_{iT}(m), i = 1, 2, 3$ can be computed using random draws from $\boldsymbol{\varepsilon} \sim N(0, \mathbf{I}_{\nu+p})$, which we denote by $\boldsymbol{\varepsilon}^{(r)}$, $r = 1, 2, ..., R$. In particular,

$$X^{(r)}_{1T}(m) = \left(\frac{\boldsymbol{\tau}'_v \mathbf{G}_0 \mathbf{c}}{v} - \mu_2\right) - \sum_{i=1}^{p}\left(\frac{\boldsymbol{\tau}'_v \mathbf{G}_i \mathbf{c}}{v} - \mu_2\right)\hat{\beta}^{(r)}_{iT}(m), \tag{29}$$

$$X^{(r)}_{2T}(m) = \frac{\boldsymbol{\tau}'_v \mathbf{G}_0 \mathbf{H} \boldsymbol{\varepsilon}^{(r)}}{v} - \sum_{i=1}^{p}\left(\frac{\boldsymbol{\tau}'_v \mathbf{G}_i \mathbf{H} \boldsymbol{\varepsilon}^{(r)}}{v}\right)\hat{\beta}^{(r)}_{iT}(m), \tag{30}$$

$$X^{(r)}_{3T}(m) = \left(\mathbf{S}_p \mathbf{G}_0 \mathbf{c} - \mu_2 \boldsymbol{\tau}_\mathbf{p}\right)'\left(\hat{\boldsymbol{\beta}}^{(r)}_T(m) - \boldsymbol{\beta}_2\right) + \left(\mathbf{S}_p \mathbf{G}_0 \mathbf{H} \boldsymbol{\varepsilon}^{(r)}\right)'\left(\hat{\boldsymbol{\beta}}^{(r)}_T(m) - \boldsymbol{\beta}_2\right), \tag{31}$$

and $\hat{\beta}^{(r)}_{iT}(m)$ denotes the estimate of β_i based on $\boldsymbol{\varepsilon}^{(r)}$. Assuming $E_\varepsilon\left[e^2_{T+1}(m)\right]$ exists, then due to the independence of $\boldsymbol{\varepsilon}^{(r)}$ across r, and the fact that $X^{(r)}_{iT}(m)$ are also independently and identically distributed across r, we have (as $R \to \infty$)

$$\hat{E}_R\left[e^2_{T+1}(m)\right] \xrightarrow{p} E_\varepsilon\left[e^2_{T+1}(m)\right].$$

The following proposition generalizes Theorem 8.5.2 in Fuller (1996, page 445) to the case where estimation has been based on an AR(p) model which has been subject to breaks in the slope coefficients and/or error variances.

Proposition 2: The one-step ahead forecast errors, $e_{T+1}(m)$, defined by (8) from the AR(p) model, (2), subject to a break in the AR coefficients ($\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$) or a break in the innovation variance ($\sigma_1^2 \neq \sigma_2^2$) are unbiased provided that:

(i) The probability distribution of $\boldsymbol{\varepsilon}^* = (\boldsymbol{\varepsilon}', \varepsilon_{T+1})'$ is symmetrically distributed around $E(\boldsymbol{\varepsilon}^*) = 0$, and its first and second order moments exist;

[5]Note that in this case $\mathbf{S}_p \mathbf{G}_0 \mathbf{c} = \mu \mathbf{S}_p \boldsymbol{\tau}_v = \mu \boldsymbol{\tau}_p$.

(ii) The first-order moments of the estimated slope coefficients, $\hat{\beta}_{iT}(m)$, exist, namely $E\left|\hat{\beta}_{iT}(m)\right| < \infty$, for $i = 1, 2, ..., p$;

(iii) There is no break in the mean of the process, $\mu_1 = \mu_2$.

Proof: Under $\mu_1 = \mu_2$, using (26) and (28), the prediction error can be written as

$$\begin{aligned} e_{T+1}(m) &= \sigma_2\varepsilon_{T+1} - \left(\hat{\boldsymbol{\beta}}_T(m) - \boldsymbol{\beta}_2\right)' \mathbf{S}_p\mathbf{G}_0\mathbf{H}\boldsymbol{\varepsilon} \\ &\quad - \left[\frac{\boldsymbol{\tau}_v'\mathbf{G}_0\mathbf{H}\boldsymbol{\varepsilon}}{v} - \sum_{i=1}^{p}\left(\frac{\boldsymbol{\tau}_v'\mathbf{G}_i\mathbf{H}\boldsymbol{\varepsilon}}{v}\right)\hat{\beta}_{iT}(m)\right]. \end{aligned}$$

It is clear that under assumption (i) the terms $\sigma_2\varepsilon_{T+1}$, $\boldsymbol{\beta}_2'\mathbf{S}_p\mathbf{G}_0\mathbf{H}\boldsymbol{\varepsilon}$, and $\boldsymbol{\tau}_v'\mathbf{G}_0\mathbf{H}\boldsymbol{\varepsilon}$, which are linear functions of $\boldsymbol{\varepsilon}^*$, have mean zero and we have

$$E_\varepsilon\left[e_{T+1}(m)\right] = -E_\varepsilon\left[\hat{\boldsymbol{\beta}}_T'(m)\mathbf{S}_p\mathbf{G}_0\mathbf{H}\boldsymbol{\varepsilon}\right] + \sum_{i=1}^{p} E_\varepsilon\left[\left(\frac{\boldsymbol{\tau}_v'\mathbf{G}_i\mathbf{H}\boldsymbol{\varepsilon}}{v}\right)\hat{\beta}_{iT}(m)\right].$$

Also under $\mu_1 = \mu_2$ and by Proposition 1, $\hat{\boldsymbol{\beta}}_T(m)$, is an even function of $\boldsymbol{\varepsilon}$. Hence, $\hat{\boldsymbol{\beta}}_T'(m)\mathbf{S}_p\mathbf{G}_0\mathbf{H}\boldsymbol{\varepsilon}$, and $(\boldsymbol{\tau}_v'\mathbf{G}_i\mathbf{H}\boldsymbol{\varepsilon})\,\hat{\beta}_{iT}(m)$ for $i = 1, 2, ..., p$ are odd functions of $\boldsymbol{\varepsilon}$, and under assumptions (i) and (ii) their expectations exist and are equal to zero by the symmetry assumption. Therefore,

$$E_\varepsilon\left[e_{T+1}(m)\right] = 0.$$

In the case where $\mu_1 \neq \mu_2$, $\hat{\beta}_{jT}(m)$ is not an even function of $\boldsymbol{\varepsilon}$, the term X_{1T} defined by (25) does not vanish and the prediction error given by (24), is no longer an odd function of $\boldsymbol{\varepsilon}$, so it will, in general, not have a zero mean. ■

Remark: Conditions under which moments of $\hat{\beta}_{iT}(m)$ exists in the case of AR(1) models with fixed coefficients have been investigated in the literature and readily extends to AR(1) models subject to breaks. For the AR(1) model under $\mu_1 = \mu_2$ we have [see (21)]

$$\hat{\beta}_T(m) = \frac{\boldsymbol{\varepsilon}'\mathbf{H}'\mathbf{G}_1'\mathbf{M}_\tau\mathbf{G}_0\mathbf{H}\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}'\mathbf{H}'\mathbf{G}_1'\mathbf{M}_\tau\mathbf{G}_1\mathbf{H}\boldsymbol{\varepsilon}}.$$

Assuming that $\boldsymbol{\varepsilon}$ is normally distributed and applying a Lemma due to Smith (1988) to $(\boldsymbol{\varepsilon}'\mathbf{H}'\mathbf{G}_1'\mathbf{M}_\tau\mathbf{G}_1\mathbf{H}\boldsymbol{\varepsilon})^{-1}$, it is easily established that the r^{th} moment of

$\hat{\beta}_T(m)$ exists if $Rank\left(\mathbf{H}'\mathbf{G}_1'\mathbf{M}_\tau\mathbf{G}_1\mathbf{H}\right) = v - 1 = T - m > 2r$.[6] Hence, $\hat{\beta}_T(m)$ has a first-order moment if $T > m + 2$. To our knowledge no such conditions are known for higher order AR processes, even with fixed coefficients.

Proposition 2 has important implications for the trade-off that exists in the estimation bias of the slope and intercept coefficients in the AR models even in the presence of breaks so long as $\mu_1 = \mu_2 = \mu$. To see this notice from (22) that

$$E\left[\hat{\alpha}_T(m) - \alpha_2\right] = -\mu\ E\left[\hat{\beta}^*_T(m) - \beta^*_2\right].$$

This provides an interesting relationship between the small sample bias of the estimator of the intercept term, $E\left[\hat{\alpha}_T(m) - \alpha_2\right]$, and the small sample bias of the long-run coefficient, $E\left[\hat{\beta}^*_T(m) - \beta^*_2\right]$. The estimator of the intercept term, $\hat{\alpha}_T(m)$, is unbiased only if the sample mean is zero. But, in general there is a spill-over effect from the bias of the slope coefficient to that of the intercept term.

For the AR(1) model the results simplify further and we have

$$E\left[\hat{\alpha}_T(m) - \alpha_2\right] = -\mu\ E\left[\hat{\beta}_T(m) - \beta_2\right]. \tag{32}$$

Since $E\left[\hat{\beta}_T(m) - \beta_2\right] < 0$, it therefore follows that

$$\begin{aligned} E\left[\hat{\alpha}_T(m) - \alpha_2\right] &> 0 \text{ if } \mu > 0, \\ E\left[\hat{\alpha}_T(m) - \alpha_2\right] &\leq 0 \text{ if } \mu \leq 0. \end{aligned}$$

Once again these results hold irrespective of whether $\beta_1 = \beta_2$ or not.

3.4. *Conditional MSFE*

As before we have

$$e_{T+1}(m) = \sigma_2\varepsilon_{T+1} - X_{1T}(m) - X_{2T}(m) - X_{3T}(m),$$

where $X_{iT}(m)$, $i = 1, 2, 3$, are defined by (25), (26), and (27). In computing the conditional MSFE, defined by $E_\varepsilon\left(e^2_{T+1}(m)\,|\Omega_T\right)$, we fix $\mathbf{x}_T$ and integrate with respect to the distribution of $\boldsymbol{\varepsilon}$. Recall that $\hat{\boldsymbol{\beta}}_T(m)$ and $\hat{\alpha}_T(m)$ as defined in (6) and (7) are only functions of $\boldsymbol{\varepsilon}$ and are hence not constrained by the terminal value,

[6] Note that $\mathbf{H}$ is full rank, $rank(G_i) = v$, and $rank(\mathbf{M}_\tau) = v - 1$.

x_T.[7] To investigate the effect of parameter estimation uncertainty we therefore draw values of ε independently of $\mathbf{x}_T$.

Once again the results simplify when $\mu_1 = \mu_2 = \mu$. In this case $X_{1T}(m) = 0$, $X_{2T}(m)$ is an odd function of $\boldsymbol{\varepsilon}$, and assuming that the distribution of $\boldsymbol{\varepsilon}$ is symmetric we have

$$E_\varepsilon[e_{T+1}(m)|\Omega_T] = -\left(\mathbf{x}_T - \mu\boldsymbol{\tau}_p\right)' E_\varepsilon\left(\hat{\boldsymbol{\beta}}_T(m) - \boldsymbol{\beta}_2\right).$$

Suppose $p = 1$, so that it is easy to characterize when $\mathbf{x}_T$ is above or below the mean. Then

$$E_\varepsilon\left[e_{T+1}(m) \mid \Omega_T\right] = -\left(y_T - \mu\right) E_\varepsilon\left(\hat{\beta}_T(m) - \beta_2\right). \tag{33}$$

Since, $E_\varepsilon\left(\hat{\beta}_T(m) - \beta_2\right) < 0$,

$$E_\varepsilon[e_{T+1}(m)|\Omega_T] = \begin{cases} > 0 & if\ y_T > \mu \\ \leq 0 & if\ y_T \leq \mu \end{cases}, \tag{34}$$

and the estimated model under-predicts if the last observation is above the unconditional mean ($y_T > \mu$), while conversely it over-predicts if the last observation is below the unconditional mean ($y_T < \mu$). Therefore, conditional predictions tend to be biased towards the unconditional mean of the process.

As with the unconditional MSFE, the computation of the conditional MSFE can also be carried out by stochastic simulations. In general, for a given value of $\mathbf{x}_T$, and using draws from $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}_{\nu+p})$ we have

$$\hat{E}_R\left(e_{T+1}^2(m)\,|\Omega_T\right) = \sigma_2^2 + \frac{1}{R}\sum_{r=1}^{R}\left[X_{1T}^{(r)}(m) + X_{2T}^{(r)}(m) + \tilde{X}_{3T}^{(r)}(m)\right]^2, \tag{35}$$

where $X_{1T}^{(r)}(m)$ and $X_{2T}^{(r)}(m)$ are given by (29) and (30), as before, with the third term, $\tilde{X}_{3T}^{(r)}(m)$, now defined by

$$\tilde{X}_{3T}^{(r)}(m) = \left(\mathbf{x}_T - \mu_2\boldsymbol{\tau}_p\right)'\left(\hat{\boldsymbol{\beta}}_T^{(r)}(m) - \boldsymbol{\beta}_2\right). \tag{36}$$

Once again as $R \to \infty$, we would expect $\hat{E}_R\left(e_{T+1}^2(m)\,|\Omega_T\right) \xrightarrow{p} E_\varepsilon\left(e_{T+1}^2(m)\,|\Omega_T\right)$.

[7] This is consistent with the approach taken in calculating asymptotic results, c.f. Fuller (1996). If we literally condition on the full path of y-values in Ω_T, then $\hat{\beta}_T(m)$ and $\hat{\alpha}_T(m)$ are of course non-random (fixed) constants and no estimation uncertainty arises.

4. **Numerical Results**

Our approach is quite general and allows us to study the small sample properties of AR models in some detail. The existing literature has focused on the AR(1) model without a break, where the key parameters affecting the properties of the OLS estimators, $\hat{\alpha}_T(m)$ and $\hat{\beta}_T(m)$, are the sample size and the persistence parameter, β_1. In our setting there are many more parameters to consider. In the absence of a break there are now p autoregressive parameters plus the intercept, α, and the innovation variance, σ^2. Under a single break, we need to consider both the pre- and post-break parameters - i.e. the AR coefficients $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, the intercepts (α_1, α_2) and the innovation variances (σ_1^2, σ_2^2). Furthermore, how the total sample divides into pre- and post-break periods (v_1 and v_2) is now crucial to the bias in the post-break parameter estimates and to the bias and variance of the forecast error.

To ensure that our results are comparable to the existing literature, our benchmark model is the AR(1) specification without a break (experiment 1 in Table 1). We study breaks in the autoregressive parameter in the form of both moderately sized (0.3) and large (0.6) breaks in either direction (experiments 2-4) as well as a unit root process in the post-break (experiment 5) or pre-break (experiment 9) period. We also consider pure breaks in the innovation variance (experiments 6 and 7), where σ changes between values of 1/4 and 1 or 4 and 1, and in the mean (experiment 8), where μ changes between 1 and 2. For convenience the parameter values assumed in each of the experiments are summarized in Table 1. Since our focus is on the effect of breaks on the bias and forecasting performance of AR models, results are presented as a function of the pre-break window size (v_1) and the post-break window size (v_2). We vary v_1 from zero (no pre-break information) through 1, 2, 3, 4, 5, 10, 20, 30, 50 and 100, while the post-break window, v_2, is set at 10, 20, 30, 50 and 100.

Simulation results are presented in Tables 2-5. Results are based on 50,000 Monte Carlo simulations with innovations drawn from an IID Gaussian distribution.[8] Table 2 shows the bias in $\hat{\beta}_1$ while Table 3 shows the conditional bias in the forecast for a situation where y_T is above its mean, i.e., $y_T = \mu_2 + \sigma_2$.[9] To measure

[8]We also considered an AR(2) specification to study the effect of higher order dynamics. Results were very similar to those reported below and are available from the authors' web site.

[9]Estimated values are computed as averages across Monte Carlo simulations relative to the

forecasting performance, Table 4 reports the unconditional RMSFE while Table 5 shows the RMSFE conditional on $y_T = \mu_2 + \sigma_2$, as functions of the pre-break (v_1) and post-break (v_2) window sizes. We condition on this particular value since if $y_T = \mu_2$ the conditional bias is zero while if $y_T = \mu_2 - \sigma_2$ the conditional bias takes the same value but with the sign reversed, c.f. (33).

4.1. *Bias Results*

First consider the bias in $\hat{\beta}_1$. In the absence of a break, $\hat{\beta}_1$ is downward biased with a bias that disappears as v_1 and v_2 increase and becomes quite small when the combined sample $v = v_1 + v_2$ is sufficiently large.[10] Notice the symmetry of the results in v_1 and v_2 which follows since (under no break) only $v_1 + v_2$ matters for the bias.[11] Once a break is introduced in the AR parameter, the bias in $\hat{\beta}_1$ continues to decline in v_2 but need no longer decline monotonically as a function of v_1. The reason for this is simple: including pre-break data generated by a different (less persistent) process introduces a new bias term in $\hat{\beta}_1$. It is only to the extent that this term is offset by a reduction in the small sample bias of the AR estimate that inclusion of pre-break data will lead to a bias reduction. Thus, when v_2 is very large (e.g., 50 or 100 post-break observations) the small sample bias in $\hat{\beta}_1$ based purely on post-break observations is already quite small. In this situation, inclusion of pre-break data will not lower the bias in $\hat{\beta}_1$. Conversely, when the post-break sample is small (i.e., $v_2 = 10 - 20$ observations), the small sample bias in $\hat{\beta}_1$ is very large and including up to 30 pre-break observations will actually reduce the bias under a moderately sized break. Naturally, if the break size is large (experiment 4), this effect is reduced since the true bias due to including pre-break observations in the estimation window dominates any reduction in the small sample bias in $\hat{\beta}_1$

---

true post-break values. To ensure comparability across the experiments they are based on the same random numbers.

[10]The bias estimates are in line with the well known Kendall (1954) approximation formula

$$E\left(\hat{\beta}_1\right) - \beta_1 = \frac{-(1+3\beta_1)}{v} + O(v^{-3/2}),\ v = v_1 + v_2.$$

[11]Recall from (32) that in the case of Gaussian errors the bias in $\hat{\alpha}_T(m)$ can be exactly inferred from the bias of $\hat{\beta}_T(m)$ when there is no break in the mean. For this reason we focus our analysis on the bias in $\hat{\beta}_T(m)$.

based solely on post-break data for all but the smallest post-break window sizes.

Interestingly, when the break is in the reverse direction (experiment 3) so that the true value of β_1 declines, including a small number of pre-break data points leads to a reduction in the bias in $\hat{\beta}_1$ even for very large post-break windows. For example, the bias in $\hat{\beta}_1$ is minimized by including 3 pre-break observations even when $v_2 = 100$. The reason is again related to the direction of the small sample bias in $\hat{\beta}_1$. Since $\hat{\beta}_1$ is downward biased, when the break is from high to low persistence, the (upward) bias introduced by inclusion of the more persistent pre-break data works in the opposite direction of the small sample (downward) bias in $\hat{\beta}_1$. For this reason the biases under a decline in β_1 tend to be smaller than the biases observed when β_1 increases at the time of the break.

Under a post-break unit root (experiment 5) the bias-minimizing pre-break window size is around 20 observations. Under a pre-break unit-root (experiment 9), bias is smallest for either $v_1 = 0$ or $v_1 = 1$. When a break occurs in the innovation variance (experiments 6 and 7), the smallest bias is always achieved by the longest pre-break windows. The only difference to the case without a break is that the bias is no longer a symmetric function of v_1 and v_2. Allowing for a break in the mean (experiment 8), the forecast error is no longer unbiased unconditionally and the optimal pre-break window size rises to 100 irrespective of the value of v_2.

Turning next to the conditional bias in the forecast, Table 3 shows that, in the absence of a break, the bias is positive when the prediction is made conditional on $y_T = \mu_2 + \sigma_2$, a value above the mean of the process. This is, of course, consistent with (34) and with the sign of the bias in $\hat{\beta}_1$. In general, the results for the conditional bias in the forecast error mirror those of the bias in $\hat{\beta}_1$, except for the case with a break in the mean. Whereas the bias in $\hat{\beta}_1$ was reduced the larger the value of v_1 when the mean increases at the time of the break, the bias in the forecast error is smallest when $v_1 = 0$ and the mean increases assuming a large post-break sample ($v_2 = 50$ or 100).

4.2. *Forecasting Performance*

To measure forecasting performance for the AR(1) model, unconditional and conditional RMSFE values are shown in Tables 4 and 5. Under no break the unconditional RMSFE is 1.15 for the smallest combined sample ($v_1 = 0, v_2 = 10$) and it declines symmetrically as a function of v_1 and v_2. In the presence of a moderate break in the AR coefficient, the unconditional RMSFE continues to decline as a function of v_2 but it no longer declines monotonically in v_1, the pre-break window. Furthermore, the unconditional RMSFE no longer converges to one - its theoretical value in the absence of parameter estimation uncertainty - provided the ratio v_1/v_2 does not go to zero. For example, when $v_1 = v_2 = 100$, the unconditional RMSFE under a moderate break in β_1 is close to 1.02 as opposed to a value of 1.006 observed in the case without a break. This difference is due to the squared bias in the AR parameters introduced by including pre-break data points. Generally, the windows that minimize the unconditional RMSFE tend to be longer than the windows that minimize the bias. Increasing the window size beyond the point that produces the smallest bias may be acceptable if it reduces the forecast error variance by more than the associated increase in the squared bias.

A moderately sized break in β_1 implies that the optimal pre-break window size declines to 10-20 observations under the unconditional RMSFE criterion although it remains much longer under the conditional RMSFE criterion. In both cases, the optimal value of v_1 is smaller, the larger the value of v_2 and the larger the size of the break in β_1 as can be seen by comparing the results from experiments 2 and 4.

Somewhat different patterns emerge when the AR model switches from having a unit root process to being stationary and vice versa. Under a post-break unit root the conditional RMSFE is minimized for rather large values of v_1, whereas the unconditional RMSFE is minimized at much smaller values of v_1, typically below 10 observations. But, under the pre-break unit root scenario, the smallest unconditional and conditional RMSFE values are produced by at most including one or two pre-break observations.

When the post-break innovation variance is higher, it is optimal to set the pre-break window as long as possible since this maximizes the length of the less noisy data and thus brings down the forecast error variance without introducing a bias in the forecast. In contrast, when the innovation variance declines at the time of the break, the optimal pre-break window size is only long provided the post-break window, v_2, is rather short and it declines to zero for larger values of v_2. Notice how the performance of the forecast can deteriorate badly upon the inclusion of a single pre-break data point even with quite long post-break windows. This is due to the extra noise introduced by using pre-break data for parameter estimation.

Under a break to the mean (experiment 8), the lowest conditional and unconditional RMSFE values are observed for the longer pre-break windows. This is an interesting finding and holds despite the fact that additional bias is introduced into the forecast. For example, in Table 4 the RMSFE is systematically reduced by increasing the pre-break window, v_1. In practice, breaks are likely to involve the means as well as the slope coefficients. In such situations our results suggest that, at least for breaks of similar size to those assumed here, it is difficult to outperform the forecasting performance generated by a model based on an expanding window of the data.

4.3. *Forecasting Performance of Rolling, Expanding and Post-break windows*

To shed light on the practical implications of our results, we next consider the out-of-sample forecasting performance of a range of widely used estimation windows. One way to deal with parameter instability is to use a rolling observation window. The size of the rolling window is often decided by *a priori* considerations. Here we consider a short rolling window using the most recent 25 observations and a relatively long rolling window based on the most recent 50 observations. If parameter instability is believed to be due to the presence of rare structural breaks, another possibility is to only use post-break data. In some cases the timing of the break may be known, but in most cases both the timing and the number of breaks must be estimated. We therefore use the Bai-Perron (1998) method to test for the presence of structural breaks and determine their timing, allowing up to three breaks and selecting the number of breaks by the Schwarz information criterion. If one or more breaks is identified at time t, this procedure uses data after the most recent break date to produce a forecast for period $t+1$. If no break is identified, an expanding data window is used to generate the forecast. Finally, as a third option an expanding window is considered. This is the most efficient estimation method in the absence of breaks and provides a natural benchmark.

We initially undertook the following simulation exercise. For each of the original AR(1) experiments we assume a break has taken place at observation 101. Our post-break forecast evaluation period runs from observations 111 to 150. For this period we computed RMSFEs of the one-step ahead forecasts obtained under different estimation windows by Monte Carlo simulation.

Panel A of Table 6 reports the results under a single break. As expected, when a break is not present the expanding window method produces the lowest RMSFE values. The expanding window also performs well when the break only affects the volatility or the mean parameter. The fact that the expanding window performs best even when the pre-break volatility is higher than the post-break volatility can be explained by the reduction in the variance of the parameter estimation error due to using a very long estimation window. The finding for a break in the mean is consistent with the simulation results in Table 4. In the experiments with a very large change in the autoregressive parameter (experiments 4-5), the short rolling window method produces the best performance, while the long rolling window works best for smaller breaks (experiments 2-3) which generate a lower squared bias.

Interestingly, the use of a post-break window with an estimated break point does not produce the lowest RMSFE performance in any of the experiments 1-8. A possible explanation of this finding lies in the modest power of break point tests to detect changes in autoregressive parameters as documented by Banerjee, Lumsdaine and Stock (1992). The only case where the post-break window method results in the lowest RMSFE is under a pre-break unit root (experiment 9). For this case the expanding window method performs quite poorly. This is consistent with our simulation results which showed that the conditional and unconditional RMSFE performance was best for very small - frequently zero - pre-break windows under a pre-break unit root. We also modified the simulation with the pre-break unit root to ensure that the point towards which the post-break process mean reverts is the terminal point of the pre-break unit root process (experiment 10) rather than simply μ_2. This is likely to generate sample paths more similar to those observed in practice, c.f. Banerjee, Lumsdaine and Stock (1992). The results show that although the expanding window method performs relatively better, it still does not produce the lowest RMSFE.

4.4. *Multiple Breaks*

So far we have focused on the case with a single structural break, but in practice the time series process under consideration may be subject to multiple breaks. Our procedure can readily be generalized to account for this possibility. Accordingly, we extended our simulation experiments to allow for two breaks occurring after 50 and 100 observations, respectively. The presence of multiple breaks raises questions concerning the process generating the breaks. Barring a general theory we consider two scenarios. The first scenario assumes that the two breaks lead to a shift in the regression coefficients in the same direction so for the AR(1) model we could have $\beta_{11} = 0.6$, $\beta_{12} = 0.75$ and $\beta_{13} = 0.9$. The second scenario assumes mean reversion in the parameters which first shift away from their initial values (at the first break date) and then revert to their original values after the second break date, so we could have $\beta_{11} = 0.6, \beta_{12} = 0.9$ and $\beta_{13} = 0.6$. Further details of the assumed parameter values are reported in the note to Table 6.

The results, provided in Panel B of Table 6, suggest that the expanding window method continues to produce the lowest RMSFEs in a number of cases including those with breaks to the volatility parameter, breaks in the mean and mean reversion in the autoregressive coefficient. Mean-reversion across breaks in the autoregressive coefficient tends to favor the expanding window method relative to the other methods since the earliest part of the sample will be similar to the final part from which the forecast is made. Adding the earliest part of the data sample prior to the first break therefore tends to pull the parameter estimate towards the value prevailing at the point of the forecast. In the few cases with multiple breaks where the expanding method does not dominate, the long rolling window method is generally best and it is frequently better to use a long rather than a short rolling window in the absence of a unit root.

5. **Empirical Analysis**

To better understand the practical implications of our theoretical analysis, we undertook a forecasting exercise using a range of macroeconomic time series. We considered forecasts of growth (log-first differences) in industrial production and real GDP, the inflation rate and short interest rates for six of the seven G7 countries, namely Canada, France, Germany, Japan, UK and the US. Italy was excluded due to incompleteness of data. All data is quarterly and covers the period 1959-1999. The data source is Stock and Watson (2003).

The forecasting exercise uses 25 initial observations for parameter estimation (or 40 observations in the case of the more heavily parameterized fourth order AR model) and considers $AR(1)$, $AR(2)$ and $AR(4)$ models. All forecasts are out-of-sample and use data up to period t to estimate the parameters of a forecasting model that is then used to generate a forecast for period $t+1$. The expanding window uses all available data up to time t. For the rolling windows we considered short and long rolling windows that use the most recent 25 and 40 observations, respectively, corresponding to roughly six and ten years of data. We also consider the two-step post-break window method described earlier, where in the first step we use the Bai-Perron procedure to identify the break point nearest to the forecast date and in the second step only use post-break data to estimate the parameters of the forecasting model.

Table 7 reports the outcome of the empirical analysis. Panel A reports RMSFE-values for the four estimation windows assuming an $AR(1)$ specification while panels B and C assume $AR(2)$ and $AR(4)$ models. For the $AR(1)$ models the post-break estimation method only produces the lowest RMSFE values in one case (Canadian inflation) out of the total of 23 cases. This happens despite the fact that breaks are identified at some point in the majority of the series, i.e. for 21 of 23 series for the AR(1) model. The short rolling window method does marginally better than the post-break method, generating the lowest RMSFE values in three of 23 cases, while the long rolling window does best in seven cases. However, by some margin, the best method turns out to be the expanding window which generates the lowest RMSFE values in 12 cases.[12]

The results are even stronger for the AR(2) and AR(4) models. For the AR(2) case the expanding window produces the best forecasting performance in 17 out of 23 cases with the long rolling window doing best in the remaining six cases. The short rolling window and the post-break window never outperform these methods. Similarly, for the AR(4) model, the expanding window generates the lowest out-of-sample RMSFE values in 18 cases while the long rolling window does so in the remaining five cases.

The theory developed in this paper and our simulation results are very useful in understanding why it is difficult to reduce the RMSFE values produced by the expanding window method. For example, in the case of the AR(1) model, we found evidence for many of the variables either that the persistence of the series declined after the most recent break or - in cases with multiple breaks - that there was mean reversion in persistence across breaks. This would explain why the long rolling window generally performs better than the short rolling window.

[12]In some cases the forecasting performance of the post-break and expanding window method is identical. This situation arises when no break point is detected so the full sample is used for estimation by the post-break method.

We also found that the 95% confidence interval for the time of the break frequently was quite wide and exceeded 10 observations. Imprecise determination of the time of the break can lead to a deterioration in the relative performance of the post-break forecasting method which will either be inefficient (if the estimated break date is later than the true date so that not all post-break data is used) or biased (if the estimated break date is premature so pre-break observations get included in the estimation window). In many cases the post-break window was also rather short, only averaging between one-half and one-third of the length of the expanding window, leading to imprecisely estimated values of the parameters of the forecasting model.

A final reason for the better overall performance of the expanding window estimation method over the other methods lies in the empirical importance of breaks to the innovation variance, σ^2. Experiments 6, 7, 16 and 17 in Table 6 showed that the expanding window method tends to be best under a volatility-only break. This finding is likely to carry over to cases with breaks in other parameters provided that the break in the volatility parameter is large relative to breaks in the other parameters. Indeed, we often observed very large variations in the estimates of σ across different break segments.

Overall, our results suggest that the squared bias arising from using pre-break data in estimation of the parameters of autoregressive forecasting models subject to breaks is less important to forecasting performance than the variance of the parameter estimation error. This would also explain the improved performance of the methods that use the longest estimation windows under the higher order AR models compared to the $AR(1)$ model since these models require estimation of more parameters.

Another interesting finding emerging from Table 7 is that, in general, variation in the out-of-sample RMSFE is greater across the various estimation windows than it is across the lag order chosen for the autoregressive model estimated through the expanding window method. A large amount of work has gone into designing methods for lag order selection. Our results suggest that the forecasting performance of autoregressive models subject to breaks could be even more affected by the length of the estimation window than by the autoregressive order and that a post-break estimation method - albeit appealing in theory - is difficult to implement successfully in practice for dynamic models. This points to the practical importance of gaining a better understanding of how to best determine the length of the data window used to estimate the parameters of the forecasting model.

6. Conclusion

This paper studied the small sample properties of forecasts from autoregressive models subject to breaks. It is insightful to compare and contrast our results for the AR(p) model to those reported by Pesaran and Timmermann (2003) under strictly exogenous regressors. Assuming strictly exogenous regressors, the OLS estimates based on post-break data are unbiased. Including pre-break data will therefore always increase the bias so that there will always be a trade-off between a larger squared bias and a smaller variance of the parameter estimates as more pre-break information is used. This trade-off can then be used to optimally determine the window size.

As we have shown in this paper, the situation can be very different for AR models due to the inherent small-sample bias in the estimates of the parameters of these models. In situations where the true AR coefficient(s) declines after a break, both the bias and the forecast error variance can in fact be reduced as a result of using pre-break data in the estimation. This is likely to be an important reason why, empirically, it is often difficult to improve forecasting performance over the expanding or long rolling window methods by only using post-break data. It also explains why forecasts based on a rolling window often perform worse than forecasts based on an expanding window of the data, particularly in cases where a short rolling window is used. These observations were confirmed empirically in an analysis of forecasts of GDP and industrial production growth, inflation and interest rates for six major economies.

More generally, we find both theoretically and empirically that there are many scenarios where the inclusion of some pre-break data for purposes of estimation of the parameters of autoregressive models leads to lower biases and lower mean squared forecast errors than if only post-break data is used. This can hold even when the post-break window is large, particularly when the post-break data generating process is highly persistent and/or has a break in the mean or variance. Our findings also indicate the possibility of a hybrid method that starts with an expanding window if the data set is relatively short and then switches to a long rolling window as the data set grows beyond a pre-specified threshold. We are currently investigating into the possible ways that such a threshold could be determined.

Several extensions to our results would be interesting to consider in future work. We have focused on the case with Gaussian innovations. Ullah (2003) observes that the bias in the forecast error is reasonably robust to skewness and kurtosis in the innovations of the AR model while, in contrast, the mean squared forecast error can be sensitive to higher order moments that arise in the non-Gaussian case. Our results could easily be extended to cover the non-normal case, for example by drawing innovations from a mixture of normals. Other possibilities are to consider the effect of additional predictors beyond autoregressive lags, multi-step ahead forecasts, and forecasts from vector autoregressive processes with or without cointegrating restrictions. Our theoretical and simulation results suggest that the empirical dominance of the expanding or long rolling window estimation method documented in this paper for univariate autoregressive models is likely to hold in these more complicated settings.

Bibliography

Abadir, K.M., 1993, OLS bias in a nonstationary autoregression. Econometric Theory 9, 81-93.

Alogoskoufis, G.S. and R. Smith, 1991, The Phillips Curve, the Persistence of Inflation, and the Lucas Critique: Evidence from Exchange Rate Regimes. American Economic Review 81, 1254-1275.

Andrews, D.W.K., 1993, Tests for Parameter Instability and Structural Change with Unknown Change Point. Econometrica 61, 821-856.

Andrews, D.W.K. and W. Ploberger, 1996, Optimal Changepoint Tests for Normal Linear Regression. Journal of Econometrics 70, 9-38.

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, pp. 1-22.

Banerjee, A., J.J. Dolado, D.F. Hendry and G.W. Smith, 1986, Exploring equilibrium relationships in economics through static models: some Monte Carlo evidence. Oxford Bulletin of Economics and Statistics 48, 253-277.

Banerjee, A., R. Lumsdaine and J.H. Stock, 1992, Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence. Journal of Business and Economic Statistics 10, 271-287.

Bartlett, M.S., 1946, On the theoretical specification and sampling properties of autocorrelated time series. Supplement to the Journal of the Royal Statistical Society 8, 27-41.

Box, G.E.P. and G.M. Jenkins, 1970, Time Series analysis: forecasting and control. San Francisco: Holden Day.

Chong, T. T-L, 2001, Structural Change in AR(1) Models. Econometric Theory 17(1), 87-155.

Chu, C-S J., M. Stinchcombe, and H. White, 1996, Monitoring Structural Change. Econometrica 64, 1045-1065.

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

Clements, M.P. and D.F. Hendry, 1999, Forecasting Non-stationary Economic Time Series. The MIT Press.

Copas, J.B., 1966, Monte Carlo results for estimation in a stable Markov time series. Journal of the Royal Statistical Society 129, 110-116.

Elliott, G. and U. Muller, 2002, Optimally Testing General Breaking Processes in Linear Time Series. Manuscript, UCSD and Princeton University.

Evans, G.B.A. and N.E. Savin, 1981, Testing unit roots: 1. Econometrica 49, 753-779.

Fuller, W.A, 1996, Introduction to Statistical Time Series. John Wiley & Sons, Inc.

Fuller, W.A. and D.P. Hasza, 1981, Properties of predictors for autoregressive time series. Journal of the American Statistical Association 76, 155-161.

Garcia, R. and P. Perron, 1996, An Analysis of the Real Interest Rate under Regime Shifts. Review of Economics and Statistics 78, 111-125.

Giacomini, R., 2002, Tests of Conditional Predictive Ability. Manuscript UCSD.

Grubb, D., Symons, J., 1987. Bias in regressions with a lagged dependent variable. Econometric Theory 3, 371-386.

Hansen, B.E., 1992, Tests for Parameter Instability in Regressions with I(1) Processes. Journal of Business and Economic Statistics 10, 321-335.

Inclan, C. and G.C. Tiao, 1994, Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance. Journal of the American Statistical Association 89, 913-923.

Kiviet, J.F., Phillips, G.D.A., 1993, Alternative bias approximations in regressions with a lagged dependent variable. Econometric Theory 9, 62-80.

Kiviet, J., and Phillips, G.D.A., 2003a, Improved Coefficient and Variance Estimation in Stable First-Order Dynamic Regression Models, UvA-Econometrics discussion paper 2002/02, version March 2003.

Kiviet, J., and Phillips, G.D.A., 2003b, Moment Approximation for Least Squares Estimators in Dynamic Regression Models with a Unit Root, version April 2003, UvA-Econometrics discussion paper 2003/03.

Pesaran, M.H. and A. Timmermann, 2002, Market timing and return prediction under model instability. Journal of Empirical Finance, 9, 495-510.

Pesaran, M.H. and A. Timmermann, 2003, Window selection with strictly exogenous regressors. Unpublished manuscript University of Cambridge and UCSD.

Phillips, P.C.B., 1977, Approximations to some finite sample distributions associated with a first-order stochastic difference equation. Econometrica 45, 463-485.

Phillips, P.C.B., 1987, Asymptotic expansions in nonstationary vector autoregression. Econometric Theory 3, 45-68.

Ploberger, W., W. Kramer, and K. Kontrus, 1989, A New Test for Structural Stability in the Linear Regression Model. Journal of Econometrics 40, 307-318.

Sawa, T., 1978, The exact moments of the least squares estimator of the autoregressive model. Journal of Econometrics 8, 159-172.

Smith M.D., 1988, Convergent series expressions for inverse moments of quadratic forms in normal variables, Australian Journal of Statistics, 30, 235-246.

Stock, J.H., 1987, Asymptotic properties of least squares estimators of cointegrating vectors. Econometrica 55, 1035-1056.

Stock, J.H. and M.W. Watson, 1996, Evidence on Structural Instability in Macroeconomic Time Series Relations. Journal of Business and Economic Statistics 14, 11-30.

Stock, J.H. and M.W. Watson, 1999, A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series. Chapter 1 of Engle, R.F. and H. White (1999), Cointegration, Causality and Forecasting. Oxford University Press.

Stock, J.H. and M.W. Watson, 2003, Combination Forecasts of Output Growth in a Seven-Country Data Set. Harvard University mimeo.

Ullah, A., 2003, Finite sample econometrics. Manuscript, University of California, Riverside.

Appendix A: Distribution of the Initial Values when the Pre-Break Process is I(1)

For the AR(2) case we first note that

$$
\begin{aligned}
y_{m-2} &= y_1 + \Delta y_2 + ... + \Delta y_{m-2}, \\
y_{m-1} &= y_1 + \Delta y_2 + ... + \Delta y_{m-2} + \Delta y_{m-1}. \qquad (37)
\end{aligned}
$$

This provides a decomposition of y_{m-i}, $i = 1, 2$ in terms of the non-stationary level component, y_1, and stationary first differences, $\Delta y_2, \Delta y_3, ...$ The distribution of (y_{m-2}, y_{m-1}) can now be derived for given assumptions concerning y_1 and Δy_2. There are many possibilities. As a simple example we consider the situation where as in the AR(1) case $y_1 \sim N(\mu_1, \omega^2)$ is distributed independently of Δy_t, $t = 2, 3, ..,$ and assume that the stationary components of y_{m-2} and y_{m-1} are started with $\Delta y_1 = 0$. Under these assumptions we have $y_1 = \omega\varepsilon_1$ and $\Delta y_2 = \sigma_1\varepsilon_2$ and, using (13),

$$\Delta y_t = \delta_{11}\Delta y_{t-1} + \sigma_1\varepsilon_t, \; t = 3, 4, ..., T_1,$$

where $|\delta_{11}| < 1$, thus ensuring that $y_t \sim I(1)$. Using these relations we have

$$
\begin{aligned}
\Delta y_2 &= \sigma_1\varepsilon_2 \\
\Delta y_3 &= \delta_{11}\sigma_1\varepsilon_2 + \sigma_1\varepsilon_3 \\
&\vdots \\
\Delta y_{m-2} &= \delta_{11}^{m-4}\sigma_1\varepsilon_2 + \delta_{11}^{m-5}\sigma_1\varepsilon_3 + ... + \delta_{11}\sigma_1\varepsilon_{m-3} + \sigma_1\varepsilon_{m-2} \\
\Delta y_{m-1} &= \delta_{11}^{m-3}\sigma_1\varepsilon_2 + \delta_{11}^{m-4}\sigma_1\varepsilon_3 + ... + \delta_{11}^{2}\sigma_1\varepsilon_{m-3} + \delta_{11}\sigma_1\varepsilon_{m-2} + \sigma_1\varepsilon_{m-1}
\end{aligned}
$$

Substituting these in (37) we now have

$$
\begin{aligned}
y_{m-2} &= y_1 + \frac{\sigma_1\varepsilon_2\left(1-\delta_{11}^{m-3}\right)}{1-\delta_{11}} + \frac{\sigma_1\varepsilon_3\left(1-\delta_{11}^{m-4}\right)}{1-\delta_{11}} + ... + \frac{\sigma_1\varepsilon_{m-2}\left(1-\delta_{11}\right)}{1-\delta_{11}} \\
y_{m-1} &= y_1 + \frac{\sigma_1\varepsilon_2\left(1-\delta_{11}^{m-2}\right)}{1-\delta_{11}} + \frac{\sigma_1\varepsilon_3\left(1-\delta_{11}^{m-3}\right)}{1-\delta_{11}} + ... \frac{\sigma_1\varepsilon_{m-2}\left(1-\delta_{11}^{2}\right)}{1-\delta_{11}} + \frac{\sigma_1\varepsilon_{m-1}\left(1-\delta_{11}\right)}{1-\delta_{11}}.
\end{aligned}
$$

Hence

$$
\begin{aligned}
Var(y_{m-1}) &= \omega^2 + \frac{\sigma_1^2 \sum_{j=1}^{m-2}\left(1-\delta_{11}^j\right)^2}{\left(1-\delta_{11}\right)^2} \\
&= \omega^2 + \frac{\sigma_1^2\left((m-2)(1-\delta_{11}^2)+\delta_{11}^2(1-\delta_{11}^{2(m-2)})-2\delta_{11}(1+\delta_{11})(1-\delta_{11}^{m-2})\right)}{\left(1-\delta_{11}\right)^2\left(1-\delta_{11}^2\right)}, \\
Var(y_{m-2}) &= \omega^2 + \frac{\sigma_1^2\left((m-3)(1-\delta_{11}^2)+\delta_{11}^2(1-\delta_{11}^{2(m-3)})-2\delta_{11}(1+\delta_{11})(1-\delta_{11}^{m-3})\right)}{\left(1-\delta_{11}\right)^2\left(1-\delta_{11}^2\right)}, \\
Cov(y_{m-1}, y_{m-2}) &= \omega^2 + \frac{\sigma_1^2\left((m-3)(1-\delta_{11}^2)+\delta_{11}^3(1-\delta_{11}^{2(m-3)})-\delta_{11}(1+\delta_{11})^2(1-\delta_{11}^{m-3})\right)}{\left(1-\delta_{11}\right)^2\left(1-\delta_{11}^2\right)},
\end{aligned} \tag{38}
$$

so that

$$
\begin{pmatrix} y_{m-2} \\ y_{m-1} \end{pmatrix} \sim N\left(\mu_1 \boldsymbol{\tau}_2, \mathbf{V}_{2,m}\right), \tag{39}
$$

where the elements in the 2×2 matrix $\mathbf{V}_{2,m}$ are given in (38). Fixed (non-stochastic) starting values can also be accommodated by setting $\omega = 0$.

Appendix B: Matrix Notations

This appendix provides details of some of the derivations from the main text. The sub-matrices $\mathbf{B}_{ij}$ in (18) are given by

$$
\underset{v_1\times p}{\mathbf{B}_{21}} = \begin{pmatrix}
-\beta_{1p} & -\beta_{1,p-1} & \cdots & & -\beta_{11} \\
0 & -\beta_{1p} & \cdots & & -\beta_{12} \\
\vdots & \vdots & \ddots & & \vdots \\
 & & & -\beta_{1p} & -\beta_{1,p-1} \\
0 & 0 & \cdots & 0 & -\beta_{1p} \\
\vdots & \vdots & \cdots & & 0 \\
0 & 0 & \cdots & & 0
\end{pmatrix},
$$

$$
\underset{v_1\times v_1}{\mathbf{B}_{22}} = \begin{pmatrix}
1 & 0 & \cdots & \cdots & & & \cdots & & & \cdots & 0 & 0 \\
-\beta_{11} & 1 & \cdots & \cdots & & & & & & & 0 & 0 \\
\vdots & \vdots & \ddots & \cdots & \vdots & & \vdots & \vdots & & & 0 & 0 \\
-\beta_{1p} & -\beta_{1,p-1} & \cdots & \cdots & -\beta_{11} & 1 & \cdots & \cdots & & & 0 & 0 \\
0 & -\beta_{1p} & \cdots & \cdots & -\beta_{12} & -\beta_{11} & 1 & \cdots & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & -\beta_{1,p-1} & \cdots & 1 & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & -\beta_{1p} & \cdots & -\beta_{11} & 1
\end{pmatrix},
$$

$$
\underset{v_2\times v_1}{\mathbf{B}_{32}} = \begin{pmatrix}
0 & 0 & \cdots & 0 & -\beta_{2p} & -\beta_{2,p-1} & \cdots & -\beta_{22} & -\beta_{21} \\
0 & 0 & \cdots & & 0 & -\beta_{2p} & \cdots & -\beta_{23} & -\beta_{22} \\
\vdots & \vdots & \ddots & \vdots & 0 & 0 & \cdots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & & & \cdots & -\beta_{2p} & -\beta_{2,p-1} \\
0 & 0 & \cdots & 0 & & & \cdots & 0 & -\beta_{2p} \\
0 & 0 & \cdots & 0 & & & \cdots & 0 & 0 \\
\vdots & \vdots & & \vdots & & & \cdots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & & & \cdots & 0 & 0
\end{pmatrix},
$$

and

$$
\underset{v_2\times v_2}{\mathbf{B}_{33}} = \begin{pmatrix}
1 & 0 & \cdots & \cdots & & & \cdots & & & \cdots & 0 & 0 \\
-\beta_{21} & 1 & \cdots & \cdots & & & & & & & 0 & 0 \\
\vdots & \vdots & \ddots & \cdots & \vdots & & & & & & 0 & 0 \\
-\beta_{2p} & -\beta_{2,p-1} & \cdots & \cdots & -\beta_{21} & 1 & \cdots & & & \cdots & 0 & 0 \\
0 & -\beta_{2p} & \cdots & \cdots & -\beta_{22} & -\beta_{21} & 1 & & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & -\beta_{2,p-1} & \cdots & 1 & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & -\beta_{2p} & \cdots & -\beta_{21} & 1
\end{pmatrix}.
$$

Matrix $\mathbf{B}$ is lower triangular with diagonal elements equal to unity and is therefore non-singular and we have

$$
\mathbf{y}_T(m-p) = \mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon},
$$

where $\mathbf{c} = \mathbf{B^{-1}d}$, and $\mathbf{H} = \mathbf{B}^{-1}\mathbf{D}$. It is now easily seen that

$$
\mathbf{y}_{T-i}(m-i) = \mathbf{G}_i\mathbf{y}_T(m-p) = \mathbf{G}_i\mathbf{c} + \mathbf{G}_i\mathbf{H}\boldsymbol{\varepsilon}, \tag{40}
$$

for $i = 0, 1, ..., p$, where $\mathbf{G}_i$ are $v \times (v+p)$ selection matrices defined by

$$\mathbf{G}_i = (\mathbf{0}_{v\times p-i} \vdots \mathbf{I}_\nu \vdots \mathbf{0}_{v\times i}), \text{ for } i = 0, 1, 2, ..., p.$$

$\mathbf{0}_{v\times p-i}$ is a $v \times (p-i)$ matrix of zeros and $\mathbf{G}_0 = (\mathbf{0}_{v\times p} \vdots \mathbf{I}_\nu)$, and $\mathbf{G}_p = \left(\mathbf{I}_\nu \vdots \mathbf{0}_{v\times p}\right)$.

The deterministic components, $\mathbf{G}_i\mathbf{c}$, in the expressions for $\mathbf{y}_{T-i}(m-i)$ simplify if there is no mean shift ($\mu_1 = \mu_2$) or if $\beta_2^* = 1$. To see this we first note that

$$\mathbf{B}\boldsymbol{\tau}_{v+p} = \begin{pmatrix} \boldsymbol{\tau}_p \\ (1-\beta_1^*)\,\boldsymbol{\tau}_{v_1} \\ (1-\beta_2^*)\,\boldsymbol{\tau}_{v_2} \end{pmatrix}. \tag{41a}$$

Also

$$\mathbf{d} = \begin{pmatrix} \mu_1\boldsymbol{\tau}_p \\ \mu_1(1-\beta_1^*)\boldsymbol{\tau}_{v_1} \\ \mu_2(1-\beta_2^*)\boldsymbol{\tau}_{v_2} \end{pmatrix} = \mu_1 \begin{pmatrix} \boldsymbol{\tau}_p \\ (1-\beta_1^*)\,\boldsymbol{\tau}_{v_1} \\ (1-\beta_2^*)\,\boldsymbol{\tau}_{v_2} \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{p\times 1} \\ \mathbf{0}_{v_1\times 1} \\ (1-\beta_2^*)\,(\mu_2-\mu_1)\,\boldsymbol{\tau}_{v_2} \end{pmatrix},$$

and using (41a)

$$\mathbf{d} = \mu_1\mathbf{B}\boldsymbol{\tau}_{v+p} + \mathbf{g},$$

where

$$\mathbf{g} = \begin{pmatrix} \mathbf{0}_{p\times 1} \\ \mathbf{0}_{v_1\times 1} \\ (1-\beta_2^*)\,(\mu_2-\mu_1)\,\boldsymbol{\tau}_{v_2} \end{pmatrix}.$$

Hence

$$\mathbf{c} = \mathbf{B}^{-1}\mathbf{d} = \mu_1\boldsymbol{\tau}_{v+p} + \mathbf{B}^{-1}\mathbf{g},$$

When there is no mean shift, $\mu_1 = \mu_2 = \mu$, then $\mathbf{g} = \mathbf{0}$, and we have

$$\mathbf{G}_i\mathbf{c} = \mathbf{G}_i\mathbf{B}^{-1}\mathbf{d} = \mu\,\mathbf{G}_i\boldsymbol{\tau}_{v+p} = \mu\boldsymbol{\tau}_v. \tag{42}$$

Similarly, if $\mu_1 \neq \mu_2$ but $\beta_2^* = 1$ we have

$$\mathbf{G}_i\mathbf{c} = \mu_1\boldsymbol{\tau}_v. \tag{43}$$

Table 1: Breakpoint Specifications by Experiments

Experiments	μ_1	μ_2	β_{11}	β_{12}	σ_1	σ_2
1: No break	1	1	0.9	0.9	1	1
2: Moderate break in β_1	1	1	0.6	0.9	1	1
3: Moderate break in β_1 (decline)	1	1	0.9	0.6	1	1
4: Large break in β_1	1	1	0.3	0.9	1	1
5: Post-break unit root	1	1	0.6	1	1	1
6: Higher post-break volatility	1	1	0.9	0.9	0.25	1
7: Lower post-break volatility	1	1	0.9	0.9	4	1
8: Break in mean (increase)	1	2	0.9	0.9	1	1
9: Pre-break unit root	1	1	1	0.6	1	1

Table 2: Small sample bias of the OLS estimate of β as a function of pre-break (v_1) and post-break (v_2) windows

Experiment no. 1: No break

v1/v2	10	20	30	50	100
0	-0.370	-0.200	-0.135	-0.081	-0.039
1	-0.344	-0.193	-0.132	-0.080	-0.039
2	-0.315	-0.184	-0.129	-0.078	-0.039
3	-0.297	-0.175	-0.125	-0.076	-0.038
4	-0.278	-0.170	-0.119	-0.076	-0.038
5	-0.262	-0.162	-0.116	-0.074	-0.037
10	-0.202	-0.136	-0.102	-0.068	-0.035
20	-0.136	-0.102	-0.081	-0.057	-0.032
30	-0.102	-0.082	-0.066	-0.050	-0.030
50	-0.067	-0.058	-0.049	-0.040	-0.026
100	-0.036	-0.033	-0.030	-0.026	-0.020

Experiment no. 4: Large break in β

v1/v2	10	20	30	50	100
0	-0.393	-0.214	-0.145	-0.085	-0.041
1	-0.389	-0.222	-0.150	-0.089	-0.042
2	-0.371	-0.217	-0.153	-0.091	-0.044
3	-0.364	-0.216	-0.154	-0.092	-0.045
4	-0.358	-0.220	-0.154	-0.095	-0.046
5	-0.355	-0.220	-0.155	-0.098	-0.047
10	-0.363	-0.233	-0.169	-0.107	-0.053
20	-0.393	-0.263	-0.196	-0.127	-0.065
30	-0.416	-0.292	-0.220	-0.147	-0.078
50	-0.452	-0.337	-0.261	-0.183	-0.099
100	-0.499	-0.402	-0.333	-0.245	-0.145

Experiment no. 7: Lower post-break volatility

v1/v2	10	20	30	50	100
0	-0.221	-0.118	-0.082	-0.052	-0.029
1	-0.326	-0.173	-0.119	-0.072	-0.036
2	-0.335	-0.192	-0.133	-0.083	-0.043
3	-0.327	-0.191	-0.137	-0.086	-0.046
4	-0.312	-0.191	-0.135	-0.090	-0.048
5	-0.298	-0.185	-0.136	-0.089	-0.049
10	-0.227	-0.155	-0.122	-0.086	-0.051
20	-0.149	-0.116	-0.095	-0.073	-0.049
30	-0.110	-0.091	-0.076	-0.062	-0.043
50	-0.071	-0.063	-0.056	-0.047	-0.035
100	-0.037	-0.035	-0.033	-0.029	-0.025

Experiment no. 2: Moderate break in β

v1/v2	10	20	30	50	100
0	-0.391	-0.213	-0.144	-0.085	-0.041
1	-0.376	-0.215	-0.146	-0.087	-0.041
2	-0.354	-0.208	-0.147	-0.088	-0.043
3	-0.340	-0.204	-0.146	-0.088	-0.043
4	-0.325	-0.203	-0.143	-0.090	-0.044
5	-0.315	-0.199	-0.142	-0.091	-0.045
10	-0.286	-0.193	-0.143	-0.094	-0.047
20	-0.271	-0.194	-0.150	-0.102	-0.055
30	-0.267	-0.200	-0.157	-0.110	-0.062
50	-0.269	-0.213	-0.172	-0.128	-0.074
100	-0.277	-0.234	-0.201	-0.156	-0.099

Experiment no. 5: Post-break unit root

v1/v2	10	20	30	50	100
0	-0.413	-0.232	-0.163	-0.102	-0.052
1	-0.394	-0.230	-0.159	-0.100	-0.052
2	-0.367	-0.217	-0.156	-0.099	-0.051
3	-0.352	-0.211	-0.152	-0.097	-0.051
4	-0.336	-0.206	-0.149	-0.095	-0.050
5	-0.327	-0.201	-0.144	-0.095	-0.050
10	-0.303	-0.191	-0.138	-0.089	-0.047
20	-0.297	-0.191	-0.138	-0.088	-0.046
30	-0.299	-0.199	-0.142	-0.090	-0.047
50	-0.312	-0.216	-0.157	-0.099	-0.048
100	-0.335	-0.248	-0.188	-0.119	-0.057

Experiment no. 8: Break in mean (increase)

v1/v2	10	20	30	50	100
0	-0.365	-0.197	-0.134	-0.080	-0.039
1	-0.337	-0.189	-0.129	-0.078	-0.038
2	-0.309	-0.179	-0.125	-0.077	-0.038
3	-0.290	-0.170	-0.122	-0.074	-0.037
4	-0.272	-0.165	-0.115	-0.073	-0.037
5	-0.256	-0.157	-0.111	-0.071	-0.036
10	-0.196	-0.130	-0.097	-0.064	-0.033
20	-0.132	-0.097	-0.076	-0.053	-0.029
30	-0.099	-0.077	-0.062	-0.045	-0.027
50	-0.066	-0.055	-0.045	-0.035	-0.021
100	-0.035	-0.031	-0.027	-0.022	-0.015

Experiment no. 3: Moderate break in β (decline)

v1/v2	10	20	30	50	100
0	-0.228	-0.124	-0.086	-0.053	-0.027
1	-0.156	-0.085	-0.058	-0.035	-0.018
2	-0.114	-0.060	-0.040	-0.023	-0.010
3	-0.081	-0.037	-0.022	-0.009	-0.002
4	-0.058	-0.020	-0.007	-0.001	0.004
5	-0.037	-0.005	0.005	0.010	0.010
10	0.040	0.053	0.052	0.047	0.037
20	0.118	0.116	0.111	0.098	0.076
30	0.161	0.155	0.148	0.132	0.105
50	0.207	0.199	0.191	0.176	0.147
100	0.250	0.243	0.237	0.226	0.200

Experiment no. 6: Higher post-break volatility

v1/v2	10	20	30	50	100
0	-0.399	-0.218	-0.147	-0.086	-0.041
1	-0.341	-0.204	-0.140	-0.085	-0.040
2	-0.302	-0.187	-0.135	-0.083	-0.041
3	-0.280	-0.176	-0.130	-0.080	-0.040
4	-0.260	-0.170	-0.123	-0.079	-0.040
5	-0.245	-0.161	-0.118	-0.078	-0.039
10	-0.198	-0.138	-0.105	-0.071	-0.037
20	-0.150	-0.112	-0.088	-0.062	-0.035
30	-0.118	-0.097	-0.076	-0.056	-0.033
50	-0.085	-0.076	-0.063	-0.049	-0.030
100	-0.048	-0.050	-0.046	-0.037	-0.025

Experiment no. 9: Pre-break unit root

v1/v2	10	20	30	50	100
0	-0.097	-0.061	-0.045	-0.031	-0.019
1	0.056	0.061	0.062	0.058	0.051
2	0.128	0.120	0.117	0.108	0.092
3	0.168	0.160	0.151	0.141	0.121
4	0.196	0.185	0.177	0.164	0.142
5	0.215	0.206	0.195	0.182	0.160
10	0.272	0.256	0.247	0.235	0.211
20	0.309	0.297	0.292	0.281	0.258
30	0.324	0.317	0.314	0.303	0.285
50	0.336	0.334	0.333	0.327	0.313
100	0.344	0.345	0.344	0.342	0.334

Note: Experiments 1 to 9 are defined in Table 1.

Table 3: Bias of forecast error conditional on $y_T = \mu_2 + \sigma_2$

Experiment no. 1:No break

v1/v2	10	20	30	50	100
0	0.370	0.200	0.135	0.081	0.039
1	0.344	0.193	0.132	0.080	0.039
2	0.315	0.184	0.129	0.078	0.039
3	0.297	0.175	0.125	0.076	0.038
4	0.278	0.170	0.119	0.076	0.038
5	0.262	0.162	0.116	0.074	0.037
10	0.202	0.136	0.102	0.068	0.035
20	0.136	0.102	0.081	0.057	0.032
30	0.102	0.082	0.066	0.050	0.030
50	0.067	0.058	0.049	0.040	0.026
100	0.036	0.033	0.030	0.026	0.020

Experiment no. 2: Moderate break in β

v1/v2	10	20	30	50	100
0	0.391	0.213	0.144	0.085	0.041
1	0.376	0.215	0.146	0.087	0.041
2	0.354	0.208	0.147	0.088	0.043
3	0.340	0.204	0.146	0.088	0.043
4	0.325	0.203	0.143	0.090	0.044
5	0.315	0.199	0.142	0.091	0.045
10	0.286	0.193	0.143	0.094	0.047
20	0.271	0.194	0.150	0.102	0.055
30	0.267	0.200	0.157	0.110	0.062
50	0.269	0.213	0.172	0.128	0.074
100	0.277	0.234	0.201	0.156	0.099

Experiment no. 3: Moderate break in β (decline)

v1/v2	10	20	30	50	100
0	0.228	0.124	0.086	0.053	0.027
1	0.156	0.085	0.058	0.035	0.018
2	0.114	0.060	0.040	0.023	0.010
3	0.081	0.037	0.022	0.009	0.002
4	0.058	0.020	0.007	0.001	-0.004
5	0.037	0.005	-0.005	-0.010	-0.010
10	-0.040	-0.053	-0.052	-0.047	-0.037
20	-0.118	-0.116	-0.111	-0.098	-0.076
30	-0.161	-0.155	-0.148	-0.132	-0.105
50	-0.207	-0.199	-0.191	-0.176	-0.147
100	-0.250	-0.243	-0.237	-0.226	-0.200

Experiment no. 4: Large break in β

v1/v2	10	20	30	50	100
0	0.393	0.214	0.145	0.085	0.041
1	0.389	0.222	0.150	0.089	0.042
2	0.371	0.217	0.153	0.091	0.044
3	0.364	0.216	0.154	0.092	0.045
4	0.358	0.220	0.154	0.095	0.046
5	0.355	0.220	0.155	0.098	0.047
10	0.363	0.233	0.169	0.107	0.053
20	0.393	0.263	0.196	0.127	0.065
30	0.416	0.292	0.220	0.147	0.078
50	0.452	0.337	0.261	0.183	0.099
100	0.499	0.402	0.333	0.245	0.145

Experiment no. 5: Post-break unit root

v1/v2	10	20	30	50	100
0	0.413	0.232	0.163	0.102	0.052
1	0.394	0.230	0.159	0.100	0.052
2	0.367	0.217	0.156	0.099	0.051
3	0.352	0.211	0.152	0.097	0.051
4	0.336	0.206	0.149	0.095	0.050
5	0.327	0.201	0.144	0.095	0.050
10	0.303	0.191	0.138	0.089	0.047
20	0.297	0.191	0.138	0.088	0.046
30	0.299	0.199	0.142	0.090	0.047
50	0.312	0.216	0.157	0.099	0.048
100	0.335	0.248	0.188	0.119	0.057

Experiment no. 6: Higher post-break volatility

v1/v2	10	20	30	50	100
0	0.399	0.218	0.147	0.086	0.041
1	0.341	0.204	0.140	0.085	0.040
2	0.302	0.187	0.135	0.083	0.041
3	0.280	0.176	0.130	0.080	0.040
4	0.260	0.170	0.123	0.079	0.040
5	0.245	0.161	0.118	0.078	0.039
10	0.198	0.138	0.105	0.071	0.037
20	0.150	0.112	0.088	0.062	0.035
30	0.118	0.097	0.076	0.056	0.033
50	0.085	0.076	0.063	0.049	0.030
100	0.048	0.050	0.046	0.037	0.025

Experiment no. 7: Lower post-break volatility

v1/v2	10	20	30	50	100
0	0.221	0.118	0.082	0.052	0.029
1	0.326	0.173	0.119	0.072	0.036
2	0.335	0.192	0.133	0.083	0.043
3	0.327	0.191	0.137	0.086	0.046
4	0.312	0.191	0.135	0.090	0.048
5	0.298	0.185	0.136	0.089	0.049
10	0.227	0.155	0.122	0.086	0.051
20	0.149	0.116	0.095	0.073	0.049
30	0.110	0.091	0.076	0.062	0.043
50	0.071	0.063	0.056	0.047	0.035
100	0.037	0.035	0.033	0.029	0.025

Experiment no. 8: Break in mean (increase)

v1/v2	10	20	30	50	100
0	0.604	0.286	0.178	0.096	0.043
1	0.583	0.284	0.178	0.098	0.044
2	0.552	0.282	0.181	0.100	0.045
3	0.529	0.276	0.182	0.100	0.046
4	0.507	0.276	0.178	0.103	0.047
5	0.486	0.269	0.178	0.103	0.047
10	0.404	0.247	0.173	0.106	0.050
20	0.309	0.214	0.161	0.106	0.054
30	0.260	0.193	0.151	0.106	0.059
50	0.208	0.169	0.140	0.106	0.064
100	0.159	0.141	0.126	0.104	0.073

Experiment no. 9: Pre-break unit root

v1/v2	10	20	30	50	100
0	0.097	0.061	0.045	0.031	0.019
1	-0.056	-0.061	-0.062	-0.058	-0.051
2	-0.128	-0.120	-0.117	-0.108	-0.092
3	-0.168	-0.160	-0.151	-0.141	-0.121
4	-0.196	-0.185	-0.177	-0.164	-0.142
5	-0.215	-0.206	-0.195	-0.182	-0.160
10	-0.272	-0.256	-0.247	-0.235	-0.211
20	-0.309	-0.297	-0.292	-0.281	-0.258
30	-0.324	-0.317	-0.314	-0.303	-0.285
50	-0.336	-0.334	-0.333	-0.327	-0.313
100	-0.344	-0.345	-0.344	-0.342	-0.334

See the note to Table 2.

Table 4: Unconditional RMSFE as a function of pre-break (v_1) and post-break window (v_2)

Experiment no. 1:No break

v1/v2	10	20	30	50	100
0	1.149	1.078	1.051	1.028	1.012
1	1.140	1.075	1.048	1.027	1.012
2	1.127	1.072	1.047	1.026	1.012
3	1.120	1.070	1.046	1.026	1.012
4	1.112	1.066	1.044	1.025	1.011
5	1.104	1.063	1.043	1.025	1.012
10	1.075	1.051	1.037	1.023	1.011
20	1.047	1.035	1.028	1.019	1.010
30	1.034	1.027	1.021	1.015	1.009
50	1.022	1.018	1.016	1.012	1.008
100	1.011	1.010	1.009	1.008	1.006

Experiment no. 2: Moderate break in β

v1/v2	10	20	30	50	100
0	1.128	1.071	1.048	1.028	1.012
1	1.120	1.069	1.047	1.028	1.013
2	1.109	1.066	1.046	1.028	1.013
3	1.100	1.064	1.046	1.027	1.012
4	1.092	1.061	1.044	1.027	1.012
5	1.085	1.059	1.042	1.026	1.012
10	1.073	1.052	1.039	1.025	1.012
20	1.070	1.048	1.037	1.025	1.012
30	1.074	1.051	1.038	1.025	1.013
50	1.085	1.059	1.045	1.030	1.015
100	1.110	1.080	1.065	1.044	1.022

Experiment no. 3: Moderate break in β (decline)

v1/v2	10	20	30	50	100
0	1.117	1.054	1.035	1.021	1.010
1	1.111	1.050	1.033	1.020	1.010
2	1.113	1.051	1.033	1.020	1.010
3	1.110	1.052	1.034	1.020	1.010
4	1.114	1.053	1.035	1.021	1.011
5	1.112	1.054	1.037	1.022	1.011
10	1.099	1.056	1.040	1.026	1.013
20	1.078	1.055	1.043	1.031	1.018
30	1.069	1.055	1.045	1.034	1.021
50	1.063	1.054	1.049	1.040	1.028
100	1.060	1.057	1.054	1.048	1.039

Experiment no. 4: Large break in β

v1/v2	10	20	30	50	100
0	1.125	1.070	1.048	1.028	1.012
1	1.122	1.071	1.048	1.029	1.013
2	1.114	1.069	1.048	1.028	1.013
3	1.106	1.067	1.048	1.028	1.013
4	1.101	1.066	1.047	1.028	1.013
5	1.097	1.065	1.046	1.028	1.013
10	1.103	1.066	1.047	1.029	1.013
20	1.132	1.080	1.055	1.033	1.015
30	1.163	1.099	1.068	1.039	1.018
50	1.216	1.139	1.096	1.058	1.024
100	1.307	1.215	1.164	1.101	1.045

Experiment no. 5: Post-break unit root

v1/v2	10	20	30	50	100
0	1.112	1.062	1.044	1.028	1.014
1	1.110	1.062	1.043	1.027	1.014
2	1.105	1.059	1.042	1.027	1.014
3	1.102	1.059	1.042	1.027	1.014
4	1.101	1.058	1.042	1.026	1.014
5	1.101	1.058	1.041	1.025	1.014
10	1.115	1.063	1.042	1.026	1.013
20	1.148	1.081	1.051	1.030	1.014
30	1.176	1.100	1.064	1.034	1.015
50	1.221	1.141	1.093	1.048	1.018
100	1.308	1.227	1.166	1.087	1.030

Experiment no. 6: Higher post-break volatility

v1/v2	10	20	30	50	100
0	1.126	1.070	1.049	1.028	1.013
1	1.107	1.063	1.043	1.027	1.012
2	1.097	1.057	1.041	1.026	1.012
3	1.090	1.054	1.040	1.024	1.012
4	1.086	1.050	1.037	1.024	1.011
5	1.082	1.047	1.035	1.023	1.011
10	1.072	1.040	1.030	1.020	1.011
20	1.058	1.034	1.024	1.017	1.009
30	1.047	1.030	1.021	1.014	1.009
50	1.036	1.024	1.019	1.012	1.008
100	1.022	1.017	1.014	1.010	1.006

Experiment no. 7: Lower post-break volatility

v1/v2	10	20	30	50	100
0	1.302	1.133	1.072	1.030	1.011
1	1.539	1.258	1.140	1.055	1.017
2	1.576	1.310	1.171	1.070	1.021
3	1.595	1.322	1.189	1.077	1.024
4	1.597	1.327	1.192	1.083	1.025
5	1.569	1.330	1.202	1.087	1.027
10	1.481	1.291	1.196	1.095	1.032
20	1.356	1.229	1.160	1.087	1.034
30	1.282	1.193	1.135	1.079	1.034
50	1.197	1.144	1.110	1.070	1.033
100	1.106	1.086	1.074	1.053	1.031

Experiment no. 8: Break in mean (increase)

v1/v2	10	20	30	50	100
0	1.149	1.078	1.050	1.028	1.012
1	1.141	1.075	1.048	1.027	1.012
2	1.129	1.072	1.047	1.026	1.012
3	1.122	1.070	1.047	1.026	1.012
4	1.115	1.066	1.044	1.025	1.011
5	1.106	1.064	1.043	1.024	1.011
10	1.079	1.053	1.037	1.023	1.011
20	1.051	1.037	1.029	1.019	1.010
30	1.039	1.030	1.023	1.016	1.009
50	1.027	1.021	1.018	1.014	1.008
100	1.017	1.014	1.012	1.010	1.007

Experiment no. 9: Pre-break unit root

v1/v2	10	20	30	50	100
0	1.072	1.037	1.026	1.016	1.008
1	1.076	1.042	1.031	1.020	1.012
2	1.085	1.052	1.040	1.029	1.018
3	1.092	1.060	1.048	1.036	1.024
4	1.098	1.067	1.055	1.042	1.029
5	1.102	1.071	1.059	1.046	1.034
10	1.112	1.085	1.073	1.062	1.048
20	1.118	1.097	1.087	1.078	1.062
30	1.118	1.105	1.096	1.084	1.073
50	1.119	1.108	1.102	1.093	1.083
100	1.113	1.105	1.101	1.096	1.088

See the note to Table 2.

Table 5. RMSFE conditional on $y_T = \mu_2 + \sigma_2$

Experiment no. 1:No break

v1/v2	10	20	30	50	100
0	1.421	1.164	1.088	1.038	1.013
1	1.375	1.153	1.085	1.037	1.012
2	1.329	1.142	1.079	1.036	1.012
3	1.301	1.130	1.075	1.034	1.012
4	1.274	1.123	1.069	1.033	1.012
5	1.248	1.115	1.067	1.032	1.011
10	1.165	1.085	1.053	1.028	1.011
20	1.086	1.054	1.037	1.021	1.009
30	1.053	1.037	1.027	1.017	1.008
50	1.027	1.021	1.017	1.012	1.007
100	1.011	1.009	1.008	1.007	1.005

Experiment no. 2: Moderate break in β

v1/v2	10	20	30	50	100
0	1.352	1.154	1.086	1.039	1.013
1	1.303	1.145	1.085	1.039	1.013
2	1.259	1.130	1.078	1.037	1.013
3	1.226	1.116	1.073	1.035	1.013
4	1.199	1.108	1.067	1.035	1.012
5	1.177	1.099	1.062	1.033	1.012
10	1.119	1.073	1.050	1.029	1.012
20	1.079	1.054	1.039	1.024	1.011
30	1.064	1.046	1.034	1.022	1.011
50	1.053	1.040	1.030	1.020	1.010
100	1.046	1.036	1.029	1.020	1.011

Experiment no. 3: Moderate break in β (decline)

v1/v2	10	20	30	50	100
0	1.161	1.066	1.039	1.021	1.009
1	1.135	1.056	1.035	1.019	1.009
2	1.126	1.053	1.033	1.019	1.009
3	1.117	1.051	1.032	1.018	1.009
4	1.115	1.050	1.032	1.018	1.009
5	1.109	1.050	1.032	1.018	1.009
10	1.088	1.046	1.032	1.020	1.010
20	1.062	1.041	1.032	1.022	1.012
30	1.051	1.038	1.032	1.024	1.015
50	1.043	1.037	1.033	1.027	1.019
100	1.040	1.038	1.036	1.032	1.025

Experiment no. 4: Large break in β

v1/v2	10	20	30	50	100
0	1.342	1.153	1.086	1.039	1.013
1	1.296	1.145	1.086	1.039	1.013
2	1.253	1.130	1.079	1.038	1.013
3	1.220	1.116	1.074	1.036	1.013
4	1.198	1.109	1.068	1.036	1.013
5	1.181	1.102	1.065	1.035	1.013
10	1.141	1.084	1.057	1.032	1.013
20	1.122	1.075	1.052	1.030	1.013
30	1.117	1.074	1.051	1.030	1.013
50	1.119	1.079	1.055	1.033	1.014
100	1.129	1.092	1.069	1.042	1.019

Experiment no. 5: Post-break unit root

v1/v2	10	20	30	50	100
0	1.457	1.264	1.183	1.121	1.063
1	1.375	1.231	1.169	1.112	1.060
2	1.318	1.198	1.148	1.099	1.056
3	1.277	1.172	1.130	1.088	1.052
4	1.242	1.155	1.116	1.080	1.048
5	1.213	1.140	1.105	1.074	1.045
10	1.141	1.098	1.075	1.055	1.035
20	1.096	1.068	1.052	1.038	1.025
30	1.080	1.056	1.043	1.031	1.020
50	1.069	1.048	1.036	1.025	1.015
100	1.066	1.045	1.033	1.021	1.011

Experiment no. 6: Higher post-break volatility

v1/v2	10	20	30	50	100
0	1.316	1.148	1.085	1.039	1.013
1	1.222	1.120	1.074	1.036	1.012
2	1.172	1.100	1.066	1.034	1.012
3	1.143	1.085	1.059	1.031	1.012
4	1.124	1.076	1.052	1.029	1.011
5	1.110	1.068	1.047	1.028	1.011
10	1.072	1.045	1.033	1.022	1.010
20	1.043	1.028	1.021	1.015	1.008
30	1.029	1.021	1.016	1.011	1.007
50	1.017	1.013	1.010	1.008	1.005
100	1.007	1.006	1.006	1.004	1.003

Experiment no. 7: Lower post-break volatility

v1/v2	10	20	30	50	100
0	1.851	1.237	1.103	1.037	1.011
1	2.339	1.419	1.189	1.064	1.016
2	2.375	1.477	1.222	1.078	1.019
3	2.357	1.488	1.239	1.084	1.021
4	2.308	1.491	1.242	1.088	1.023
5	2.217	1.488	1.251	1.093	1.023
10	1.927	1.408	1.234	1.099	1.028
20	1.567	1.303	1.188	1.090	1.030
30	1.396	1.241	1.154	1.082	1.030
50	1.240	1.166	1.120	1.070	1.030
100	1.111	1.090	1.075	1.053	1.028

Experiment no. 8: Break in mean (increase)

v1/v2	10	20	30	50	100
0	1.507	1.189	1.098	1.041	1.013
1	1.464	1.179	1.096	1.041	1.013
2	1.416	1.170	1.092	1.040	1.013
3	1.381	1.159	1.089	1.039	1.013
4	1.352	1.152	1.083	1.038	1.013
5	1.322	1.144	1.081	1.037	1.012
10	1.225	1.113	1.067	1.034	1.012
20	1.129	1.077	1.051	1.027	1.011
30	1.088	1.057	1.040	1.023	1.010
50	1.051	1.038	1.029	1.019	1.009
100	1.025	1.021	1.018	1.013	1.008

Experiment no. 9: Pre-break unit root

v1/v2	10	20	30	50	100
0	1.088	1.043	1.028	1.016	1.008
1	1.060	1.031	1.022	1.014	1.008
2	1.056	1.033	1.025	1.018	1.012
3	1.058	1.036	1.029	1.022	1.015
4	1.059	1.039	1.032	1.025	1.018
5	1.062	1.042	1.035	1.028	1.021
10	1.070	1.052	1.046	1.039	1.031
20	1.078	1.064	1.057	1.050	1.042
30	1.081	1.070	1.064	1.057	1.049
50	1.083	1.074	1.070	1.065	1.057
100	1.079	1.072	1.069	1.065	1.060

See the note to Table 2.

Table 6: RMSFE-values

A. Single break										
Experiment no.	1a	2a	3a	4[a]	5a	6a	7a	8[a]	9a	10a
	No break	Moderate Break in β	Decline In β	Large break in β	Post-break unit root	Higher post-break vol.	Lower post-break vol.	Increase in mean	Pre-break unit root	Pre-break unit root
Expanding window	1.006	1.066	1.052	1.186	1.157	1.007	1.066	1.010	1.104	1.096
Rolling window (short)	1.053	1.058	1.046	1.059	1.058	1.049	1.146	1.056	1.067	1.041
Rolling window (long)	1.020	1.040	1.037	1.071	1.061	1.021	1.147	1.026	1.092	1.045
Post-break window	1.011	1.086	1.058	1.133	1.115	1.124	1.104	1.021	1.051	1.072
B. Multiple breaks										
Experiment no.	11	12	13	14	15	16	17	18	19	20
	Continued rise in β	Mean-Reverting β	Continued decline in β	Mean-reverting β	Large rise In β	Volatility break	Volatility break	Mean reverting intercept	Pre-break unit root	Pre-break unit root
Expanding window	1.038	1.028	1.033	1.011	1.120	1.035	1.006	1.003	1.026	1.026
Rolling window (short)	1.051	1.049	1.040	1.051	1.060	1.141	1.056	1.053	1.053	1.053
Rolling window (long)	1.024	1.039	1.017	1.036	1.041	1.140	1.027	1.023	1.058	1.058
Post-break window	1.055	1.037	1.035	1.030	1.105	1.202	1.044	1.014	1.084	1.065

Note: Experimental designs are described in Table 1 for experiments 1-9.
Experiment 10 assumes mean-reversion to the end-point of the pre-break unit root process.
Experiments 11-20 assume two breaks occurring after 50 and 100 periods.
Experiment 11 lets the AR(1) coefficient go from 0.6 to 0.75 to 0.9.
Experiment 12 lets the AR(1) coefficient go from 0.6 to 0.9 to 0.6.
Experiment 13 lets the AR(1) coefficient go from 0.9 to 0.75 to 0.6.
Experiment 14 lets the AR(1) coefficient go from 0.9 to 0.6 to 0.9.
Experiment 15 lets the AR(1) coefficient go from 0.3 to 0.6 to 0.9.
Experiment 16 lets volatility go from 1 to 4 to 1.
Experiment 17 lets volatility go from 1 to 0.25 to 1.
Experiment 18 lets the intercept go from 1 to 2 to 1.
Experiments 19 and 20 assume unit roots before the first and after the second break with AR(1) parameters of 1, 0.6 and 1, respectively.
Experiment 19 lets the mean of the stationary segment be independent of the terminal point of the unit root process, while experiment 20 assumes the stationary part of the process has a mean determined by the end-point of the unit root process. Parameters not changed in the experiments are kept at their base values with an AR(1) parameter of 0.9, volatility of 1 and an intercept of 1.

Table 7: Out-of-sample forecasting performance

RMSFE-values						
A. AR(1) Models	**Canada**	**France**	**Germany**	**Japan**	**UK**	**US**
Expanding window						
Inflation rate	0.451	0.401	0.471	0.927	0.833	0.446
Industrial production growth	1.712	1.599	1.825	1.710	1.977	1.494
Real GDP growth	0.917	0.631	1.299	1.300	1.050	0.867
Interest rate	1.291	1.179	0.548	NA	1.239	1.298
Short rolling window						
Inflation rate	0.456	0.398	0.459	0.900	0.842	0.441
Industrial production growth	1.798	1.605	1.854	1.695	2.046	1.552
Real GDP growth	0.901	0.641	1.348	1.104	1.055	0.915
Interest rate	1.347	1.266	0.597	NA	1.257	1.421
Long rolling window						
Inflation rate	0.453	0.400	0.465	0.890	0.845	0.449
Industrial production growth	1.723	1.599	1.846	1.683	1.988	1.513
Real GDP growth	0.874	0.636	1.286	1.076	1.047	0.880
Interest rate	1.300	1.176	0.551	NA	1.236	1.356
Post-break window						
Inflation rate	0.437	0.418	0.540	1.037	0.933	0.477
Industrial production growth	1.854	1.712	1.846	1.828	2.442	1.550
Real GDP growth	0.955	0.638	1.299	1.164	1.050	0.875
Interest rate	1.538	1.277	0.661	NA	1.385	1.508
B. AR(2) Models	**Canada**	**France**	**Germany**	**Japan**	**UK**	**US**
Expanding window						
Inflation rate	0.437	0.399	0.469	0.890	0.844	0.448
Industrial production growth	1.711	1.565	1.837	1.723	2.015	1.484
Real GDP growth	0.916	0.610	1.308	1.167	1.057	0.865
Interest rate	1.309	1.208	0.529	NA	1.254	1.344
Short rolling window						
Inflation rate	0.451	0.412	0.483	0.927	0.875	0.457
Industrial production growth	1.815	1.622	1.906	1.743	2.144	1.571
Real GDP growth	0.917	0.629	1.367	1.112	1.090	0.923
Interest rate	1.397	1.294	0.587	NA	1.287	1.459
Long rolling window						
Inflation rate	0.448	0.399	0.483	0.883	0.872	0.452
Industrial production growth	1.738	1.569	1.888	1.706	2.056	1.529
Real GDP growth	0.877	0.617	1.306	1.069	1.058	0.882
Interest rate	1.334	1.198	0.536	NA	1.252	1.408
Post-break window						
Inflation rate	0.438	0.416	0.571	1.084	1.353	0.508
Industrial production growth	1.826	1.640	2.368	1.743	3.807	1.484
Real GDP growth	0.938	0.610	1.308	1.201	1.057	0.865
Interest rate	1.759	1.753	0.603	NA	1.536	1.556

Table 7: Out-of-sample forecasting performance (continued)

C. AR(4) Models	Canada	France	Germany	Japan	UK	US
Expanding window						
Inflation rate	0.447	0.411	0.472	0.863	0.925	0.437
Industrial production growth	1.721	1.626	1.731	1.806	2.174	1.512
Real GDP growth	0.924	0.532	1.056	1.034	1.104	0.898
Interest rate	1.379	1.287	0.468	NA	1.264	1.463
Short rolling window						
Inflation rate	0.499	0.464	0.562	0.907	0.980	0.456
Industrial production growth	1.897	1.740	1.867	1.790	2.388	1.718
Real GDP growth	0.963	0.585	1.224	1.061	1.209	0.984
Interest rate	1.544	1.488	0.523	NA	1.334	1.633
Long rolling window						
Inflation rate	0.474	0.424	0.524	0.880	0.966	0.434
Industrial production growth	1.771	1.620	1.823	1.757	2.243	1.644
Real GDP growth	0.895	0.565	1.062	1.029	1.136	0.930
Interest rate	1.431	1.288	0.472	NA	1.274	1.561
Post-break window						
Inflation rate	0.449	0.436	0.818	1.525	4.952	0.515
Industrial production growth	1.828	2.903	1.731	2.185	7.100	2.749
Real GDP growth	0.924	0.532	1.056	1.034	1.816	0.898
Interest rate	2.389	1.377	0.472	NA	2.155	3.494