

# Small Sample Properties of Forecasts from Autoregressive Models under Structural Breaks

*M. Hashem Pesaran and Allan Timmermann*

June 2003

DAE Working Paper No. 0331

**Not to be quoted without permission**

## ABSTRACT

Autoregressive models are used routinely in forecasting and often lead to better performance than more complicated models. However, empirical evidence is also suggesting that the autoregressive representations of many macroeconomic and financial time series are likely to be subject to structural breaks. This paper develops a theoretical framework for the analysis of small-sample properties of forecasts from general autoregressive models under a structural break. Our approach is quite general and allows for unit roots both pre- and post-break. We derive finite-sample results for the mean squared forecast error of one-step-ahead forecasts, both conditionally and unconditionally and present numerical results for different types of break specifications. Implications of breaks for the determination of the optimal window size are also discussed.

*JEL Classifications:* C22, C53.

*Key Words:* Small sample properties of forecasts, RMSFE, structural breaks, autoregression.

# Small Sample Properties of Forecasts from Autoregressive Models under Structural Breaks\*

M. Hashem Pesaran

Faculty of Economics and Politics

University of Cambridge

Allan Timmermann

University of California, San Diego

June 3, 2003

## 1 Introduction

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

Despite their relative 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), Garcia and Perron (1996) and Pesaran and Timmermann (2003a) are other examples of

---

\*We are grateful to seminar participants at Cass Business School for comments on a previous version of this paper. We would also like to thank Mutita Akusuwan for excellent research assistance.

studies that document instability related to the autoregressive terms in forecasting models used routinely throughout economics and finance. 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 this flurry of 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 typically be much smaller 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 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. 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).

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 both considers stationary AR models as well as models with unit root dynamics. 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 also extend Fuller (1996)'s result on the absence of a bias in the forecast in the presence of an intercept in the AR model to cover breaks in autoregressive coefficients. Third, we present extensive numerical results quantifying the effect of the size of the pre-break and post-break data window on parameter bias and

RMSFE.

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 concludes with 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  $\hat{\alpha}_T$ ,  $\hat{\beta}_T$ , the Ordinary Least Squares (OLS) estimators of  $\alpha$  and  $\beta$ , in the first-order autoregressive (AR(1)) model

$$y_t = \alpha + \beta y_{t-1} + \sigma \varepsilon_t, \quad t = 1, 2, \dots, T, \quad (1)$$

where  $\varepsilon_t \sim iid(0, 1)$ . Early analysis of the small sample bias of  $\hat{\beta}_T$  include Bartlett (1946), Hurwicz (1950), Kendall (1954), Marriott and Pope (1954) and White (1961). These studies focus on the case where  $a = 0$  and  $|\beta| < 1$ , namely a stationary AR(1) model without an intercept. Extensions to higher order models with intercepts have been proposed by Orcutt and Winoker (1969), Sawa (1978), Hoque (1985), and Bao and Ullah (2002). Hoque and Peters (1986), Grubb and Symons (1987), Kiviet and Phillips (1993, 2003a) further included exogenous regressors in the model and consider the so-called ARX(1) specification. Assuming stationarity ( $|\beta| < 1$ ),  $\hat{\beta}_T$  has been shown to have an asymptotic normal distribution and its finite-sample distribution has been further studied by Phillips (1977, 1978) and Evans and Savin (1981). The case with a unit root,  $\beta = 1$ , has been studied by, inter alia, Bannerjee, 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. Based on the sample observations,  $(y_0, y_1, \dots, y_T)$ , the one-step-ahead forecast of  $y_{T+1}$ ,  $\hat{y}_{T+1} = \hat{\alpha}_T + \hat{\beta}_T y_T$  and the associated forecast error,  $y_{T+1} - \hat{y}_{T+1}$ , have also received considerable atten-

tion. 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  $y_T$ , we have

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

which yields the more familiar unconditional result<sup>1</sup>

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

Generalizations to AR(p) and multi-step ahead 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 in fact unbiased in small samples assuming  $\varepsilon_t$  has a symmetric distribution and  $E(|\hat{y}_{T+1}|) < \infty$ . This is particularly interesting considering the often large small sample bias associated with the 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), 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 these pioneering literatures we now consider the small sample problem of estimation and forecasting with AR(p) models in the presence of structural

---

<sup>1</sup>Ullah (2003) provides an extensive discussion and survey of the properties of forecasts from the AR(1) model.

breaks. For this purpose, we consider the following AR(p) model defined over the period  $t = 1, 2, \dots, T$ ; and assumed to have been subject to a single structural break at the end of time  $T_1$  :

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

where 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,  $\alpha_i$ ,  $i = 1, 2$ :

$$\alpha_i = \mu_i(1 - \beta_i^*), \quad (3)$$

where  $-(1 - \beta_i^*)$  represents the coefficient of  $y_{t-1}$  in the error correction representation of (2). In particular

$$\beta_i^* = \sum_{j=1}^p \beta_{ij}. \quad (4)$$

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

$$\sum_{j=1}^p \lambda^j \beta_{ij} - 1 = 0, \text{ for } i = 1, 2, \quad (5)$$

lie on or outside the unit circle.<sup>2</sup> The intercepts  $\alpha_i = (1 - \beta_i^*)$  are therefore unrestricted when the underlying AR processes are stationary and are set to zero in the presence of unit roots to avoid the possibility of generating linear trends in the  $y_t$  process. In the stationary case  $\mu_i$  represents the unconditional mean of  $y_t$  under regime  $i$ . In the unit root case  $\mu_i$  is not identified and we have  $E(\Delta y_t) = 0$ .

Suppose  $y_t$  is observed over the period  $t = 1, 2, \dots, T$ , and the object of interest is the point (or probability) forecast of  $y_{T+1}$ , conditional on  $\mathbf{y}_T(1) = (y_1, y_2, \dots, y_T)'$ , and the above autoregressive specification, equation (2), subject to the regime switch at the end of time  $t = T_1$ . 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

---

<sup>2</sup>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.

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 exercise on pre-break as well as the post-break observations. The number of pre-break observations, which we denote by  $v_1$ , will become a choice parameter. In what follows we assume  $T_1$  is known but consider forecasting  $y_{T+1}$  using the observations  $\mathbf{y}_T(m-p) = (y_{m-p}, y_{m-p+1}, \dots, y_{T_1}, y_{T_1+1}, \dots, y_T)'$ , with  $y_{m-p}, y_{m-p+1}, \dots, y_{m-1}$  treated as given initial values.<sup>3</sup> The pre-break number of time periods 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  $v_1 \geq p$ .

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  $\mathbf{x}_T = (y_T, y_{T-1}, \dots, y_{T-p+1})'$ ,  $\hat{\boldsymbol{\beta}}_T(m) = (\hat{\beta}_{1T}(m), \hat{\beta}_{2T}(m), \dots, \hat{\beta}_{pT}(m))'$

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

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

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

$$\mathbf{M}_\tau = \mathbf{I}_v - \boldsymbol{\tau}_v (\boldsymbol{\tau}'_v \boldsymbol{\tau}_v)^{-1} \boldsymbol{\tau}'_v,$$

and  $\boldsymbol{\tau}_v = (1, 1, \dots, 1)'$ . The one-step ahead forecast error is

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

where

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

$\boldsymbol{\beta}_2 = (\beta_{21}, \beta_{22}, \dots, \beta_{2p})'$  and  $\alpha_2 = \mu_2 (1 - \beta_2)$ . The size of the forecast error varies with  $m$ , and the aim is to choose  $m$  such that  $E(e_{T+1}^2(m) | \mathbf{x}_T)$  is minimized:

$$m^* = \arg \min_{m=1, \dots, T_1+1} \{E[e_{T+1}^2(m) | \mathbf{x}_T]\}. \quad (10)$$

To this end we first note that  $\varepsilon_{T+1}$  and  $\xi_T(m)$  are independently distributed and

$$E(e_{T+1}^2(m) | \mathbf{x}_T) = \sigma_2^2 + E_\varepsilon(\xi_T^2(m) | \mathbf{x}_T).$$

---

<sup>3</sup>Throughout the paper we shall use the notation  $\mathbf{y}_T(k) = (y_k, \dots, y_T)'$ .



The expectations operator  $E_\varepsilon(\cdot)$  is defined with respect to the distribution of the innovations  $\varepsilon_t$ . Therefore, 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  $\hat{\beta}_T(m)$  and  $\hat{\alpha}_T(m)$  depends on the distribution of the initial values  $\mathbf{y}_{m-1}(m-p) = (y_{m-p}, y_{m-p+1}, \dots, y_{m-1})'$ . In the case where the pre-break regime is stationary the distribution of  $\mathbf{y}_{m-1}(m-p)$  is given by

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

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

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

A similar assumption concerning the initial values can be made if the pre-break process contains a unit root. However, the covariance of  $\mathbf{y}_{m-1}(m-p)$  is no longer given by  $\sigma_1^2 \mathbf{V}_p$ . In this case  $\beta_1^* = 1$  and the pre-break process is given by

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

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,  $(\Delta y_2, \Delta y_3, \dots, \Delta y_p)$ , using (13), and assuming that  $y_1$  is distributed as  $N(\bar{y}_0, \omega^2)$ , where  $\bar{y}_0$  and  $\omega$  are treated as free parameters. For example, in the AR(1) case

$$y_{m-1} \sim N[\bar{y}_0, \omega^2 + (m-2)\sigma_1^2]. \quad (14)$$

For the AR(2) case we first note that

$$\begin{aligned} y_{m-2} &= y_1 + \Delta y_2 + \dots + \Delta y_{m-2}, \\ y_{m-1} &= y_1 + \Delta y_2 + \dots + \Delta y_{m-2} + \Delta y_{m-1}, \end{aligned} \quad (15)$$

which provide a decomposition in terms of the non-stationary level component,  $y_1$ , and stationary first differences,  $\Delta y_2, \Delta y_3, \dots$ . Also using (13) we have

$$\Delta y_t = \delta_{11} \Delta y_{t-1} + \sigma_1 \varepsilon_t, \quad t = 2, 3, \dots, T_1$$

where  $|\delta_{11}| < 1$ , thus ensuring that  $y_t \sim I(1)$ . The distribution of  $(y_{m-2}, y_{m-1})$  can now be derived for given assumptions concerning  $y_1$  and  $\Delta y_1$ . There are many possibilities. Here as a simple example we consider the situation where as in the AR(1) case  $y_1 \sim N(\bar{y}_0, \omega^2)$ , distributed independently of  $\Delta y_t$ ,  $t = 2, 3, \dots$ , and assume that the stationary components of  $y_{m-2}$  and  $y_{m-1}$  are started with  $\Delta y_1 = 0$ . Under the latter we have

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

Substituting these in (15) we now have

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

Hence

$$\begin{aligned}Var(y_{m-1}) &= \omega^2 + \frac{\sigma_1^2 \sum_{j=1}^{m-2} (1 - \delta_{11}^j)^2}{(1 - \delta_{11})^2} \\ &= \omega^2 + \frac{\sigma_1^2 \left( (m-2)(1 - \delta_{11}^2) + \delta_{11}^2 (1 - \delta_{11}^{2(m-2)}) - 2\delta_{11}(1 + \delta_{11})(1 - \delta_{11}^{m-2}) \right)}{(1 - \delta_{11})^2 (1 - \delta_{11}^2)}, \\ Var(y_{m-2}) &= \omega^2 + \frac{\sigma_1^2 \left( (m-3)(1 - \delta_{11}^2) + \delta_{11}^2 (1 - \delta_{11}^{2(m-3)}) - 2\delta_{11}(1 + \delta_{11})(1 - \delta_{11}^{m-3}) \right)}{(1 - \delta_{11})^2 (1 - \delta_{11}^2)}, \\ Cov(y_{m-1}, y_{m-2}) &= \omega^2 + \frac{\sigma_1^2 \left( (m-3)(1 - \delta_{11}^2) + \delta_{11}^3 (1 - \delta_{11}^{2(m-3)}) - \delta_{11}(1 + \delta_{11})^2 (1 - \delta_{11}^{m-3}) \right)}{(1 - \delta_{11})^2 (1 - \delta_{11}^2)},\end{aligned}\tag{16}$$

so that

$$\begin{pmatrix} y_{m-2} \\ y_{m-1} \end{pmatrix} \sim N(\bar{y}_0 \boldsymbol{\tau}_2, \mathbf{V}_2),$$

where the elements in the  $2 \times 2$  matrix  $\mathbf{V}_2$  are given in (16).

Fixed (non-stochastic) starting values can also be accommodated by setting  $\omega = 0$ . In what follows we focus on the case where the pre-break regime is stationary

initialized stochastically according to the initial value distribution defined by (12), but allow the post-break regime to switch into (possibly) a process with a single unit root.

Using (12) in conjunction with (2) for  $t = m, m + 1, \dots, T$ , in matrix notations we have

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

where

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

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

$\mathbf{I}_{\nu_1}$  and  $\mathbf{I}_{\nu_2}$  are identity matrices of order  $\nu_1$  and  $\nu_2$ , respectively,

$$\boldsymbol{\varepsilon} = (\varepsilon_{m-p}, \varepsilon_{m-p+1}, \dots, \varepsilon_T)' \sim N(0, \mathbf{I}_{\nu+p}),$$

$\boldsymbol{\psi}_p$  is a lower triangular Cholesky factor of  $\mathbf{V}_p$ , namely  $\mathbf{V}_p = \boldsymbol{\psi}_p \boldsymbol{\psi}_p'$ , and the submatrices  $\mathbf{B}_{ij}$  are defined by

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

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

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

and

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

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

$$\mathbf{c} = \mathbf{B}^{-1}\mathbf{d}, \text{ and } \mathbf{H} = \mathbf{B}^{-1}\mathbf{D}.$$

It is now easily seen that

$$\mathbf{y}_{T-i}(m-i) = \mathbf{G}_i \mathbf{y}_T(m-p) = \mathbf{G}_i \mathbf{c} + \mathbf{G}_i \mathbf{H} \boldsymbol{\varepsilon}, \quad (17)$$

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

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

Note that  $\mathbf{0}_{v \times p-i}$  is a  $v \times (p-i)$  matrix of zeros. In particular  $\mathbf{G}_0 = (\mathbf{0}_{v \times p} : \mathbf{I}_\nu)$ , and  $\mathbf{G}_p = (\mathbf{I}_\nu : \mathbf{0}_{v \times p})$ .

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 in the model, i.e., if  $\mu_1 = \mu_2 = \mu$ .<sup>4</sup> First, it is easily verified that

$$\mathbf{B} \boldsymbol{\tau}_{v+p} = \begin{pmatrix} \boldsymbol{\tau}_p \\ (1 - \beta_1^*) \boldsymbol{\tau}_{v_1} \\ (1 - \beta_2^*) \boldsymbol{\tau}_{v_2} \end{pmatrix}. \quad (18a)$$

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 (18a)

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

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

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

and when  $\mu_1 = \mu_2 = \mu$  (or if  $\beta_2^* = 1$ ) we have (noting that  $\mathbf{G}_i \boldsymbol{\tau}_{v+p} = \boldsymbol{\tau}_v$ )

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

Therefore

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

Using these results the  $(i, j)$  element of the product moment matrix,  $\mathbf{X}'_T(m) \mathbf{M}_\tau \mathbf{X}_T(m)$ , is given by  $\boldsymbol{\varepsilon}' \mathbf{H}' \mathbf{G}'_i \mathbf{M}_\tau \mathbf{G}_j \mathbf{H} \boldsymbol{\varepsilon}$ , for  $i, j = 1, 2, \dots, p$ , and the  $j^{\text{th}}$  element of  $\mathbf{X}'_T(m) \mathbf{M}_\tau \mathbf{y}_T(\mathbf{m})$  is given by  $\boldsymbol{\varepsilon}' \mathbf{H}' \mathbf{G}'_j \mathbf{M}_\tau \mathbf{G}_0 \mathbf{H} \boldsymbol{\varepsilon}$ , for  $j = 1, 2, \dots, p$ . Hence,  $\hat{\boldsymbol{\beta}}_T(m)$  is a non-linear function of the quadratic forms  $\boldsymbol{\varepsilon}' \mathbf{H}' \mathbf{G}'_i \mathbf{M}_\tau \mathbf{G}_j \mathbf{H} \boldsymbol{\varepsilon}$ , for  $i = 1, 2, \dots, p$ , and  $j = 0, 1, \dots, p$ , with known matrices  $\mathbf{H}$  and  $\mathbf{G}_i$ , and  $\boldsymbol{\varepsilon} \sim N(0, \mathbf{I}_{\nu+p})$ . For example, for  $p = 1$  we

---

<sup>4</sup>Notice that a break in the slopes induces a change in the intercepts even if the mean of the process remains unchanged.

have<sup>5</sup>

$$\hat{\beta}_T(m) = \frac{\boldsymbol{\varepsilon}'\mathbf{H}'\mathbf{G}'_1\mathbf{M}_\tau\mathbf{G}_0\mathbf{H}\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}'\mathbf{H}'\mathbf{G}'_1\mathbf{M}_\tau\mathbf{G}_1\mathbf{H}\boldsymbol{\varepsilon}}, \quad (20)$$

and for  $p = 2$

$$\hat{\beta}_{1T}(m) = \frac{(\boldsymbol{\varepsilon}'\mathbf{H}'\mathbf{G}'_2\mathbf{M}_\tau\mathbf{G}_2\mathbf{H}\boldsymbol{\varepsilon})(\boldsymbol{\varepsilon}'\mathbf{H}'\mathbf{G}'_1\mathbf{M}_\tau\mathbf{G}_0\mathbf{H}\boldsymbol{\varepsilon}) - (\boldsymbol{\varepsilon}'\mathbf{H}'\mathbf{G}'_1\mathbf{M}_\tau\mathbf{G}_2\mathbf{H}\boldsymbol{\varepsilon})(\boldsymbol{\varepsilon}'\mathbf{H}'\mathbf{G}'_2\mathbf{M}_\tau\mathbf{G}_0\mathbf{H}\boldsymbol{\varepsilon})}{(\boldsymbol{\varepsilon}'\mathbf{H}'\mathbf{G}'_1\mathbf{M}_\tau\mathbf{G}_1\mathbf{H}\boldsymbol{\varepsilon})(\boldsymbol{\varepsilon}'\mathbf{H}'\mathbf{G}'_2\mathbf{M}_\tau\mathbf{G}_2\mathbf{H}\boldsymbol{\varepsilon}) - (\boldsymbol{\varepsilon}'\mathbf{H}'\mathbf{G}'_1\mathbf{M}_\tau\mathbf{G}_2\mathbf{H}\boldsymbol{\varepsilon})^2}, \quad (21)$$

and

$$\hat{\beta}_{2T}(m) = \frac{(\boldsymbol{\varepsilon}'\mathbf{H}'\mathbf{G}'_1\mathbf{M}_\tau\mathbf{G}_1\mathbf{H}\boldsymbol{\varepsilon})(\boldsymbol{\varepsilon}'\mathbf{H}'\mathbf{G}'_2\mathbf{M}_\tau\mathbf{G}_0\mathbf{H}\boldsymbol{\varepsilon}) - (\boldsymbol{\varepsilon}'\mathbf{H}'\mathbf{G}'_2\mathbf{M}_\tau\mathbf{G}_1\mathbf{H}\boldsymbol{\varepsilon})(\boldsymbol{\varepsilon}'\mathbf{H}'\mathbf{G}'_1\mathbf{M}_\tau\mathbf{G}_0\mathbf{H}\boldsymbol{\varepsilon})}{(\boldsymbol{\varepsilon}'\mathbf{H}'\mathbf{G}'_1\mathbf{M}_\tau\mathbf{G}_1\mathbf{H}\boldsymbol{\varepsilon})(\boldsymbol{\varepsilon}'\mathbf{H}'\mathbf{G}'_2\mathbf{M}_\tau\mathbf{G}_2\mathbf{H}\boldsymbol{\varepsilon}) - (\boldsymbol{\varepsilon}'\mathbf{H}'\mathbf{G}'_1\mathbf{M}_\tau\mathbf{G}_2\mathbf{H}\boldsymbol{\varepsilon})^2}. \quad (22)$$

From the above results it is now clear that (i) the probability distribution of  $\hat{\beta}_T(m)$  will depend only on the ratio of the error variances,  $\sigma_1^2/\sigma_2^2$ , and does not depend on their scale. Therefore, in the case where  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ , the distribution of  $\hat{\beta}_T(m)$  will be invariant to  $\sigma^2$ . (ii) In the absence of a mean shift or if  $\beta_1^* < 1$  but  $\beta_2^* = 1$ ,  $\hat{\beta}_T(m)$  will not depend on the unconditional mean(s) of the underlying autoregressive process. (iii) Finally, in the case where  $\mu_1 = \mu_2$ ,  $\hat{\beta}_T(m)$  is an even function of  $\boldsymbol{\varepsilon}$  and this property is unaffected by whether the slope coefficients and/or the error variances are subject to structural breaks.

Consider now the distribution of  $\hat{\alpha}_T(m)$  given by (7), and to simplify the exposition assume that  $\mu_1 = \mu_2$  (or  $\beta_2^* = 1$ ). Using (19) we first note that

$$\boldsymbol{\tau}'_v\mathbf{y}_T(m) = v\mu + \boldsymbol{\tau}'_v\mathbf{G}_0\mathbf{H}\boldsymbol{\varepsilon},$$

Similarly

$$\boldsymbol{\tau}'_v\mathbf{X}_T(m)\hat{\beta}_T(m) = \sum_{j=1}^p \hat{\beta}_{jT}(m) [v\mu + \boldsymbol{\tau}'_v\mathbf{G}_j\mathbf{H}\boldsymbol{\varepsilon}].$$

Using these results we have

$$\hat{\alpha}_T(m) = \mu \left(1 - \hat{\beta}_T^*(m)\right) + v^{-1} \left[ (\boldsymbol{\tau}'_v\mathbf{G}_0\mathbf{H}\boldsymbol{\varepsilon}) - \sum_{j=1}^p (\boldsymbol{\tau}'_v\mathbf{G}_j\mathbf{H}\boldsymbol{\varepsilon}) \hat{\beta}_{jT}(m) \right], \quad (23)$$

---

<sup>5</sup>Allowing for a mean shift and setting  $p = 1$ , the expressions for  $\hat{\beta}_T(m)$  and  $\hat{\alpha}_T(m)$  become

$$\begin{aligned} \hat{\beta}_T(m) &= \frac{(\boldsymbol{\varepsilon}'\mathbf{H}' + \mathbf{c}')\mathbf{G}'_1\mathbf{M}_\tau\mathbf{G}_0(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon})}{(\boldsymbol{\varepsilon}'\mathbf{H}' + \mathbf{c}')\mathbf{G}'_1\mathbf{M}_\tau\mathbf{G}_1(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon})}, \\ \hat{\alpha}_T(m) &= v^{-1}\boldsymbol{\tau}'_v[\mathbf{G}_0(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon}) - \mathbf{G}_1(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon})\hat{\beta}_T(m)]. \end{aligned}$$

where  $\hat{\beta}_T^*(m) = \sum_{j=1}^p \hat{\beta}_{jT}(m)$ . Therefore, unlike  $\hat{\beta}_T(m)$  and perhaps not surprisingly the distribution of  $\hat{\alpha}_T(m)$  does depend on the unconditional mean of the process even under  $\mu_1 = \mu_2 = \mu$ .

### 3.1 An Unconditional Expression for $e_{T+1}^2(m)$

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

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

Therefore, using (19) we have (noting that  $\mathbf{S}_p \boldsymbol{\tau}_v = \boldsymbol{\tau}_p$ )

$$\mathbf{x}_T = \mathbf{S}_p \mathbf{y}_T(m) = \mathbf{S}_p (\mu \boldsymbol{\tau}_v + \mathbf{G}_0 \mathbf{H} \boldsymbol{\varepsilon}),$$

or

$$\mathbf{x}_T - \mu \boldsymbol{\tau}_p = \mathbf{S}_p \mathbf{G}_0 \mathbf{H} \boldsymbol{\varepsilon}.$$

Hence unconditionally we have

$$\xi_T(m) = \left( \hat{\beta}_T(m) - \beta_2 \right)' \mathbf{S}_p \mathbf{G}_0 \mathbf{H} \boldsymbol{\varepsilon} + v^{-1} \left[ (\boldsymbol{\tau}'_v \mathbf{G}_0 \mathbf{H} \boldsymbol{\varepsilon}) - \sum_{j=1}^p (\boldsymbol{\tau}'_v \mathbf{G}_j \mathbf{H} \boldsymbol{\varepsilon}) \hat{\beta}_{jT}(m) \right]. \quad (24)$$

Since under  $\mu_1 = \mu_2 = \mu$  (or  $\beta_2 = 1$ ),  $\hat{\beta}_T(m)$  does not depend on  $\alpha$  (or  $\alpha_1$ ) it also follows that unconditionally  $\xi_T(m)$ , and hence  $E[e_{T+1}^2(m)]$ , will not depend on the unconditional mean of the autoregressive process.

In the case with a break in the mean,  $\mu_1 \neq \mu_2$ , we have

$$\begin{aligned} \xi_T(m) &= \left( \hat{\beta}_T(m) - \beta_2 \right)' \mathbf{S}_p \mathbf{G}_0 (\mathbf{c} + \mathbf{H} \boldsymbol{\varepsilon}) \\ &+ v^{-1} \left[ \boldsymbol{\tau}'_v \mathbf{G}_0 (\mathbf{c} + \mathbf{H} \boldsymbol{\varepsilon}) - \sum_{j=1}^p \boldsymbol{\tau}'_v \mathbf{G}_j (\mathbf{c} + \mathbf{H} \boldsymbol{\varepsilon}) \hat{\beta}_{jT}(m) \right] - \mu_2. \end{aligned} \quad (25)$$

The computation of  $E[e_{T+1}^2(m)]$  can be carried out by stochastic simulations. We have

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

where  $\xi_T^{(r)}(m)$  is given by

$$\begin{aligned} \xi_T^{(r)}(m) &= \left( \hat{\beta}_T^{(r)}(m) - \beta_2 \right)' \mathbf{S}_p \mathbf{G}_0 \mathbf{H} \boldsymbol{\varepsilon}^{(r)} \\ &\quad + v^{-1} \left[ \left( \boldsymbol{\tau}'_v \mathbf{G}_0 \mathbf{H} \boldsymbol{\varepsilon}^{(r)} \right) - \sum_{j=1}^p \left( \boldsymbol{\tau}'_v \mathbf{G}_j \mathbf{H} \boldsymbol{\varepsilon}^{(r)} \right) \hat{\beta}_{jT}^{(r)}(m) \right]. \end{aligned}$$

Due to the independence of  $\boldsymbol{\varepsilon}^{(r)}$  across  $r$ , and the fact that  $\xi_T^{(r)}(m)$  are independently and identically distributed across  $r$ , by a standard law of large numbers we have

$$\hat{E}_R [e_{T+1}^2(m)] \xrightarrow{p} E [e_{T+1}^2(m)],$$

so long as  $E [e_{T+1}^2(m)]$  does in fact exist. This clearly requires restrictions on the size of  $v$ , the estimation window. We can provide exact conditions in the simple case where  $p = 1$ , but there appears to be no results for the existence of moment conditions in the more general case of  $p \geq 2$ .

The following proposition generalizes Theorem 8.5.2 in Fuller (1996, page 445) to the case with a break in the slope coefficient:

*Proposition: The unconditional expectation of the forecast errors from the AR( $p$ ) model (2) subject to a break in the AR coefficients ( $\beta_1 \neq \beta_2$ ) 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;*
- (ii)  *$E |e_{T+1}(m)| < \infty$ ; and*
- (iii) *there is no break in the mean of the process,  $\mu_1 = \mu_2$ .*

Proof: It has already been established that when  $\mu_1 = \mu_2$ ,  $\hat{\beta}_T(m)$  can be written as a non-linear function of quadratic forms in  $\boldsymbol{\varepsilon}$ , and is therefore an even function of  $\boldsymbol{\varepsilon}$ . In the case where  $\mu_1 = \mu_2$ , using (8) and (24), the prediction error can be written as

$$\begin{aligned} e_{T+1}(m) &= \sigma_2 \varepsilon_{T+1} - \left( \hat{\beta}_T(m) - \beta_2 \right)' \mathbf{S}_p \mathbf{G}_0 \mathbf{H} \boldsymbol{\varepsilon} \\ &\quad - v^{-1} \left[ \left( \boldsymbol{\tau}'_v \mathbf{G}_0 \mathbf{H} \boldsymbol{\varepsilon} \right) - \sum_{j=1}^p \left( \boldsymbol{\tau}'_v \mathbf{G}_j \mathbf{H} \boldsymbol{\varepsilon} \right) \hat{\beta}_{jT}(m) \right]. \end{aligned}$$

Hence

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



Since  $\hat{\beta}_{jT}(m)$ ,  $j = 1, 2, \dots, p$  are even functions of  $\varepsilon$ , and  $(\tau'_v \mathbf{G}_j \mathbf{H} \varepsilon)$  and  $\hat{\beta}'_T(m) \mathbf{S}_p \mathbf{G}_0 \mathbf{H} \varepsilon$  are odd functions of  $\varepsilon$  all the terms inside the above expectations are odd functions of  $e_{T+1}(m)$  and (by assumptions (i) and (ii)) exist, we must have

$$E[e_{T+1}(m)] = 0.$$

In the case where  $\mu_1 \neq \mu_2$  the expression for the prediction error is given by (25) and is no longer an odd function of  $\varepsilon$ , so it will, in general, not have mean zero.

One important implication of the above result is 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  $\mu_1 = \mu_2$ . To see this using (23) we have

$$E[\hat{\alpha}_T(m)] = \mu \left\{ 1 - E[\hat{\beta}_T^*(m)] \right\},$$

which can equivalently be written as

$$E[\hat{\alpha}_T(m) - \mu(1 - \beta_2^*)] = -\mu E[\hat{\beta}_T^*(m) - \beta_2^*].$$

This provides an interesting relationship between the small sample bias of the estimator of the intercept term,  $E[\hat{\alpha}_T(m) - \mu(1 - \beta_2^*)]$ , and the small sample bias of the long-run coefficient,  $E[\hat{\beta}_T^*(m) - \beta_2^*]$ . The estimator of the intercept term,  $\hat{\alpha}_T(m)$ , is unbiased only if the sample mean is zero. But, in general there is an spill-over effect from the bias of the slope coefficient to that of the intercept term.

In the case of the AR(1) model the results simplify further and we have

$$E[\hat{\alpha}_T(m) - \alpha_2] = -\mu E[\hat{\beta}_T(m) - \beta_2]. \quad (26)$$

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

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

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

### 3.2 A Conditional Expression for $e_{T+1}^2(m)$

As before we have

$$e_{T+1}(m) = \sigma_2 \varepsilon_{T+1} - \xi_T(m).$$

Using (23) in (9) and after some algebra we have (under  $\mu_1 = \mu_2 = \mu$ )<sup>6</sup>

$$\xi_T(m) = (\mathbf{x}_T - \mu\boldsymbol{\tau}_p)' \left( \hat{\boldsymbol{\beta}}_T(m) - \boldsymbol{\beta}_2 \right) + v^{-1} \left[ (\boldsymbol{\tau}'_v \mathbf{G}_0 \mathbf{H} \boldsymbol{\varepsilon}) - \sum_{j=1}^p (\boldsymbol{\tau}'_v \mathbf{G}_j \mathbf{H} \boldsymbol{\varepsilon}) \hat{\beta}_{jT}(m) \right]. \quad (27)$$

Suppose  $p = 1$  so that it is easy to characterize when  $\mathbf{x}_T$  is above or below the mean and assume that the distribution of  $\boldsymbol{\varepsilon}$  is symmetric. Then

$$E[e_{T+1}(m) | y_T] = (y_T - \mu) E \left( \hat{\beta}_T(m) - \beta_2 \right).$$

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

$$E[e_{T+1}(m) | y_T] = \begin{cases} < 0 & \text{if } y_T > \mu \\ \geq 0 & \text{if } y_T \leq \mu \end{cases},$$

and the estimated model underpredicts if the last observation is above the unconditional mean ( $y_T > \mu$ ), while conversely it overpredicts if the last observation is below the unconditional mean ( $y_T < \mu$ ). In general we have

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

To compute conditional MSFE, we note that conditional on  $\mathbf{x}_T$ ,  $\xi_T(m)$  depends on  $\mu, \sigma_1, \sigma_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2$  and can be simulated for given values of  $\mu, \sigma_1, \sigma_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ , and  $\mathbf{x}_T$  and draws from  $\boldsymbol{\varepsilon} \sim N(0, \mathbf{I}_{\nu+p})$ . Denoting the  $r^{\text{th}}$  draw of  $\boldsymbol{\varepsilon}$  by  $\boldsymbol{\varepsilon}^{(r)}$ ,  $r = 1, 2, \dots, R$ , the conditional MSFE can be computed by

$$\hat{E}_R(e_{T+1}^2(m) | \mathbf{x}_T) = \sigma_2^2 + \hat{E}_R(\xi_T^2(m) | \mathbf{x}_T),$$

where

$$\begin{aligned} \hat{E}_R(\xi_T^2(m) | \mathbf{x}_T) &= \frac{1}{R} \sum_{r=1}^R \left[ \xi_T^{(r)}(m) \right]^2, \\ \xi_T^{(r)}(m) &= (\mathbf{x}_T - \mu\boldsymbol{\tau}_p)' \left( \hat{\boldsymbol{\beta}}_T^{(r)}(m) - \boldsymbol{\beta}_2 \right) \\ &\quad + v^{-1} \left[ (\boldsymbol{\tau}'_v \mathbf{G}_0 \mathbf{H} \boldsymbol{\varepsilon}^{(r)}) - \sum_{j=1}^p (\boldsymbol{\tau}'_v \mathbf{G}_j \mathbf{H} \boldsymbol{\varepsilon}^{(r)}) \hat{\beta}_{jT}^{(r)}(m) \right], \end{aligned}$$

and  $\hat{\boldsymbol{\beta}}_T^{(r)}(m)$  is given by expressions such as (20), or (21) and (22) with  $\boldsymbol{\varepsilon}$ , replaced by  $\boldsymbol{\varepsilon}^{(r)}$ .

---

<sup>6</sup> Allowing for a break in the intercept, the corresponding expression becomes

$$\xi_T(m) = \mathbf{x}'_T \left( \hat{\boldsymbol{\beta}}_T(m) - \boldsymbol{\beta}_2 \right)' + v^{-1} \boldsymbol{\tau}'_v \left[ \mathbf{G}_0(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon}) - \sum_{j=1}^p \mathbf{G}_j(\mathbf{c} + \mathbf{H}\boldsymbol{\varepsilon}) \hat{\beta}_{jT}(m) \right] - \mu_2(1 - \beta_2^*).$$

## 4 Numerical Results

Our approach is quite general and allows us to study the small sample properties of the AR model 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,  $\hat{\alpha}$  and  $\hat{\beta}$ , are the sample size and the persistence parameter,  $\beta_1$ . 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,  $\alpha$ , and the innovation variance,  $\sigma^2$ . Under a break, we need to consider both the pre- and post-break parameters - i.e. the AR coefficients  $(\beta_1, \beta_2)$ , the intercepts  $(\alpha_1, \alpha_2)$  and the innovation variances  $(\sigma_1^2, \sigma_2^2)$ . 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 1a in Table 1a). We next introduce breaks in this model and extend the results to cover the AR(2) specification. This allows us to consider the effect of higher order dynamics. In all cases we present results for 50,000 Monte Carlo simulations and draw innovations from an IID Gaussian distribution. We study breaks in the autoregressive parameters in the form of both moderately sized (0.3) and large (0.6) breaks in either direction (experiments 2a-4a) as well as a shift to a unit root process in the post-break period (experiment 5a). We also consider pure breaks in the innovation variance (experiments 6a and 7a), where  $\sigma$  changes between values of 1/2 and 2, and in the mean (experiments 8a and 9a), where  $\mu$  changes between 1 and 2. For convenience the parameter values assumed in each of the experiments are summarized in Table 1a. 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.

### 4.1 Results for the AR(1) Model

Results for the AR(1) model are presented in Tables 2-5. As a measure of bias in the parameter estimates and in the resulting forecast, 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.,  $y_T = \alpha_2 + \sigma_2$ .<sup>7</sup> To measure forecasting performance, Table 4 reports the unconditional RMSFE while Table 5 shows the conditional RMSFE when  $y_T = \alpha_2 + \sigma_2$  as a function of the pre-break ( $v_1$ ) and post-break window size ( $v_2$ ).

#### 4.1.1 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 large.<sup>8</sup> Notice the symmetry of the results in  $v_1$  and  $v_2$  which follows since (under no break) only  $v_1 + v_2$  matters to the bias.<sup>9</sup> 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 hence 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 4a), 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  $\hat{\beta}_1$  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 3a) so that the true value of  $\beta_1$  declines, including a small number of pre-break data points

---

<sup>7</sup>Values are computed as averages across Monte Carlo simulations relative to the true post-break values.

<sup>8</sup>The bias estimates are in line with the well known Kendall (1954) approximation given by

$$\hat{\beta}_1 - \beta_1 = \frac{-(1 + 3\beta_1)}{v} + O(v^{-3/2}),$$

where  $v = v_1 + v_2$ .

<sup>9</sup>Recall from (26) 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  $\hat{\beta}_T(m)$ .

leads to a reduction in the bias in  $\hat{\beta}_1$  even for the very large post-break windows considered here. 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  $\hat{\beta}_1$ . As  $\hat{\beta}_1$  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  $\beta_1$  tend to be smaller than the biases observed when  $\beta_1$  increases at the time of the break.

Under a post-break unit root (experiment 5a) the bias-minimizing pre-break window size is quite constant around 20 observations. When a break occurs in the innovation variance (experiments 6a and 7a), the smallest bias is always achieved by the longest pre- and post-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 in either direction (experiments 8a and 9a), the forecast error is no longer unbiased unconditionally and the optimal window size rises to 100 in both experiments 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 negative when the prediction is made conditional on a value above the mean of the process,  $y_T = \alpha_2 + \sigma_2$ . This is, of course, consistent 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 in the case with breaks 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) or for a pre-break window size around 10-20 observations under a decrease in the mean.

#### 4.1.2 Forecasting Performance

To measure forecasting performance under 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 again 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 the pre-break window,  $v_1$ . Furthermore, the unconditional RMSFE no longer converges to one -

its theoretical value in the absence of parameter estimation uncertainty - whenever the pre-break window,  $v_1$ , measured as a proportion of the post-break window,  $v_2$ , does not go to zero. For example, when  $v_1 = v_2 = 100$ , the unconditional RMSFE under a moderate break in  $\beta_1$  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 smaller 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.

The presence of a moderately sized break in  $\beta_1$  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 higher the value of  $v_2$  and the larger the size of the break in  $\beta_1$  as can be seen by comparing experiments 2a and 4a.

Somewhat different patterns emerge under a post-break unit root where the conditional RMSFE is minimized for the largest values of  $v_1$ , whereas the unconditional RMSFE is minimized at much smaller values of  $v_1$ , typically below 10 observations.

When the innovation variance is higher post-break, it is optimal to set the pre-break window as large 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  $v_2$ . Under breaks to the mean, the lowest conditional and unconditional RMSFE values are observed for the longer pre-break windows. This is an interesting finding. When breaks occur in practice, they are likely to affect the mean. 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.

## 4.2 Results for the AR(2) Model

Results for the AR(2) model are presented in Tables 6-10, while Panel b in Table 1 shows the parameter configurations used in the experiments labelled 1b to 9b.

It is no longer so straightforward to summarize the results as there are now more parameters. With two AR parameters we report the bias separately for  $\hat{\beta}_1$  and for  $\hat{\beta}^* = \hat{\beta}_1 + \hat{\beta}_2$ , the latter providing a measure of persistence. Our baseline scenario assumes autoregressive roots of 0.95 and -0.20 so that  $\beta_1 = 0.75$  and  $\beta_2 = -0.19$ . A large break occurs when these coefficients shift to 0.45 and 0, while a moderate break assumes that  $\beta_1 = 0.65, \beta_2 = -0.29$ . Breaks to the mean or to the innovation variance are identical to those assumed for the AR(1) model.

#### 4.2.1 Bias in AR coefficients

Many findings are qualitatively similar to those reported above for the AR(1) model so we simply summarize the main results here. Table 6 shows that under no break or a break in  $\sigma$  (experiments 1b, 6b and 7b) the larger the pre-break window,  $v_1$ , the smaller the bias in  $\hat{\beta}_1$ . A moderate increase in  $\beta_1$  (experiment 4b) now means that the smallest bias in  $\hat{\beta}_1$  is observed when no pre-break observations are included, i.e. for  $v_1 = 0$ , while under a moderate or large decline in  $\beta$  the bias is generally smallest for small (but non-zero) pre-break windows. Under a break in the mean, the smallest bias in  $\hat{\beta}_1$  is observed for  $v_1 = 100$  only when  $v_2 = 10$ , while the optimal value of the pre-break window,  $v_1$ , is around 10-20 observations and declines for larger values of  $v_2$ .

Turning to the bias in  $\hat{\beta}^*$ , Table 7 shows that the results are qualitatively very similar to those reported for the AR(1) case. The main exceptions are that the bias-minimizing value of the pre-break window,  $v_1$ , is now quite large even under the large break in  $\beta$ . Conversely, it is no longer the case that the largest value of  $v_1$  minimizes the bias in  $\hat{\beta}^*$  in the case with a break in the mean.

Table 8 shows that the conditional forecast error bias continues to be negative in the absence of a break or under a break in the innovation variance. For these experiments, the forecast error bias is generally smaller the larger the value of the pre-break window,  $v_1$ . Upon introducing breaks, the bias-minimizing pre-break window continues to decline as  $v_2$  gets larger and in some cases (experiments 2b and 3b), equals zero when the post-break window,  $v_2 \geq 50$ .

#### 4.2.2 Forecasting performance

The forecasting performance results reported in Tables 9 and 10 are qualitatively similar to those obtained for the AR(1) model. Long pre-break windows,  $v_1$ , are generally optimal in the absence of a break, under higher post-break volatility and

for increases as well as decreases in  $\beta_1$  at the time of the break. Smaller pre-break windows minimize the RMSFE as  $v_2$  gets larger. A smaller pre-break window continues to be called for as the size of the break increases irrespective of whether the break occurs in the AR coefficients or in the mean. However, the optimal window length continues to be quite long in many experiments even when  $v_2$  is large.

## 5 Conclusion

This paper studied the small sample properties of forecasts from autoregressive models subject to breaks. It is insightful to compare our results for the AR(p) model to those reported derived by Pesaran and Timmermann (2003b) 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 optimal 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 decline as a result of using pre-break data in the estimation. This is likely to be an important reason why, empirically, it is often quite difficult to improve forecasting performance over the expanding window method by only using post-break data.

More generally, we find 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.

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 MSE 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 the innovations



from a mixture of normals. Another possibility would be to consider the effect of adding additional predictors beyond autoregressive lags as well as extensions to cases with multiple breaks.

## 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 Change-point 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.
- Bannerjee, 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.
- Bao, Y. and A. Ullah, 2002, The second-order bias and mean squared error of non-linear estimators in time series. Manuscript, University of California, Riverside.
- 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.
- 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.
- Hoque, A., 1985, The Exact moments of Forecast error in the general dynamic models. *Sankhya* 47, 128-143.
- Hoque, A., and T. Peters, 1986, Finite sample analysis of the ARMAX models. *Sankhya* 48, 266-283.
- Hurwicz, L., 1950, Least squares bias in time series. In: Koopmans, T. (ed) *Statistical inference in dynamic economic models*. Wiley, New York, 365-383.

- 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.
- Kendall, M.G., 1954, Note on bias in the estimation of autocorrelation. *Biometrika* 61, 403-404.
- 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.
- Marriott, F.H.C., Pope, J.A., 1954. Bias in the estimation of autocorrelations. *Biometrika* 61, 393-403.
- Orcutt, G.H. and H.S. Winokur, 1969, First order autoregression: Inference, estimation and prediction. *Econometrica* 37, 1-14.
- Pesaran, M.H. and A. Timmermann, 2003a, Model instability and choice of observation window. Unpublished manuscript University of Cambridge and UCSD.
- Pesaran, M.H. and A. Timmermann, 2003b, 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., 1978, Edgeworth and saddlepoint approximations in the first-order non circular autoregression. *Biometrika* 65, 91-98.
- 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.
- 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.
- Ullah, A., 2003, Finite sample econometrics. Manuscript, University of California, Riverside.
- White, J.S., 1961, Asymptotic expansions for the mean and variance of the serial correlation coefficient. *Biometrika* 48, 85-94.

**Table 1a: Breakpoint Specifications by Experiments**  
**AR(1) Model**

Experiments	$\mu_1$	$\mu_2$	$\beta_{11}$	$\beta_{12}$	$\sigma_1$	$\sigma_2$
1a: No break	1	1	0.9	0.9	1	1
2a: Moderate break in $\beta_1$	1	1	0.6	0.9	1	1
3a: Moderate break in $\beta_1$ (decline)	1	1	0.9	0.6	1	1
4a: Large break in $\beta_1$	1	1	0.3	0.9	1	1
5a: Post-break unit root	1	1	0.6	1	1	1
6a: Higher post-break volatility	1	1	0.9	0.9	0.5	2
7a: Lower post-break volatility	1	1	0.9	0.9	2	0.5
8a: Break in mean (increase)	1	2	0.9	0.9	1	1
9a: Break in mean (decrease)	2	1	0.9	0.9	1	1

**Table 1b: AR(2) Model**

Experiments	$\mu_1$	$\mu_2$	$\beta_{11}$	$\beta_{12}$	$\beta_{21}$	$\beta_{22}$	$\sigma_1$	$\sigma_2$
1b: No break	1	1	0.75	-0.19	0.75	-0.19	1	1
2b: Moderate break in $\beta$	1	1	0.65	-0.29	0.75	-0.19	1	1
3b: Large increase in $\beta_1$	1	1	0.45	0.00	0.75	-0.19	1	1
4b: Large decrease in $\beta_1$	1	1	0.75	-0.19	0.45	0.00	1	1
5b: Post-break unit root	1	1	0.60	0.20	0.80	0.20	1	1
6b: Higher post-break volatility	1	1	0.75	-0.19	0.75	-0.19	0.5	2
7b: Lower post-break volatility	1	1	0.75	-0.19	0.75	-0.19	2	0.5
8b: Break in mean (increase)	1	2	0.75	-0.19	0.75	-0.19	1	1
9b: Break in mean (decrease)	2	1	0.75	-0.19	0.75	-0.19	1	1

**Table 2: Small sample bias of the OLS estimate of  $\beta$  as a function of pre-break ( $v_1$ ) and post-break ( $v_2$ ) windows - (AR(1) model).**

Experiment no. 1a : No break						Experiment no. 4a : Large break in $\beta$						Experiment no. 7a : Lower post-break volatility					
$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100
0	-0.370	-0.200	-0.135	-0.081	-0.039	0	-0.396	-0.214	-0.145	-0.086	-0.041	0	-0.224	-0.118	-0.080	-0.052	-0.028
1	-0.344	-0.193	-0.132	-0.080	-0.039	1	-0.386	-0.221	-0.150	-0.089	-0.042	1	-0.326	-0.176	-0.118	-0.072	-0.037
2	-0.315	-0.184	-0.129	-0.078	-0.039	2	-0.371	-0.218	-0.151	-0.091	-0.044	2	-0.338	-0.188	-0.134	-0.082	-0.042
3	-0.297	-0.175	-0.125	-0.076	-0.038	3	-0.362	-0.219	-0.151	-0.093	-0.045	3	-0.330	-0.190	-0.134	-0.086	-0.046
4	-0.278	-0.170	-0.119	-0.076	-0.038	4	-0.358	-0.221	-0.154	-0.095	-0.046	4	-0.316	-0.188	-0.135	-0.089	-0.048
5	-0.262	-0.162	-0.116	-0.074	-0.037	5	-0.357	-0.222	-0.156	-0.096	-0.048	5	-0.299	-0.185	-0.135	-0.091	-0.050
10	-0.202	-0.136	-0.102	-0.068	-0.035	10	-0.363	-0.233	-0.170	-0.107	-0.054	10	-0.226	-0.157	-0.121	-0.087	-0.051
20	-0.136	-0.102	-0.081	-0.057	-0.032	20	-0.392	-0.262	-0.195	-0.127	-0.066	20	-0.150	-0.116	-0.095	-0.073	-0.048
30	-0.102	-0.082	-0.066	-0.050	-0.030	30	-0.418	-0.292	-0.220	-0.147	-0.078	30	-0.110	-0.090	-0.078	-0.061	-0.043
50	-0.067	-0.058	-0.049	-0.040	-0.026	50	-0.453	-0.334	-0.262	-0.181	-0.100	50	-0.070	-0.063	-0.057	-0.048	-0.036
100	-0.036	-0.033	-0.030	-0.026	-0.020	100	-0.499	-0.401	-0.332	-0.246	-0.145	100	-0.037	-0.035	-0.033	-0.029	-0.024
Experiment no. 2a : Moderate break in $\beta$						Experiment no. 5a : Post-break unit root						Experiment no. 8a : Break in mean (increase)					
$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100
0	-0.392	-0.214	-0.144	-0.085	-0.041	0	-0.413	-0.233	-0.163	-0.101	-0.052	0	-0.366	-0.198	-0.134	-0.080	-0.038
1	-0.375	-0.214	-0.146	-0.087	-0.042	1	-0.391	-0.227	-0.159	-0.100	-0.052	1	-0.333	-0.189	-0.131	-0.078	-0.039
2	-0.354	-0.209	-0.144	-0.087	-0.043	2	-0.367	-0.220	-0.157	-0.098	-0.051	2	-0.311	-0.177	-0.126	-0.075	-0.038
3	-0.337	-0.204	-0.144	-0.088	-0.044	3	-0.351	-0.211	-0.152	-0.096	-0.051	3	-0.289	-0.169	-0.120	-0.075	-0.037
4	-0.324	-0.200	-0.145	-0.089	-0.044	4	-0.338	-0.207	-0.149	-0.095	-0.051	4	-0.270	-0.164	-0.115	-0.071	-0.037
5	-0.317	-0.200	-0.143	-0.091	-0.045	5	-0.325	-0.199	-0.146	-0.094	-0.050	5	-0.257	-0.156	-0.113	-0.071	-0.036
10	-0.286	-0.190	-0.144	-0.094	-0.048	10	-0.303	-0.190	-0.138	-0.089	-0.048	10	-0.196	-0.130	-0.097	-0.063	-0.033
20	-0.269	-0.191	-0.149	-0.102	-0.055	20	-0.295	-0.190	-0.138	-0.088	-0.046	20	-0.133	-0.097	-0.076	-0.053	-0.029
30	-0.268	-0.200	-0.158	-0.111	-0.062	30	-0.301	-0.197	-0.143	-0.091	-0.047	30	-0.100	-0.077	-0.062	-0.045	-0.027
50	-0.269	-0.213	-0.174	-0.126	-0.074	50	-0.313	-0.215	-0.158	-0.098	-0.048	50	-0.066	-0.054	-0.045	-0.034	-0.022
100	-0.276	-0.233	-0.200	-0.156	-0.100	100	-0.337	-0.247	-0.188	-0.119	-0.056	100	-0.035	-0.031	-0.028	-0.022	-0.015
Experiment no. 3a : Break in $\beta$ (decline)						Experiment no. 6a : Higher post-break volatility						Experiment no. 9a : Break in mean (decrease)					
$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100
0	-0.226	-0.125	-0.087	-0.053	-0.028	0	-0.399	-0.219	-0.147	-0.087	-0.041	0	-0.365	-0.197	-0.134	-0.081	-0.039
1	-0.156	-0.084	-0.057	-0.036	-0.017	1	-0.340	-0.202	-0.140	-0.083	-0.041	1	-0.334	-0.187	-0.128	-0.079	-0.038
2	-0.113	-0.057	-0.036	-0.022	-0.010	2	-0.301	-0.188	-0.134	-0.083	-0.041	2	-0.312	-0.178	-0.123	-0.075	-0.038
3	-0.084	-0.037	-0.021	-0.010	-0.003	3	-0.278	-0.177	-0.129	-0.080	-0.040	3	-0.289	-0.171	-0.120	-0.074	-0.037
4	-0.059	-0.020	-0.007	0.001	0.004	4	-0.263	-0.170	-0.125	-0.080	-0.039	4	-0.271	-0.164	-0.115	-0.072	-0.036
5	-0.036	-0.003	0.005	0.010	0.011	5	-0.244	-0.161	-0.118	-0.077	-0.039	5	-0.255	-0.156	-0.111	-0.070	-0.037
10	0.040	0.051	0.052	0.048	0.036	10	-0.196	-0.134	-0.105	-0.071	-0.038	10	-0.195	-0.130	-0.097	-0.064	-0.034
20	0.117	0.117	0.112	0.098	0.076	20	-0.148	-0.111	-0.089	-0.062	-0.035	20	-0.133	-0.096	-0.076	-0.053	-0.029
30	0.161	0.156	0.147	0.133	0.106	30	-0.122	-0.095	-0.077	-0.057	-0.033	30	-0.100	-0.077	-0.063	-0.045	-0.027
50	0.208	0.200	0.190	0.177	0.147	50	-0.085	-0.074	-0.063	-0.048	-0.029	50	-0.066	-0.054	-0.045	-0.035	-0.021
100	0.250	0.243	0.237	0.225	0.200	100	-0.049	-0.050	-0.045	-0.037	-0.025	100	-0.035	-0.031	-0.028	-0.022	-0.015

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

**Table 3: Bias of forecast error conditional on  $y = \alpha_2 + \sigma_2$  - (AR(1) model)**

Experiment no. 1a : No break						Experiment no. 4a : Large break in $\beta$						Experiment no. 7a : Lower post-break volatility					
$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100
0	-0.370	-0.200	-0.135	-0.081	-0.039	0	-0.396	-0.214	-0.145	-0.086	-0.041	0	-0.112	-0.059	-0.040	-0.026	-0.014
1	-0.344	-0.193	-0.132	-0.080	-0.039	1	-0.386	-0.221	-0.150	-0.089	-0.042	1	-0.163	-0.088	-0.059	-0.036	-0.018
2	-0.315	-0.184	-0.129	-0.078	-0.039	2	-0.371	-0.218	-0.151	-0.091	-0.044	2	-0.169	-0.094	-0.067	-0.041	-0.021
3	-0.297	-0.175	-0.125	-0.076	-0.038	3	-0.362	-0.219	-0.151	-0.093	-0.045	3	-0.165	-0.095	-0.067	-0.043	-0.023
4	-0.278	-0.170	-0.119	-0.076	-0.038	4	-0.358	-0.221	-0.154	-0.095	-0.046	4	-0.158	-0.094	-0.068	-0.045	-0.024
5	-0.262	-0.162	-0.116	-0.074	-0.037	5	-0.357	-0.222	-0.156	-0.096	-0.048	5	-0.149	-0.092	-0.068	-0.045	-0.025
10	-0.202	-0.136	-0.102	-0.068	-0.035	10	-0.363	-0.233	-0.170	-0.107	-0.054	10	-0.113	-0.078	-0.061	-0.044	-0.026
20	-0.136	-0.102	-0.081	-0.057	-0.032	20	-0.392	-0.262	-0.195	-0.127	-0.066	20	-0.075	-0.058	-0.048	-0.036	-0.024
30	-0.102	-0.082	-0.066	-0.050	-0.030	30	-0.418	-0.292	-0.220	-0.147	-0.078	30	-0.055	-0.045	-0.039	-0.031	-0.021
50	-0.067	-0.058	-0.049	-0.040	-0.026	50	-0.453	-0.334	-0.262	-0.181	-0.100	50	-0.035	-0.031	-0.028	-0.024	-0.018
100	-0.036	-0.033	-0.030	-0.026	-0.020	100	-0.499	-0.401	-0.332	-0.246	-0.145	100	-0.018	-0.017	-0.016	-0.015	-0.012
Experiment no. 2a : Moderate break in $\beta$						Experiment no. 5a : Post-break unit root						Experiment no. 8a : Break in mean (increase)					
$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100
0	-0.392	-0.214	-0.144	-0.085	-0.041	0	-0.413	-0.233	-0.163	-0.101	-0.052	0	-0.607	-0.288	-0.178	-0.096	-0.042
1	-0.375	-0.214	-0.146	-0.087	-0.042	1	-0.391	-0.227	-0.159	-0.100	-0.052	1	-0.576	-0.285	-0.181	-0.098	-0.044
2	-0.354	-0.209	-0.144	-0.087	-0.043	2	-0.367	-0.220	-0.157	-0.098	-0.051	2	-0.556	-0.280	-0.181	-0.098	-0.045
3	-0.337	-0.204	-0.144	-0.088	-0.044	3	-0.351	-0.211	-0.152	-0.096	-0.051	3	-0.527	-0.276	-0.180	-0.101	-0.046
4	-0.324	-0.200	-0.145	-0.089	-0.044	4	-0.338	-0.207	-0.149	-0.095	-0.051	4	-0.505	-0.274	-0.178	-0.100	-0.046
5	-0.317	-0.200	-0.143	-0.091	-0.045	5	-0.325	-0.199	-0.146	-0.094	-0.050	5	-0.488	-0.267	-0.179	-0.103	-0.047
10	-0.286	-0.190	-0.144	-0.094	-0.048	10	-0.303	-0.190	-0.138	-0.089	-0.048	10	-0.402	-0.247	-0.173	-0.105	-0.050
20	-0.269	-0.191	-0.149	-0.102	-0.055	20	-0.295	-0.190	-0.138	-0.088	-0.046	20	-0.310	-0.213	-0.161	-0.105	-0.055
30	-0.268	-0.200	-0.158	-0.111	-0.062	30	-0.301	-0.197	-0.143	-0.091	-0.047	30	-0.261	-0.194	-0.152	-0.105	-0.059
50	-0.269	-0.213	-0.174	-0.126	-0.074	50	-0.313	-0.215	-0.158	-0.098	-0.048	50	-0.208	-0.168	-0.140	-0.104	-0.064
100	-0.276	-0.233	-0.200	-0.156	-0.100	100	-0.337	-0.247	-0.188	-0.119	-0.056	100	-0.158	-0.141	-0.126	-0.104	-0.073
Experiment no. 3a : Break in $\beta$ (decline)						Experiment no. 6a : Higher post-break volatility						Experiment no. 9a : Break in mean (decrease)					
$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100
0	-0.226	-0.125	-0.087	-0.053	-0.028	0	-0.798	-0.438	-0.295	-0.174	-0.083	0	-0.124	-0.109	-0.091	-0.064	-0.035
1	-0.156	-0.084	-0.057	-0.036	-0.017	1	-0.681	-0.403	-0.279	-0.167	-0.082	1	-0.091	-0.092	-0.079	-0.058	-0.032
2	-0.113	-0.057	-0.036	-0.022	-0.010	2	-0.602	-0.377	-0.268	-0.165	-0.082	2	-0.067	-0.076	-0.069	-0.052	-0.031
3	-0.084	-0.037	-0.021	-0.010	-0.003	3	-0.556	-0.354	-0.259	-0.161	-0.080	3	-0.051	-0.064	-0.060	-0.048	-0.029
4	-0.059	-0.020	-0.007	0.001	0.004	4	-0.526	-0.340	-0.250	-0.160	-0.079	4	-0.037	-0.053	-0.052	-0.043	-0.027
5	-0.036	-0.003	0.005	0.010	0.011	5	-0.488	-0.322	-0.236	-0.154	-0.079	5	-0.026	-0.044	-0.045	-0.038	-0.025
10	0.040	0.051	0.052	0.048	0.036	10	-0.393	-0.269	-0.211	-0.143	-0.076	10	0.011	-0.014	-0.021	-0.023	-0.017
20	0.117	0.117	0.112	0.098	0.076	20	-0.295	-0.223	-0.177	-0.123	-0.070	20	0.044	0.021	0.009	0.000	-0.004
30	0.161	0.156	0.147	0.133	0.106	30	-0.244	-0.190	-0.153	-0.113	-0.066	30	0.060	0.039	0.027	0.015	0.005
50	0.208	0.200	0.190	0.177	0.147	50	-0.171	-0.148	-0.126	-0.096	-0.058	50	0.076	0.060	0.049	0.036	0.021
100	0.250	0.243	0.237	0.225	0.200	100	-0.098	-0.100	-0.090	-0.075	-0.051	100	0.089	0.079	0.071	0.060	0.043

See the note to Table 2.

**Table 4: Unconditional RMSFE as a function of pre-break ( $v_1$ ) and post-break window ( $v_2$ ) - (AR(1) model)**

Experiment no. 1a : No break

Experiment no. 4a : large break in  $\beta$

Experiment no. 7a : Lower post-break volatility

$v_1/v_2$	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

$v_1/v_2$	10	20	30	50	100
0	1.129	1.070	1.048	1.028	1.013
1	1.123	1.070	1.049	1.028	1.013
2	1.112	1.070	1.049	1.028	1.013
3	1.107	1.068	1.048	1.029	1.013
4	1.101	1.066	1.046	1.028	1.013
5	1.101	1.065	1.046	1.028	1.013
10	1.101	1.066	1.046	1.028	1.013
20	1.132	1.079	1.056	1.033	1.015
30	1.163	1.100	1.068	1.040	1.017
50	1.217	1.138	1.096	1.057	1.024
100	1.303	1.216	1.163	1.103	1.045

$v_1/v_2$	10	20	30	50	100
0	0.648	0.567	0.535	0.515	0.506
1	0.769	0.630	0.568	0.526	0.509
2	0.797	0.656	0.585	0.535	0.510
3	0.796	0.661	0.593	0.539	0.512
4	0.795	0.666	0.595	0.542	0.513
5	0.789	0.668	0.600	0.544	0.514
10	0.739	0.648	0.595	0.549	0.516
20	0.678	0.615	0.580	0.544	0.517
30	0.642	0.595	0.569	0.540	0.517
50	0.598	0.571	0.555	0.534	0.516
100	0.552	0.543	0.537	0.527	0.515

Experiment no. 2a : Moderate break in  $\beta$

$v_1/v_2$	10	20	30	50	100
0	1.135	1.073	1.048	1.028	1.013
1	1.120	1.070	1.047	1.028	1.013
2	1.109	1.067	1.046	1.027	1.013
3	1.100	1.063	1.045	1.027	1.013
4	1.092	1.061	1.044	1.026	1.013
5	1.087	1.059	1.043	1.026	1.012
10	1.072	1.051	1.039	1.024	1.012
20	1.070	1.048	1.037	1.025	1.012
30	1.075	1.051	1.039	1.026	1.013
50	1.084	1.060	1.046	1.030	1.015
100	1.110	1.081	1.063	1.045	1.023

Experiment no. 5a : Post-break unit root

$v_1/v_2$	10	20	30	50	100
0	1.116	1.064	1.044	1.028	1.014
1	1.109	1.063	1.043	1.027	1.014
2	1.101	1.061	1.042	1.027	1.014
3	1.101	1.058	1.041	1.026	1.014
4	1.103	1.058	1.041	1.026	1.014
5	1.104	1.057	1.040	1.026	1.014
10	1.116	1.062	1.042	1.026	1.013
20	1.145	1.079	1.052	1.029	1.013
30	1.177	1.100	1.064	1.034	1.015
50	1.223	1.140	1.094	1.047	1.018
100	1.306	1.228	1.162	1.087	1.030

Experiment no. 8a : Break in mean (increase)

$v_1/v_2$	10	20	30	50	100
0	1.152	1.078	1.050	1.028	1.012
1	1.139	1.075	1.048	1.027	1.012
2	1.129	1.071	1.046	1.026	1.012
3	1.122	1.069	1.045	1.025	1.011
4	1.114	1.068	1.043	1.025	1.012
5	1.108	1.064	1.043	1.024	1.011
10	1.078	1.053	1.038	1.022	1.011
20	1.051	1.038	1.030	1.019	1.010
30	1.039	1.030	1.024	1.017	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. 3a : Break in  $\beta$  (decline)

$v_1/v_2$	10	20	30	50	100
0	1.118	1.054	1.035	1.020	1.010
1	1.113	1.053	1.032	1.020	1.010
2	1.112	1.050	1.033	1.020	1.010
3	1.113	1.051	1.034	1.020	1.010
4	1.111	1.052	1.035	1.021	1.010
5	1.112	1.053	1.036	1.022	1.011
10	1.098	1.056	1.039	1.026	1.013
20	1.079	1.055	1.043	1.030	1.018
30	1.070	1.054	1.045	1.034	1.021
50	1.064	1.056	1.049	1.041	1.028
100	1.060	1.057	1.054	1.049	1.038

Experiment no. 6a : Higher post-break volatility

$v_1/v_2$	10	20	30	50	100
0	2.263	2.145	2.096	2.055	2.025
1	2.217	2.125	2.088	2.052	2.024
2	2.194	2.113	2.081	2.050	2.024
3	2.185	2.107	2.077	2.049	2.024
4	2.172	2.100	2.073	2.048	2.024
5	2.167	2.097	2.070	2.045	2.022
10	2.145	2.080	2.059	2.040	2.022
20	2.117	2.065	2.050	2.033	2.019
30	2.097	2.059	2.044	2.030	2.017
50	2.073	2.050	2.036	2.026	2.015
100	2.043	2.034	2.028	2.019	2.012

Experiment no. 9a : Break in mean (decrease)

$v_1/v_2$	10	20	30	50	100
0	1.154	1.079	1.050	1.027	1.012
1	1.142	1.074	1.048	1.027	1.012
2	1.132	1.071	1.047	1.026	1.012
3	1.122	1.069	1.045	1.026	1.012
4	1.114	1.066	1.044	1.025	1.011
5	1.106	1.064	1.042	1.024	1.011
10	1.079	1.052	1.037	1.022	1.011
20	1.051	1.038	1.030	1.019	1.010
30	1.039	1.029	1.024	1.017	1.009
50	1.027	1.021	1.018	1.013	1.008
100	1.017	1.014	1.012	1.010	1.007

See the note to Table 2.



**Table 5. RMSFE conditional on  $y = \alpha_2 + \sigma_2$  - (AR(1) model)**

Experiment no. 1a : No break						Experiment no. 4a : Large break in $\beta$						Experiment no. 7a : Lower post-break volatility					
$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100
0	1.421	1.164	1.088	1.038	1.013	0	1.344	1.152	1.086	1.039	1.013	0	0.922	0.619	0.551	0.519	0.506
1	1.375	1.153	1.085	1.037	1.012	1	1.293	1.146	1.085	1.039	1.013	1	1.177	0.710	0.591	0.530	0.508
2	1.329	1.142	1.079	1.036	1.012	2	1.254	1.128	1.078	1.037	1.013	2	1.208	0.742	0.609	0.539	0.509
3	1.301	1.130	1.075	1.034	1.012	3	1.220	1.118	1.073	1.037	1.013	3	1.178	0.745	0.618	0.542	0.510
4	1.274	1.123	1.069	1.033	1.012	4	1.197	1.109	1.070	1.035	1.013	4	1.156	0.746	0.620	0.545	0.511
5	1.248	1.115	1.067	1.032	1.011	5	1.182	1.102	1.066	1.034	1.013	5	1.118	0.744	0.623	0.547	0.512
10	1.165	1.085	1.053	1.028	1.011	10	1.141	1.084	1.057	1.032	1.013	10	0.959	0.709	0.614	0.550	0.514
20	1.086	1.054	1.037	1.021	1.009	20	1.120	1.074	1.051	1.030	1.013	20	0.786	0.651	0.593	0.545	0.515
30	1.053	1.037	1.027	1.017	1.008	30	1.118	1.074	1.051	1.030	1.013	30	0.701	0.618	0.577	0.541	0.515
50	1.027	1.021	1.017	1.012	1.007	50	1.119	1.078	1.055	1.033	1.014	50	0.618	0.581	0.560	0.535	0.515
100	1.011	1.009	1.008	1.007	1.005	100	1.129	1.091	1.068	1.043	1.019	100	0.555	0.545	0.538	0.526	0.514

  

Experiment no. 2a : Moderate break in $\beta$						Experiment no. 5a : Post-break unit root						Experiment no. 8a : Break in mean (increase)					
$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100
0	1.352	1.158	1.087	1.039	1.013	0	1.458	1.263	1.188	1.119	1.063	0	1.514	1.190	1.098	1.041	1.013
1	1.303	1.145	1.083	1.038	1.013	1	1.372	1.229	1.167	1.110	1.060	1	1.460	1.180	1.095	1.040	1.013
2	1.260	1.129	1.076	1.037	1.013	2	1.321	1.200	1.148	1.097	1.056	2	1.419	1.169	1.092	1.039	1.013
3	1.228	1.117	1.073	1.035	1.012	3	1.276	1.172	1.130	1.087	1.052	3	1.379	1.159	1.087	1.038	1.013
4	1.197	1.106	1.068	1.034	1.012	4	1.242	1.157	1.119	1.081	1.048	4	1.348	1.152	1.083	1.037	1.013
5	1.180	1.098	1.063	1.034	1.012	5	1.214	1.141	1.106	1.075	1.046	5	1.326	1.143	1.081	1.037	1.012
10	1.118	1.073	1.050	1.029	1.012	10	1.140	1.098	1.076	1.054	1.035	10	1.221	1.113	1.068	1.033	1.012
20	1.079	1.053	1.039	1.024	1.011	20	1.095	1.068	1.052	1.038	1.025	20	1.129	1.077	1.051	1.027	1.011
30	1.065	1.046	1.034	1.022	1.011	30	1.080	1.056	1.043	1.030	1.020	30	1.089	1.058	1.040	1.023	1.010
50	1.053	1.039	1.031	1.020	1.010	50	1.069	1.048	1.036	1.025	1.015	50	1.051	1.037	1.029	1.018	1.009
100	1.046	1.036	1.029	1.020	1.011	100	1.066	1.045	1.033	1.021	1.011	100	1.025	1.021	1.018	1.013	1.008

  

Experiment no. 3a : Break in $\beta$ (decline)						Experiment no. 6a : Higher post-break volatility						Experiment no. 9a : Break in mean (decrease)					
$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100
0	1.161	1.066	1.039	1.021	1.010	0	2.634	2.295	2.173	2.077	2.025	0	1.372	1.143	1.080	1.036	1.012
1	1.136	1.057	1.034	1.019	1.009	1	2.443	2.238	2.146	2.071	2.025	1	1.329	1.132	1.074	1.033	1.012
2	1.126	1.053	1.033	1.018	1.009	2	2.344	2.199	2.130	2.067	2.024	2	1.290	1.120	1.068	1.031	1.011
3	1.121	1.052	1.032	1.018	1.009	3	2.286	2.172	2.116	2.062	2.023	3	1.257	1.110	1.063	1.030	1.011
4	1.114	1.050	1.032	1.018	1.009	4	2.250	2.151	2.106	2.059	2.023	4	1.231	1.103	1.058	1.029	1.010
5	1.111	1.049	1.032	1.019	1.009	5	2.218	2.135	2.095	2.055	2.022	5	1.211	1.096	1.056	1.027	1.010
10	1.087	1.047	1.032	1.020	1.010	10	2.143	2.088	2.067	2.043	2.020	10	1.139	1.070	1.043	1.023	1.009
20	1.062	1.041	1.032	1.022	1.012	20	2.085	2.055	2.043	2.030	2.016	20	1.073	1.043	1.030	1.017	1.008
30	1.051	1.038	1.031	1.024	1.015	30	2.060	2.040	2.031	2.023	2.013	30	1.045	1.030	1.022	1.014	1.007
50	1.044	1.037	1.033	1.027	1.019	50	2.034	2.026	2.021	2.016	2.010	50	1.025	1.019	1.015	1.010	1.006
100	1.040	1.038	1.036	1.032	1.025	100	2.014	2.013	2.011	2.009	2.006	100	1.013	1.011	1.009	1.007	1.005

See the note to Table 2.

Table 6: Small sample bias of the OLS estimate of  $\beta_1$  - (AR(2) Model)

Experiment no. 1b : No break						Experiment no.4b : Large decrease in $\beta$						Experiment no. 7b : Lower post-break volatility					
$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100
0	-0.193	-0.088	-0.058	-0.034	-0.015	0	-0.168	-0.080	-0.052	-0.030	-0.015	0	-0.136	-0.068	-0.046	-0.029	-0.015
1	-0.165	-0.081	-0.052	-0.032	-0.017	1	-0.134	-0.061	-0.038	-0.023	-0.011	1	-0.144	-0.084	-0.058	-0.037	-0.019
2	-0.157	-0.080	-0.055	-0.033	-0.016	2	-0.097	-0.046	-0.031	-0.018	-0.008	2	-0.148	-0.090	-0.066	-0.043	-0.021
3	-0.145	-0.076	-0.050	-0.031	-0.015	3	-0.062	-0.031	-0.018	-0.010	-0.003	3	-0.138	-0.088	-0.064	-0.045	-0.025
4	-0.129	-0.072	-0.049	-0.031	-0.015	4	-0.037	-0.015	-0.010	-0.006	-0.001	4	-0.125	-0.089	-0.064	-0.049	-0.025
5	-0.118	-0.069	-0.048	-0.030	-0.016	5	-0.015	-0.006	-0.001	0.002	0.002	5	-0.122	-0.085	-0.064	-0.046	-0.027
10	-0.087	-0.054	-0.042	-0.027	-0.013	10	0.063	0.046	0.036	0.026	0.015	10	-0.091	-0.068	-0.058	-0.043	-0.027
20	-0.059	-0.039	-0.034	-0.022	-0.013	20	0.145	0.112	0.091	0.064	0.039	20	-0.059	-0.051	-0.044	-0.034	-0.026
30	-0.042	-0.034	-0.028	-0.020	-0.013	30	0.183	0.148	0.125	0.094	0.059	30	-0.043	-0.038	-0.035	-0.028	-0.021
50	-0.027	-0.024	-0.020	-0.016	-0.010	50	0.224	0.193	0.168	0.135	0.091	50	-0.030	-0.026	-0.023	-0.021	-0.015
100	-0.014	-0.014	-0.012	-0.011	-0.008	100	0.258	0.236	0.218	0.189	0.142	100	-0.016	-0.014	-0.013	-0.011	-0.010
Experiment no. 2b : Small break in $\beta$						Experiment no. 5b : Break in $\beta$ (post-break unit root)						Experiment no. 8b : Break in mean (increase)					
$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100
0	-0.193	-0.087	-0.057	-0.033	-0.017	0	-0.353	-0.179	-0.126	-0.077	-0.038	0	-0.186	-0.087	-0.055	-0.033	-0.016
1	-0.185	-0.088	-0.058	-0.035	-0.017	1	-0.325	-0.182	-0.124	-0.078	-0.038	1	-0.148	-0.074	-0.047	-0.028	-0.014
2	-0.170	-0.087	-0.057	-0.035	-0.018	2	-0.306	-0.174	-0.122	-0.076	-0.041	2	-0.125	-0.062	-0.042	-0.025	-0.013
3	-0.158	-0.087	-0.059	-0.036	-0.018	3	-0.292	-0.172	-0.121	-0.077	-0.041	3	-0.105	-0.053	-0.037	-0.022	-0.012
4	-0.154	-0.087	-0.061	-0.037	-0.019	4	-0.275	-0.168	-0.122	-0.078	-0.040	4	-0.092	-0.047	-0.031	-0.019	-0.009
5	-0.146	-0.087	-0.061	-0.037	-0.020	5	-0.267	-0.167	-0.122	-0.077	-0.042	5	-0.081	-0.041	-0.027	-0.017	-0.008
10	-0.129	-0.085	-0.063	-0.041	-0.022	10	-0.228	-0.154	-0.117	-0.079	-0.043	10	-0.046	-0.020	-0.011	-0.005	-0.001
20	-0.114	-0.086	-0.067	-0.048	-0.027	20	-0.202	-0.148	-0.118	-0.083	-0.048	20	-0.021	-0.002	0.004	0.007	0.007
30	-0.108	-0.087	-0.071	-0.053	-0.031	30	-0.193	-0.147	-0.121	-0.087	-0.053	30	-0.013	0.003	0.008	0.013	0.014
50	-0.104	-0.090	-0.076	-0.060	-0.039	50	-0.188	-0.153	-0.128	-0.098	-0.063	50	-0.006	0.008	0.014	0.020	0.020
100	-0.102	-0.091	-0.083	-0.071	-0.052	100	-0.190	-0.163	-0.142	-0.115	-0.079	100	-0.001	0.008	0.014	0.020	0.026
Experiment no. 3b : Large increase in $\beta$						Experiment no. 6b : Higher post-break volatility						Experiment no. 9b : Break in mean (decrease)					
$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100
0	-0.199	-0.088	-0.057	-0.033	-0.016	0	-0.204	-0.092	-0.057	-0.033	-0.016	0	-0.189	-0.084	-0.054	-0.032	-0.016
1	-0.199	-0.099	-0.067	-0.039	-0.018	1	-0.183	-0.087	-0.055	-0.032	-0.015	1	-0.154	-0.074	-0.046	-0.027	-0.014
2	-0.201	-0.108	-0.070	-0.043	-0.021	2	-0.162	-0.084	-0.056	-0.032	-0.016	2	-0.126	-0.062	-0.042	-0.025	-0.012
3	-0.209	-0.115	-0.079	-0.049	-0.025	3	-0.148	-0.083	-0.053	-0.032	-0.016	3	-0.107	-0.053	-0.037	-0.023	-0.010
4	-0.212	-0.121	-0.084	-0.052	-0.026	4	-0.136	-0.075	-0.053	-0.031	-0.015	4	-0.091	-0.048	-0.031	-0.018	-0.010
5	-0.214	-0.127	-0.091	-0.056	-0.030	5	-0.131	-0.071	-0.051	-0.030	-0.016	5	-0.079	-0.039	-0.027	-0.018	-0.008
10	-0.233	-0.154	-0.115	-0.077	-0.040	10	-0.103	-0.062	-0.044	-0.030	-0.015	10	-0.045	-0.021	-0.013	-0.006	-0.002
20	-0.249	-0.187	-0.151	-0.108	-0.063	20	-0.073	-0.052	-0.039	-0.026	-0.013	20	-0.023	-0.003	0.003	0.007	0.008
30	-0.264	-0.209	-0.175	-0.130	-0.080	30	-0.060	-0.044	-0.036	-0.025	-0.014	30	-0.013	0.004	0.009	0.012	0.013
50	-0.272	-0.235	-0.205	-0.164	-0.109	50	-0.047	-0.039	-0.031	-0.020	-0.013	50	-0.006	0.007	0.014	0.018	0.020
100	-0.286	-0.261	-0.241	-0.207	-0.155	100	-0.027	-0.025	-0.023	-0.016	-0.011	100	-0.003	0.008	0.014	0.021	0.026

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

**Table 7: Small sample bias of the OLS estimate of  $\beta_1 + \beta_2$  - (AR(2) model)**

Experiment no. 1b : No break						Experiment no. 4b : Large decrease in $\beta$						Experiment no. 7b : Lower post-break volatility					
$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100
0	-0.315	-0.152	-0.098	-0.059	-0.028	0	-0.342	-0.173	-0.117	-0.071	-0.035	0	-0.159	-0.091	-0.065	-0.043	-0.024
1	-0.277	-0.143	-0.096	-0.057	-0.029	1	-0.302	-0.159	-0.106	-0.065	-0.032	1	-0.193	-0.116	-0.083	-0.056	-0.030
2	-0.256	-0.137	-0.094	-0.056	-0.028	2	-0.267	-0.146	-0.099	-0.060	-0.031	2	-0.205	-0.127	-0.095	-0.063	-0.034
3	-0.237	-0.131	-0.088	-0.055	-0.028	3	-0.231	-0.129	-0.090	-0.055	-0.028	3	-0.202	-0.131	-0.099	-0.068	-0.037
4	-0.220	-0.127	-0.087	-0.054	-0.028	4	-0.204	-0.117	-0.084	-0.053	-0.026	4	-0.195	-0.132	-0.100	-0.072	-0.039
5	-0.202	-0.118	-0.083	-0.052	-0.028	5	-0.182	-0.110	-0.078	-0.047	-0.024	5	-0.192	-0.129	-0.102	-0.071	-0.041
10	-0.147	-0.097	-0.073	-0.048	-0.025	10	-0.102	-0.066	-0.047	-0.032	-0.017	10	-0.151	-0.109	-0.093	-0.070	-0.044
20	-0.100	-0.072	-0.058	-0.041	-0.024	20	-0.023	-0.015	-0.009	-0.006	-0.002	20	-0.102	-0.084	-0.071	-0.056	-0.041
30	-0.072	-0.058	-0.048	-0.036	-0.022	30	0.013	0.014	0.013	0.012	0.009	30	-0.076	-0.066	-0.058	-0.049	-0.034
50	-0.049	-0.041	-0.036	-0.029	-0.018	50	0.047	0.043	0.040	0.035	0.026	50	-0.049	-0.044	-0.041	-0.035	-0.028
100	-0.026	-0.024	-0.022	-0.020	-0.014	100	0.076	0.072	0.068	0.062	0.049	100	-0.027	-0.025	-0.024	-0.022	-0.018

  

Experiment no. 2b : Small break in $\beta$						Experiment no. 5b : Break in $\beta$ (post-break unit root)						Experiment no. 8b : Break in mean (increase)					
$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100
0	-0.317	-0.154	-0.098	-0.059	-0.030	0	-0.544	-0.290	-0.202	-0.125	-0.063	0	-0.289	-0.144	-0.096	-0.056	-0.028
1	-0.303	-0.154	-0.102	-0.061	-0.030	1	-0.503	-0.284	-0.195	-0.122	-0.062	1	-0.236	-0.121	-0.083	-0.049	-0.025
2	-0.286	-0.154	-0.101	-0.063	-0.031	2	-0.465	-0.271	-0.190	-0.120	-0.062	2	-0.202	-0.104	-0.071	-0.043	-0.021
3	-0.272	-0.154	-0.106	-0.063	-0.033	3	-0.437	-0.261	-0.184	-0.117	-0.061	3	-0.174	-0.090	-0.062	-0.036	-0.019
4	-0.264	-0.153	-0.108	-0.066	-0.034	4	-0.415	-0.250	-0.180	-0.115	-0.060	4	-0.154	-0.079	-0.052	-0.030	-0.015
5	-0.253	-0.152	-0.108	-0.067	-0.036	5	-0.395	-0.243	-0.175	-0.112	-0.060	5	-0.138	-0.069	-0.045	-0.026	-0.013
10	-0.232	-0.154	-0.114	-0.075	-0.040	10	-0.327	-0.214	-0.159	-0.105	-0.057	10	-0.084	-0.035	-0.018	-0.007	-0.001
20	-0.211	-0.157	-0.125	-0.090	-0.051	20	-0.269	-0.186	-0.142	-0.097	-0.053	20	-0.042	-0.006	0.007	0.016	0.017
30	-0.207	-0.163	-0.135	-0.100	-0.059	30	-0.244	-0.175	-0.135	-0.093	-0.051	30	-0.027	0.006	0.018	0.028	0.028
50	-0.201	-0.171	-0.147	-0.116	-0.074	50	-0.221	-0.167	-0.132	-0.091	-0.050	50	-0.014	0.013	0.026	0.039	0.041
100	-0.200	-0.179	-0.164	-0.139	-0.102	100	-0.207	-0.167	-0.137	-0.097	-0.054	100	-0.006	0.014	0.026	0.041	0.052

  

Experiment no. 3b : Large increase in $\beta$						Experiment no. 6b : Higher post-break volatility						Experiment no. 9b : Break in mean (decrease)					
$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100
0	-0.323	-0.153	-0.100	-0.058	-0.029	0	-0.365	-0.166	-0.105	-0.061	-0.029	0	-0.289	-0.142	-0.094	-0.055	-0.028
1	-0.304	-0.151	-0.102	-0.059	-0.030	1	-0.327	-0.156	-0.102	-0.058	-0.028	1	-0.238	-0.122	-0.081	-0.050	-0.025
2	-0.282	-0.147	-0.100	-0.060	-0.029	2	-0.292	-0.150	-0.101	-0.058	-0.028	2	-0.201	-0.106	-0.072	-0.043	-0.021
3	-0.268	-0.147	-0.100	-0.060	-0.030	3	-0.265	-0.144	-0.095	-0.057	-0.029	3	-0.175	-0.091	-0.061	-0.038	-0.018
4	-0.258	-0.143	-0.100	-0.060	-0.031	4	-0.250	-0.137	-0.095	-0.057	-0.028	4	-0.154	-0.078	-0.053	-0.031	-0.016
5	-0.245	-0.140	-0.098	-0.059	-0.031	5	-0.232	-0.130	-0.090	-0.055	-0.028	5	-0.138	-0.069	-0.045	-0.027	-0.013
10	-0.209	-0.134	-0.099	-0.064	-0.033	10	-0.183	-0.116	-0.081	-0.053	-0.027	10	-0.081	-0.035	-0.019	-0.007	-0.001
20	-0.172	-0.126	-0.098	-0.066	-0.037	20	-0.132	-0.093	-0.070	-0.047	-0.026	20	-0.043	-0.006	0.008	0.016	0.017
30	-0.160	-0.120	-0.097	-0.069	-0.040	30	-0.106	-0.078	-0.062	-0.043	-0.024	30	-0.027	0.004	0.019	0.028	0.029
50	-0.140	-0.116	-0.099	-0.076	-0.046	50	-0.080	-0.064	-0.052	-0.037	-0.022	50	-0.013	0.013	0.026	0.038	0.042
100	-0.127	-0.112	-0.102	-0.084	-0.059	100	-0.046	-0.044	-0.038	-0.029	-0.019	100	-0.007	0.014	0.026	0.041	0.053

See the note to Table 6.

**Table 8: Bias of forecast error conditional on  $y = \alpha_2 + \sigma_2$  (AR(2) model)**

Experiment no. 1b: No break

$v_1/v_2$	10	20	30	50	100
0	-0.315	-0.152	-0.098	-0.059	-0.028
1	-0.277	-0.143	-0.096	-0.057	-0.029
2	-0.256	-0.137	-0.094	-0.056	-0.028
3	-0.237	-0.131	-0.088	-0.055	-0.028
4	-0.220	-0.127	-0.087	-0.054	-0.028
5	-0.202	-0.118	-0.083	-0.052	-0.028
10	-0.147	-0.097	-0.073	-0.048	-0.025
20	-0.100	-0.072	-0.058	-0.041	-0.024
30	-0.072	-0.058	-0.048	-0.036	-0.022
50	-0.049	-0.041	-0.036	-0.029	-0.018
100	-0.026	-0.024	-0.022	-0.020	-0.014

Experiment no. 4b: Large decrease in  $\beta$ 

$v_1/v_2$	10	20	30	50	100
0	-0.342	-0.173	-0.117	-0.071	-0.035
1	-0.302	-0.159	-0.106	-0.065	-0.032
2	-0.267	-0.146	-0.099	-0.060	-0.031
3	-0.231	-0.129	-0.090	-0.055	-0.028
4	-0.204	-0.117	-0.084	-0.053	-0.026
5	-0.182	-0.110	-0.078	-0.047	-0.024
10	-0.102	-0.066	-0.047	-0.032	-0.017
20	-0.023	-0.015	-0.009	-0.006	-0.002
30	0.013	0.014	0.013	0.012	0.009
50	0.047	0.043	0.040	0.035	0.026
100	0.076	0.072	0.068	0.062	0.049

Experiment no. 7b: Lower post-break volatility

$v_1/v_2$	10	20	30	50	100
0	-0.079	-0.045	-0.033	-0.021	-0.012
1	-0.096	-0.058	-0.042	-0.028	-0.015
2	-0.102	-0.064	-0.047	-0.031	-0.017
3	-0.101	-0.066	-0.049	-0.034	-0.019
4	-0.098	-0.066	-0.050	-0.036	-0.020
5	-0.096	-0.064	-0.051	-0.035	-0.020
10	-0.076	-0.055	-0.046	-0.035	-0.022
20	-0.051	-0.042	-0.035	-0.028	-0.020
30	-0.038	-0.033	-0.029	-0.024	-0.017
50	-0.025	-0.022	-0.021	-0.017	-0.014
100	-0.013	-0.012	-0.012	-0.011	-0.009

Experiment no. 2b: Small break in  $\beta$ 

$v_1/v_2$	10	20	30	50	100
0	-0.317	-0.154	-0.098	-0.059	-0.030
1	-0.303	-0.154	-0.102	-0.061	-0.030
2	-0.286	-0.154	-0.101	-0.063	-0.031
3	-0.272	-0.154	-0.106	-0.063	-0.033
4	-0.264	-0.153	-0.108	-0.066	-0.034
5	-0.253	-0.152	-0.108	-0.067	-0.036
10	-0.232	-0.154	-0.114	-0.075	-0.040
20	-0.211	-0.157	-0.125	-0.090	-0.051
30	-0.207	-0.163	-0.135	-0.100	-0.059
50	-0.201	-0.171	-0.147	-0.116	-0.074
100	-0.200	-0.179	-0.164	-0.139	-0.102

Experiment no. 5b: Break in  $\beta$  (post-break unit root)

$v_1/v_2$	10	20	30	50	100
0	-0.544	-0.290	-0.202	-0.125	-0.063
1	-0.503	-0.284	-0.195	-0.122	-0.062
2	-0.465	-0.271	-0.190	-0.120	-0.062
3	-0.437	-0.261	-0.184	-0.117	-0.061
4	-0.415	-0.250	-0.180	-0.115	-0.060
5	-0.395	-0.243	-0.175	-0.112	-0.060
10	-0.327	-0.214	-0.159	-0.105	-0.057
20	-0.269	-0.186	-0.142	-0.097	-0.053
30	-0.244	-0.175	-0.135	-0.093	-0.051
50	-0.221	-0.167	-0.132	-0.091	-0.050
100	-0.207	-0.167	-0.137	-0.097	-0.054

Experiment no. 8b: Break in mean (increase)

$v_1/v_2$	10	20	30	50	100
0	-0.344	-0.160	-0.103	-0.059	-0.029
1	-0.339	-0.161	-0.106	-0.061	-0.030
2	-0.340	-0.164	-0.108	-0.064	-0.031
3	-0.338	-0.168	-0.111	-0.065	-0.033
4	-0.341	-0.173	-0.114	-0.066	-0.033
5	-0.342	-0.176	-0.117	-0.069	-0.035
10	-0.349	-0.195	-0.133	-0.082	-0.041
20	-0.363	-0.227	-0.165	-0.104	-0.054
30	-0.377	-0.253	-0.192	-0.126	-0.067
50	-0.392	-0.290	-0.231	-0.161	-0.091
100	-0.411	-0.340	-0.290	-0.223	-0.139

Experiment no. 3b: Large increase in  $\beta$ 

$v_1/v_2$	10	20	30	50	100
0	-0.323	-0.153	-0.100	-0.058	-0.029
1	-0.304	-0.151	-0.102	-0.059	-0.030
2	-0.282	-0.147	-0.100	-0.060	-0.029
3	-0.268	-0.147	-0.100	-0.060	-0.030
4	-0.258	-0.143	-0.100	-0.060	-0.031
5	-0.245	-0.140	-0.098	-0.059	-0.031
10	-0.209	-0.134	-0.099	-0.064	-0.033
20	-0.172	-0.126	-0.098	-0.066	-0.037
30	-0.160	-0.120	-0.097	-0.069	-0.040
50	-0.140	-0.116	-0.099	-0.076	-0.046
100	-0.127	-0.112	-0.102	-0.084	-0.059

Experiment no. 6b: Higher post-break volatility

$v_1/v_2$	10	20	30	50	100
0	-0.730	-0.332	-0.209	-0.123	-0.057
1	-0.655	-0.311	-0.204	-0.117	-0.056
2	-0.584	-0.299	-0.201	-0.117	-0.056
3	-0.529	-0.288	-0.190	-0.113	-0.057
4	-0.499	-0.273	-0.190	-0.115	-0.056
5	-0.465	-0.261	-0.180	-0.111	-0.056
10	-0.366	-0.232	-0.161	-0.105	-0.054
20	-0.263	-0.186	-0.140	-0.094	-0.053
30	-0.213	-0.157	-0.125	-0.086	-0.048
50	-0.161	-0.128	-0.105	-0.074	-0.045
100	-0.092	-0.088	-0.076	-0.058	-0.038

Experiment no. 9b: Break in mean (decrease)

$v_1/v_2$	10	20	30	50	100
0	-0.234	-0.126	-0.087	-0.052	-0.028
1	-0.134	-0.082	-0.058	-0.038	-0.020
2	-0.065	-0.046	-0.035	-0.023	-0.012
3	-0.013	-0.014	-0.011	-0.009	-0.004
4	0.033	0.014	0.008	0.005	0.002
5	0.067	0.038	0.027	0.016	0.009
10	0.183	0.124	0.096	0.067	0.039
20	0.279	0.216	0.180	0.136	0.087
30	0.323	0.265	0.228	0.181	0.123
50	0.364	0.317	0.284	0.238	0.174
100	0.399	0.367	0.342	0.304	0.244

See the note to Table 6.

**Table 9: Unconditional root mean squared forecast error - (AR(2) model)**

Experiment no. 1b : No break						Experiment no. 4b : Large decrease in $\beta$						Experiment no. 7b : Lower post-break volatility					
$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100
0	1.192	1.083	1.054	1.031	1.015	0	1.180	1.080	1.052	1.031	1.015	0	0.569	0.531	0.522	0.513	0.507
1	1.166	1.078	1.051	1.030	1.015	1	1.160	1.077	1.051	1.029	1.015	1	0.604	0.549	0.533	0.520	0.510
2	1.152	1.075	1.048	1.030	1.015	2	1.145	1.074	1.049	1.029	1.015	2	0.619	0.558	0.539	0.524	0.512
3	1.139	1.071	1.048	1.029	1.015	3	1.131	1.070	1.048	1.029	1.015	3	0.628	0.562	0.542	0.526	0.514
4	1.125	1.068	1.046	1.028	1.015	4	1.122	1.067	1.047	1.029	1.015	4	0.633	0.566	0.546	0.528	0.515
5	1.115	1.066	1.045	1.029	1.014	5	1.115	1.064	1.044	1.028	1.014	5	0.635	0.568	0.547	0.529	0.516
10	1.082	1.054	1.040	1.026	1.014	10	1.086	1.056	1.040	1.026	1.014	10	0.635	0.572	0.551	0.533	0.518
20	1.052	1.039	1.031	1.022	1.012	20	1.068	1.047	1.036	1.025	1.014	20	0.615	0.570	0.551	0.532	0.518
30	1.039	1.031	1.025	1.019	1.011	30	1.058	1.042	1.036	1.024	1.014	30	0.594	0.563	0.547	0.532	0.518
50	1.025	1.022	1.019	1.015	1.010	50	1.052	1.042	1.034	1.025	1.015	50	0.567	0.551	0.541	0.529	0.517
100	1.014	1.013	1.012	1.010	1.008	100	1.047	1.041	1.036	1.028	1.018	100	0.539	0.533	0.529	0.523	0.515

  

Experiment no. 2b : Small break in $\beta$						Experiment no. 5b : Break in $\beta$ (post-break unit root)						Experiment no. 8b : Break in mean (increase)					
$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100
0	1.190	1.083	1.053	1.031	1.015	0	1.177	1.089	1.061	1.038	1.019	0	1.185	1.083	1.052	1.031	1.015
1	1.170	1.080	1.052	1.031	1.015	1	1.165	1.086	1.059	1.037	1.019	1	1.166	1.075	1.049	1.030	1.015
2	1.152	1.075	1.051	1.030	1.015	2	1.149	1.081	1.057	1.037	1.019	2	1.148	1.074	1.049	1.029	1.015
3	1.137	1.073	1.048	1.029	1.015	3	1.140	1.079	1.057	1.035	1.019	3	1.140	1.069	1.047	1.028	1.015
4	1.124	1.070	1.048	1.029	1.015	4	1.129	1.076	1.055	1.035	1.019	4	1.134	1.068	1.047	1.028	1.015
5	1.114	1.066	1.046	1.028	1.015	5	1.124	1.075	1.052	1.034	1.018	5	1.125	1.066	1.045	1.028	1.015
10	1.083	1.056	1.041	1.027	1.014	10	1.102	1.067	1.048	1.032	1.017	10	1.105	1.062	1.043	1.027	1.015
20	1.059	1.045	1.034	1.024	1.013	20	1.084	1.058	1.045	1.030	1.016	20	1.092	1.059	1.044	1.029	1.015
30	1.050	1.038	1.031	1.023	1.013	30	1.077	1.058	1.044	1.030	1.017	30	1.088	1.059	1.046	1.031	1.017
50	1.039	1.033	1.027	1.021	1.013	50	1.077	1.059	1.048	1.032	1.017	50	1.086	1.063	1.050	1.035	1.020
100	1.033	1.028	1.025	1.019	1.013	100	1.089	1.072	1.060	1.041	1.021	100	1.087	1.070	1.059	1.044	1.027

  

Experiment no. 3b : Large increase in $\beta$						Experiment no. 6b : Higher post-break volatility						Experiment no. 9b : Break in mean (decrease)					
$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100
0	1.194	1.084	1.053	1.031	1.015	0	2.428	2.176	2.111	2.063	2.030	0	1.184	1.080	1.052	1.030	1.015
1	1.175	1.079	1.053	1.031	1.015	1	2.367	2.164	2.107	2.062	2.030	1	1.164	1.075	1.050	1.030	1.015
2	1.156	1.076	1.051	1.030	1.015	2	2.332	2.159	2.104	2.061	2.031	2	1.151	1.072	1.048	1.029	1.015
3	1.143	1.073	1.049	1.030	1.015	3	2.315	2.147	2.100	2.060	2.030	3	1.140	1.070	1.047	1.029	1.015
4	1.130	1.071	1.049	1.029	1.015	4	2.302	2.145	2.096	2.058	2.029	4	1.132	1.068	1.047	1.028	1.014
5	1.122	1.068	1.047	1.029	1.015	5	2.287	2.140	2.096	2.058	2.029	5	1.124	1.067	1.046	1.028	1.014
10	1.097	1.061	1.043	1.028	1.014	10	2.238	2.124	2.087	2.055	2.029	10	1.104	1.061	1.044	1.027	1.014
20	1.074	1.053	1.040	1.026	1.015	20	2.194	2.106	2.076	2.049	2.027	20	1.091	1.059	1.044	1.029	1.015
30	1.065	1.048	1.038	1.026	1.015	30	2.161	2.095	2.068	2.046	2.026	30	1.088	1.060	1.045	1.031	1.017
50	1.055	1.046	1.037	1.027	1.016	50	2.122	2.081	2.061	2.042	2.024	50	1.086	1.063	1.050	1.035	1.020
100	1.050	1.044	1.038	1.030	1.019	100	2.080	2.061	2.049	2.034	2.021	100	1.087	1.070	1.059	1.044	1.027

See the note to Table 6.

**Table 10: Root mean squared forecast error conditional on  $y = \alpha_2 + \sigma_2$  - (AR(2) model)**

Experiment no. 1b : No break						Experiment no. 4b : Large decrease in $\beta$						Experiment no. 7b : Lower post-break volatility					
$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100
0	1.206	1.074	1.043	1.022	1.010	0	1.214	1.084	1.051	1.027	1.012	0	0.572	0.528	0.517	0.510	0.505
1	1.172	1.069	1.042	1.022	1.010	1	1.185	1.078	1.047	1.026	1.012	1	0.607	0.543	0.525	0.514	0.506
2	1.155	1.066	1.040	1.022	1.010	2	1.163	1.073	1.045	1.025	1.012	2	0.621	0.549	0.530	0.516	0.507
3	1.139	1.062	1.038	1.021	1.010	3	1.142	1.067	1.043	1.024	1.011	3	0.628	0.553	0.533	0.518	0.508
4	1.125	1.059	1.037	1.020	1.010	4	1.128	1.063	1.041	1.023	1.011	4	0.633	0.556	0.534	0.519	0.509
5	1.111	1.054	1.035	1.020	1.010	5	1.114	1.059	1.039	1.023	1.011	5	0.635	0.559	0.537	0.520	0.509
10	1.074	1.043	1.030	1.018	1.009	10	1.073	1.045	1.031	1.020	1.010	10	0.634	0.565	0.541	0.523	0.510
20	1.043	1.030	1.022	1.015	1.008	20	1.041	1.030	1.023	1.016	1.009	20	0.612	0.565	0.544	0.525	0.512
30	1.030	1.023	1.018	1.013	1.007	30	1.028	1.022	1.018	1.014	1.008	30	0.590	0.559	0.542	0.526	0.512
50	1.018	1.015	1.013	1.010	1.006	50	1.018	1.016	1.014	1.011	1.007	50	0.564	0.548	0.537	0.525	0.513
100	1.009	1.008	1.007	1.006	1.005	100	1.012	1.011	1.010	1.009	1.006	100	0.537	0.531	0.527	0.520	0.512

  

Experiment no. 2b : Small break in $\beta$						Experiment no. 5b : Break in $\beta$ (post-break unit root)						Experiment no. 8b : Break in mean (increase)					
$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100
0	1.207	1.076	1.043	1.023	1.010	0	1.623	1.319	1.219	1.133	1.066	0	1.233	1.080	1.045	1.023	1.010
1	1.183	1.072	1.043	1.023	1.010	1	1.523	1.288	1.191	1.125	1.064	1	1.215	1.077	1.044	1.023	1.010
2	1.160	1.068	1.041	1.022	1.010	2	1.436	1.248	1.177	1.111	1.061	2	1.199	1.075	1.043	1.023	1.010
3	1.145	1.067	1.041	1.022	1.010	3	1.376	1.219	1.157	1.103	1.059	3	1.188	1.073	1.043	1.023	1.010
4	1.129	1.063	1.039	1.022	1.010	4	1.331	1.196	1.144	1.096	1.054	4	1.177	1.072	1.042	1.022	1.010
5	1.118	1.060	1.039	1.021	1.010	5	1.295	1.176	1.129	1.088	1.051	5	1.169	1.069	1.041	1.022	1.010
10	1.085	1.050	1.034	1.020	1.010	10	1.185	1.117	1.088	1.062	1.039	10	1.140	1.064	1.039	1.022	1.010
20	1.058	1.040	1.029	1.019	1.009	20	1.104	1.069	1.053	1.039	1.026	20	1.117	1.060	1.039	1.022	1.010
30	1.047	1.034	1.027	1.018	1.009	30	1.072	1.050	1.039	1.029	1.020	30	1.107	1.060	1.040	1.022	1.010
50	1.037	1.029	1.024	1.017	1.009	50	1.048	1.035	1.027	1.020	1.013	50	1.100	1.062	1.044	1.025	1.011
100	1.029	1.024	1.021	1.016	1.010	100	1.031	1.024	1.019	1.014	1.009	100	1.095	1.069	1.053	1.034	1.016

  

Experiment no. 3b : Large increase in $\beta$						Experiment no. 6b : Higher post-break volatility						Experiment no. 9b : Break in mean (decrease)					
$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100	$v_1/v_2$	10	20	30	50	100
0	1.210	1.075	1.044	1.022	1.010	0	2.443	2.159	2.089	2.047	2.020	0	1.159	1.064	1.039	1.021	1.010
1	1.185	1.072	1.043	1.022	1.010	1	2.356	2.142	2.085	2.044	2.020	1	1.118	1.053	1.033	1.019	1.009
2	1.160	1.068	1.041	1.022	1.010	2	2.296	2.132	2.081	2.043	2.020	2	1.096	1.046	1.030	1.018	1.009
3	1.145	1.065	1.040	1.022	1.010	3	2.259	2.122	2.075	2.042	2.019	3	1.083	1.041	1.027	1.017	1.008
4	1.132	1.061	1.039	1.021	1.010	4	2.230	2.113	2.073	2.041	2.019	4	1.073	1.038	1.026	1.016	1.008
5	1.120	1.058	1.037	1.021	1.010	5	2.209	2.106	2.069	2.040	2.019	5	1.068	1.036	1.025	1.015	1.008
10	1.087	1.049	1.033	1.019	1.009	10	2.148	2.083	2.057	2.035	2.018	10	1.056	1.033	1.023	1.015	1.008
20	1.055	1.037	1.027	1.017	1.009	20	2.097	2.061	2.044	2.029	2.016	20	1.059	1.038	1.028	1.018	1.010
30	1.042	1.030	1.023	1.015	1.008	30	2.074	2.049	2.037	2.025	2.014	30	1.064	1.045	1.035	1.023	1.012
50	1.029	1.023	1.018	1.013	1.008	50	2.051	2.037	2.029	2.021	2.013	50	1.072	1.055	1.045	1.033	1.019
100	1.018	1.015	1.013	1.010	1.007	100	2.028	2.024	2.020	2.015	2.010	100	1.080	1.069	1.060	1.048	1.032

See the note to Table 6.