Department of Economics, Hitotsubashi University

Discussion Paper Series #2014-16

Improving the Finite Sample Performance of Tests for a Shift in Mean

Daisuke Yamazaki and Eiji Kurozumi

November 2014

Improving the Finite Sample Performance of Tests for a Shift in Mean*

Daisuke Yamazaki†
Department of Economics
Hitotsubashi University

Eiji Kurozumi‡
Department of Economics
Hitotsubashi University

November 10, 2014

Abstract

It is widely known that structural break tests based on the long-run variance estimator, which is estimated under the alternative, suffer from serious size distortion when the errors are serially correlated. In this paper, we propose bias-corrected tests for a shift in mean by correcting the bias of the long-run variance estimator up to $O(1/T)$. Simulation results show that the proposed tests have good size and high power.

JEL classification: C12, C22

Key words: structural change, long-run variance, bias correction

*The authors are grateful to Graham Elliott, Dukpa Kim, Yohei Yamamoto, and the participants at SETA 2014, IAAE 2014 annual conference, and the conference participants at the University of Tokyo, Kyoto University, and the seminar participants at Hitotsubashi University. All errors are our responsibility.

†Yamazaki gratefully acknowledges the financial support by the JSPS Research Fellowship for Young Scientists. Address correspondence to Daisuke Yamazaki, Department of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan; E-mail: ed121003@g.hit-u.ac.jp.

‡Kurozumi's research was supported by JSPS KAKENHI Grant Numbers 23243038 and 25285067. E-mail: kurozumi@stat.hit-u.ac.jp.

1 Introduction

Testing for structural breaks has been a longstanding problem and various tests have been proposed in the econometric and statistical literature. One of the frequently used tests for parameter constancy against the general alternative is the CUSUM test based on recursive residuals proposed by Brown, Durbin, and Evans (1975), and this test was further developed based on OLS residuals by Ploberger and Krämer (1992). By specifying a random walk as the alternative, optimal tests for parameter constancy were investigated by Nyblom and Mäkeläinen (1983), Nyblom (1986, 1989), and Nabeya and Tanaka (1988), among others, while the point optimal test for general regression models was studied by Elliott and Müller (2006). On the other hand, it is often the case that a one-time structural change with an unknown change point is considered as the alternative and the sup-type test by Andrews (1993) and the mean- and exponential-type tests developed by Andrews and Ploberger (1994) and Andrews, Lee, and Ploberger (1996) are widely used in practical analyses. For a general discussion on structural changes, see, for example, Csörgő and Horváth (1997), Perron (2006), and Aue and Horváth (2013).

In practice, when we test for structural breaks in time-series models, we need to take serial correlation into account, and thus we have to estimate the long-run variance of the errors. If we estimate the long-run variance under the null hypothesis of no structural breaks, then it is known that the above tests suffer from the so-called non-monotonic power problem, that is, the power initially rises under the alternative, but as the magnitude of the break increases, the power eventually falls and tends to zero. This problem was investigated by Vogelsang (1999), Crainiceanu and Vogelsang (2007), Deng and Perron (2008), and Perron and Yamamoto (2014). The reason for this problem is that the long-run variance estimator takes significantly large values as the magnitude of the break increases.

On the other hand, if we estimate the long-run variance under the alternative, then the tests suffer from size distortion; they tend to over-reject the null hypothesis. This is because the long-run variance is under-estimated, so that the test statistics tend to take large values under the null hypothesis of no break.

In order to cope with the problem associated with the estimation of the long-run variance, several methods have been proposed. Kejriwal (2009) proposed to estimate the long-run variance using the residuals under both the null and alternative hypotheses. By using this hybrid estimator, we can reduce size distortion, but the power becomes extremely low when the error is strongly serially correlated. Juhl and Xiao (2009) proposed to estimate the long-run variance using the residuals of the nonparametric regression to mitigate the non-monotonic power problem. However, the finite sample performance of this test crucially depends on the choice of the bandwidth in the nonparametric regression. While these papers tried to improve the accuracy of the long-run variance estimator, there are several methods with which we do not have to consistently estimate the long-run variance. Sayginsoy and Vogelsang (2011) and Yang and Vogelsang (2011) proposed fixed-b sup-Wald and fixed-b sup-LM tests, respectively, which are robust to I(0)/I(1) errors. The fixed-b framework is based on Kiefer and Vogelsang (2005), which used an inconsistent long-run variance estimator where the bandwidth is proportional to the sample size. The fixed-b sup-Wald and sup-LM tests have relatively good sizes under the null hypothesis, but there is a loss of power due to the inconsistent estimation of the long-run variance. On the other hand, Shao and Zhang (2010) proposed a self-normalized test based on the CUSUM test. The basic idea of self-normalization is similar to the fixed-b approach. Although the finite sample performance of these tests are improved, compared to the frequently used tests, such as the original CUSUM and sup-type tests, the existing methods do not seem to be satisfactory in terms of both size and power.

In this paper, we develop an accurate long-run variance estimator and propose to use it to improve the finite sample property of the structural change tests. This estimator can be obtained by correcting the bias up to $O(T^{-1})$, where T is the sample size. The key feature of our method is that bias correction is achieved by taking a structural break into account. The advantage of our method is that tests with our long-run variance estimator can control the empirical size well, while maintaining high power. The simulation results show that the proposed tests have a higher power than other tests, such as the fixed-b test. Moreover, the power difference between our bias-corrected tests and the original (bias-uncorrected) tests is very minor, and it becomes negligible as the sample size increases. This result is in contrast to some other tests, which suffer from asymptotic power loss.

The remainder of this paper is organized as follows. In Section 2, we introduce the model and the test statistic. The derivation of the bias term is discussed in Section 3, and the bias correction method is explained in Section 4. The case with general error processes is discussed in Section 5. Simulation results are given in Section 6, and Section 7 concludes the paper. All mathematical proofs are delegated to the appendix.

2 Model and Test Statistic

Let us consider the following mean-shift model:

$$y_t = \mu + \delta \cdot DU_t(T_b^0) + u_t, \quad t = 1, \cdots, T, \tag{1}$$

where $DU_t(T_b^0) = 1\{t > T_b^0\}$, and $1\{\cdot\}$ is the indicator function. We assume that u_t is a zero-mean stationary process and that the break date T_b^0 is unknown.

The testing problem is

$$H_0 : \quad \delta = 0 \qquad \text{vs.} \qquad H_1 : \quad \delta \neq 0. \tag{2}$$

Under H_0, there is no shift in mean, whereas under H_1, there is a one-time break.

In order to test for a shift in mean, we need to estimate the long-run variance of u_t defined by $\omega = \sum_{\ell=-\infty}^{\infty} E(u_t u_{t-\ell})$ for the scale adjustment, which can be consistently estimated by the kernel method. As it is known that tests with ω estimated under the null hypothesis suffer from the non-monotonic power problem, as pointed out by Vogelsang (1999), we exclude the case where the long-run variance is estimated under the null hypothesis, and focus on the case where it is estimated under the alternative of a one-time break. That is, we consider the following kernel estimator of ω as a benchmark:

$$\hat{\omega}(T_b) = \hat{\gamma}_0 + 2 \sum_{j=1}^{T-1} k\left(\frac{j}{m}\right) \hat{\gamma}_j, \tag{3}$$

where $k(\cdot)$ is the kernel function, m is the bandwidth, $\hat{\gamma}_j$ is the estimator of the jth autocovariance of u_t defined by $\hat{\gamma}_j = T^{-1} \sum_{t=j+1}^{T} \hat{u}_t \hat{u}_{t-j}$, the residuals $\hat{u}_t$ are obtained under the alternative with the supposed break date T_b, and

$$\hat{u}_t = \begin{cases} y_t - \bar{y}_1 & \text{for } t = 1, \cdots, T_b, \\ y_t - \bar{y}_2 & \text{for } t = T_b + 1, \cdots, T, \end{cases} \tag{4}$$

where $\bar{y}_1 = T_b^{-1} \sum_{t=1}^{T_b} y_t$ and $\bar{y}_2 = (T - T_b)^{-1} \sum_{t=T_b+1}^{T} y_t$. Note that T_b is specified by a researcher and it is not necessarily consistent with T_b^0. We suppress the dependency of $\hat{\gamma}_j$ and $\hat{u}_t$ on T_b for notational simplicity.

When the parametric structure is framed for u_t, we may use, instead of the kernel estimator, the autoregressive spectral density estimator of ω based on the AR(p) model given by

$$\hat{\omega}_{AR}(T_b) = \frac{\hat{\sigma}_\varepsilon^2}{\left(1 - \sum_{j=1}^{p} \hat{\phi}_j\right)^2}, \tag{5}$$

where $\hat{u}_t = \sum_{j=1}^{p} \hat{\phi}_j \hat{u}_{t-j} + \hat{\varepsilon}_t$ with $\hat{\phi}_j$ $(j = 1, \cdots, p)$ being the OLS estimator, and $\hat{\sigma}_\varepsilon^2 = (T - p)^{-1} \sum_{t=p+1}^{T} \hat{\varepsilon}_t^2$.

In this paper, we mainly consider the following two structural change tests, which have been commonly used in many practical analyses, with $\hat{\omega}^*(T_b)$ denoting either $\hat{\omega}(T_b)$ in (3) or $\hat{\omega}_{AR}(T_b)$ in (5), as the estimator of ω.

Sup-Wald test

Following Andrews (1993), the sup-Wald statistic for testing problem (2) is given by

$$\text{sup-}W = \max_{T_b \in [\varepsilon T, (1-\varepsilon)T]} W(T_b), \quad \text{where} \quad W(T_b) = \frac{SSR_0 - SSR(T_b)}{\hat{\omega}^*(T_b)}, \tag{6}$$

where SSR_0 is the sum of squared residuals under H_0, $SSR(T_b)$ is the sum of squared residuals under the alternative of a one-time break with the break date T_b, and ε is the trimming parameter.

CUSUM test

The CUSUM test statistic proposed by Ploberger and Krämer (1992) is originally defined as

$$CUSUM = \max_{T_b \in [1, T-1]} \left| \frac{T^{-1/2} \sum_{t=1}^{T_b} \tilde{u}_t}{\sqrt{\tilde{\omega}}} \right|,$$

where $\tilde{u}_t$ is the residual under H_0, and the long-run variance estimator $\tilde{\omega}$ is estimated under the null hypothesis. As explained in Crainiceanu and Vogelsang (2007) and Deng and Perron (2008), this test suffers from the non-monotonic power problem because the long-run variance is estimated under the null hypothesis of no break. In order to avoid this problem, we again consider estimating the long-run variance under the alternative of a one-time break. Then, the test statistic should be modified as

$$CUSUM_{H_1} = \max_{T_b \in [1, T-1]} \left| \frac{T^{-1/2} \sum_{t=1}^{T_b} \tilde{u}_t}{\sqrt{\hat{\omega}^*(T_b)}} \right|. \tag{7}$$

3 Derivation of the Bias

In this section, we derive the bias of the reciprocal of the long-run variance estimator up to $O(T^{-1})$, ignoring the $o_p(T^{-1})$ terms under the assumption that the correct specification for u_t is the AR(p) model. The case with general error processes will be discussed later. Note that since our purpose is to control the size of the tests by precisely estimating the long-run variance, the bias is derived under the null hypothesis of no break, whereas $\hat{u}_t$ is obtained assuming a one-time break at T_b, which is given by

$$\hat{u}_t = \begin{cases} u_t - \bar{u}_1 & \text{for } t = 1, \cdots, T_b, \\ u_t - \bar{u}_2 & \text{for } t = T_b + 1, \cdots, T. \end{cases} \tag{8}$$

To derive the bias term, we make the following assumptions when $p \geq 1$:

Assumption 1 *$\{u_t\}$ follows a zero-mean stationary AR(p) process: $u_t = \sum_{j=1}^{p} \phi_j u_{t-j} + \varepsilon_t$, where $1 - \sum_{j=1}^{p} \phi_j z^j \neq 0$ for $|z| \leq 1$, and $\{\varepsilon_t\}$ is a martingale difference sequence with a finite 4th moment, which satisfies $E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma_\varepsilon^2$ and $E(\varepsilon_t^3 | \mathcal{F}_{t-1}) = \kappa_3$.*

Assumption 2 *$T_b/T \to \lambda \in (0, 1)$ as $T \to \infty$.*

When $p = 0$, we use the following Assumption 1', instead of Assumption 1.

Assumption 1' *$u_t = \varepsilon_t$ for all t, where $\{\varepsilon_t\}$ is a martingale difference sequence with a finite 4th moment, which satisfies $E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma_\varepsilon^2$.*

Assumptions 1 and 1' exclude the case where $\{u_t\}$ is a unit root process. Assumption 2 is standard for structural break models.

3.1 Bias of the OLS estimator of the autoregressive coefficients

First, we derive the bias of the OLS estimator of ϕ_j $(j = 1, \cdots, p)$ for $p \geq 1$, which is given by

$$\hat{\phi} = \begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{bmatrix} = \begin{bmatrix} \hat{r}_{11} & \hat{r}_{12} & \cdots & \hat{r}_{1p} \\ \hat{r}_{21} & \hat{r}_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \hat{r}_{p-1,p} \\ \hat{r}_{p1} & \cdots & \hat{r}_{p,p-1} & \hat{r}_{pp} \end{bmatrix}^{-1} \begin{bmatrix} \hat{r}_{10} \\ \hat{r}_{20} \\ \vdots \\ \hat{r}_{p0} \end{bmatrix},$$

where $\hat{r}_{ij} = (T-p)^{-1} \sum_{t=p+1}^{T} \hat{u}_{t-i}\hat{u}_{t-j}$.

In order to derive the bias of $\hat{\phi}$, we define the following three $(p+1) \times (p+1)$ matrices, based on Stine and Shaman (1989) and Patterson (2000):

$$\begin{aligned} B_{1p} &= \mathrm{diag}\{0, 1, \cdots, p\}, \\ B_{2p} &= \begin{cases} \left[-e_0, -e_1, \cdots, -e_{\frac{p}{2}-1}, 0_{(p+1)\times 1}, e_{\frac{p}{2}-1}, \cdots, e_1, e_0\right] & \text{when } p \text{ is even,} \\ \left[-d_1, -d_2, \cdots, -d_{\frac{p-1}{2}}, 0_{(p+1)\times 1}, d_{\frac{p-1}{2}}, \cdots, d_1, d_0\right] & \text{when } p \text{ is odd,} \end{cases} \\ (B_{3p})_{ij} &= \begin{cases} -1 & \text{for } j < i \leq p-j+2, \\ 1 & \text{for } p-j+2 < i \leq j, \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

where $0_{(p+1)\times 1}$ is a $(p+1) \times 1$ vector of zeros, e_j is a $(p+1) \times 1$ vector with ones in rows $j+3, j+5, \cdots, p+1-j$ and zeros elsewhere, and d_j is a $(p+1) \times 1$ vector with ones in rows $j+2, j+4, \cdots, p+1-j$ and zeros elsewhere. For example, $d_0 = [0, 1, 0, 1]'$ and $d_1 = [0, 0, 1, 0]'$ for $p = 3$, while $e_0 = [0, 0, 1, 0, 1]'$ and $e_1 = [0, 0, 0, 1, 0]'$ for $p = 4$.

Let $D_p = B_{1p} + B_{2p} + 2B_{3p}$, and we divide D_p into four blocks as follows:

$$D_p = \left[\begin{array}{c|c} 0_{1\times 1} & 0_{1\times p} \\ \hline -K_p & B_p \end{array}\right]. \tag{9}$$

where K_p and B_p are $p \times 1$ and $p \times p$, respectively, that is, K_p is (-1) times the $p \times 1$ lower-left block element of D_p, and B_p is the $p \times p$ lower-right block element of D_p. The values of K_p and B_p for $p = 1, \cdots, 5$ are given in Table 1.

The following theorem gives the bias of the OLS estimator $\hat{\phi}$.

Theorem 1 *Under Assumptions 1 and 2, the expectation of the OLS estimator* $\hat{\phi}$ *up to* $O(T^{-1})$ *is given by*

$$E(\hat{\phi}) = \phi - \frac{1}{T-p}(K_p + B_p\phi) + o(T^{-1}), \tag{10}$$

where $\phi = [\phi_1, \cdots, \phi_p]'$.

Remark 1 *The first-order bias of the OLS estimator does not depend on the maintained break fraction* $\lambda = \lim_{T\to\infty} T_b/T$.

Remark 2 *When* $p = 1$, *by Theorem 1, the expectation of the OLS estimator with a one-time structural break in mean reduces to*

$$E(\hat{\phi}_1) = \phi_1 - \frac{1}{T-1}(2 + 4\phi_1) + o(T^{-1}),$$

whereas the well-known bias formula without a structural break is

$$E(\hat{\phi}_1) = \phi_1 - \frac{1}{T-1}(1 + 3\phi_1) + o(T^{-1}).$$

Hence, when $\phi_1 > 0$, *we can see that the OLS estimator with a break has a larger downward bias than the one without a break, which also leads to a downward bias in* (5).

Remark 3 *This result can be easily extended to the case with multiple structural breaks. In this case, we obtain the residuals assuming structural breaks at* $t = T_1, \cdots, T_m$, *where* m *is the number of structural breaks, such that* $\lim_{T\to\infty} T_i/T = \lambda_i$ *and* $0 < \lambda_1 < \cdots < \lambda_m < 1$, *and the residual is given by*

$$\hat{u}_t = u_t - \bar{u}_i \quad \text{for} \quad t = T_{i-1}+1, \cdots, T_i \quad (i = 1, \cdots, m+1), \tag{11}$$

where $\bar{u}_i = (T_i - T_{i-1})^{-1} \sum_{t=T_{i-1}+1}^{T_i} u_t$, $T_0 = 0$, *and* $T_{m+1} = T$. *Then, the expectation of* $\hat{\phi}$ *up to* $O(T^{-1})$ *is given by*

$$E(\hat{\phi}) = \phi - \frac{1}{T-p}\left(K_p^{(m)} + B_p^{(m)}\phi\right) + o(T^{-1}),$$

where $D_p^{(m)} = B_{1p} + B_{2p} + (m+1)B_{3p}$, *and* $K_p^{(m)}$ *and* $B_p^{(m)}$ *are defined in the same way as in* (9). *We do not give the proof of the above result in detail because we consider only a one-time break in this paper.*

3.2 Bias of the reciprocal of the long-run variance estimator

Next, we derive the bias of the reciprocal of $\hat{\omega}_{AR}$, which is given by

$$\frac{1}{\hat{\omega}_{AR}} = \begin{cases} \dfrac{\left(1-\sum_{j=1}^{p}\hat{\phi}_j\right)^2}{\hat{\sigma}_\varepsilon^2} & \text{for } p \geq 1, \\ \dfrac{1}{\hat{\sigma}_\varepsilon^2} & \text{for } p = 0. \end{cases} \tag{12}$$

Here, we consider the bias of the reciprocal of $\hat{\omega}_{AR}$ because the long-run variance estimator is placed in the denominator of the test statistics.

In general, when random variables X and Y satisfy $X - E(X) = O_p(T^{-1/2})$, $Y - E(Y) = O_p(T^{-1/2})$, $E(X) \neq 0$, and $E(Y) \neq 0$, the following relation holds:

$$E\left(\frac{X}{Y}\right) = \frac{E(X)}{E(Y)}\left[1 - \frac{Cov(X,Y)}{E(X)E(Y)} + \frac{Var(Y)}{\{E(Y)\}^2}\right] + o(T^{-1}), \tag{13}$$

which can be obtained by the Taylor expansion of $f(x,y) = x/y$ around $(x,y) = (E(X), E(Y))$, and by taking expectations, ignoring the $o_p(T^{-1})$ terms. See Mood, Graybill, and Boes (1974, p.181).

Therefore, in order to derive the bias of (12) up to $O(T^{-1})$, we need to obtain $E[(1 - \sum_{j=1}^{p}\hat{\phi}_j)^2]$, $E[\hat{\sigma}_\varepsilon^2]$, $Var[\hat{\sigma}_\varepsilon^2]$, and $Cov[\hat{\sigma}_\varepsilon^2, (1-\sum_{j=1}^{p}\hat{\phi}_j)^2]$ for $p \geq 1$. When $p = 0$, we only need $E[\hat{\sigma}_\varepsilon^2]$ and $Var[\hat{\sigma}_\varepsilon^2]$.

The following lemma gives the results for $p \geq 1$:

Lemma 1 *Under Assumptions 1 and 2, the following relations hold:*

(a) $E\left[\left(1-\sum_{j=1}^{p}\hat{\phi}_j\right)^2\right] = (1-\iota'\phi)^2 + \dfrac{1}{T-p}\left\{2(1-\iota'\phi)\iota'(K_p + B_p\phi) + \sigma_\varepsilon^2\iota' R^{-1}\iota\right\} + o(T^{-1}),$

(b) $E\left[\hat{\sigma}_\varepsilon^2\right] = \sigma_\varepsilon^2 - \dfrac{p+2}{T-p}\sigma_\varepsilon^2 + o(T^{-1}),$

(c) $Var\left[\hat{\sigma}_\varepsilon^2\right] = \dfrac{1}{T-p}\left\{E(\varepsilon_t^4) - \sigma_\varepsilon^4\right\} + o(T^{-1}),$

(d) $Cov\left[\hat{\sigma}_\varepsilon^2, \left(1-\sum_{j=1}^{p}\hat{\phi}_j\right)^2\right] = o(T^{-1}),$

where R is a $p \times p$ matrix whose (i,j) element is given by $\gamma_{|i-j|} = E(u_t u_{t-|i-j|})$, and ι is a $p \times 1$ vector of ones.

By (13) and Lemma 1, we obtain the first-order bias of the reciprocal of the long-run variance estimator for $p \geq 1$:

Theorem 2 *Under Assumptions 1 and 2, the expectation of* $1/\hat{\omega}_{AR}$ *up to* $O(T^{-1})$ *is given by*

$$E\left[\frac{1}{\hat{\omega}_{AR}}\right] = \frac{(1-\iota'\phi)^2}{\sigma_\varepsilon^2} + \frac{1}{T-p}\left[\frac{1}{\sigma_\varepsilon^2}\left\{2(1-\iota'\phi)\iota'(K_p + B_p\phi) + \sigma_\varepsilon^2\iota' R^{-1}\iota + (p+2)(1-\iota'\phi)^2\right\} + \frac{(1-\iota'\phi)^2}{\sigma_\varepsilon^2}\left\{\frac{E(\varepsilon_t^4)}{\sigma_\varepsilon^4} - 1\right\}\right] + o(T^{-1}).$$

Remark 4 *When* $p = 1$, *the expectation of* $1/\hat{\omega}_{AR}$ *is given by*

$$E\left[\frac{1}{\hat{\omega}_{AR}}\right] = \frac{(1-\phi_1)^2}{\sigma_\varepsilon^2} + \frac{1}{T-1}\cdot\frac{1}{\sigma_\varepsilon^2}\left[(1-\phi_1)(8+6\phi_1) + (1-\phi_1)^2\left\{\frac{E(\varepsilon_t^4)}{\sigma_\varepsilon^4} - 1\right\}\right] + o(T^{-1}),$$

if the long-run variance is estimated using the residuals under the alternative.

On the other hand, if we use the residuals under the null hypothesis of no structural break to estimate the long-run variance, the expectation can be shown to be given by

$$E\left[\frac{1}{\hat{\omega}_{AR}}\right] = \frac{(1-\phi_1)^2}{\sigma_\varepsilon^2} + \frac{1}{T-1}\cdot\frac{1}{\sigma_\varepsilon^2}\left[(1-\phi_1)(5+5\phi_1) + (1-\phi_1)^2\left\{\frac{E(\varepsilon_t^4)}{\sigma_\varepsilon^4} - 1\right\}\right] + o(T^{-1}).$$

Therefore, we can see that the first-order bias of $1/\hat{\omega}_{AR}$ *with the residuals under the alternative hypothesis is larger than the one with the residuals under the null hypothesis.*

Similarly, when $p = 0$, we obtain the following lemma and theorem:

Lemma 1' *Under Assumptions 1' and 2, the following relations hold:*

$$(a)\quad E\left[\hat{\sigma}_\varepsilon^2\right] = \sigma_\varepsilon^2 - \frac{2}{T}\sigma_\varepsilon^2 + o(T^{-1}),$$

$$(b)\quad Var\left[\hat{\sigma}_\varepsilon^2\right] = \frac{1}{T}\left\{E(\varepsilon_t^4) - \sigma_\varepsilon^4\right\} + o(T^{-1}).$$

Theorem 2' *Under Assumptions 1' and 2, the expectation of* $1/\hat{\omega}_{AR}$ *up to* $O(T^{-1})$ *is given by*

$$E\left[\frac{1}{\hat{\omega}_{AR}}\right] = \frac{1}{\sigma_\varepsilon^2} + \frac{1}{T}\left[\frac{2}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\varepsilon^2}\left\{\frac{E(\varepsilon_t^4)}{\sigma_\varepsilon^4} - 1\right\}\right] + o(T^{-1}).$$

Remark 5 *The first-order bias of* $1/\hat{\omega}_{AR}$ *does not depend on the maintained break fraction* λ.

4 Bias-Corrected Test

In this section, we propose the correction of the bias of (12) using Theorems 2 and 2', and explain how to use our bias-corrected estimator in order to test for a shift in mean.

4.1 Bias correction of the reciprocal of the long-run variance estimator

In this subsection, we obtain the bias-corrected estimator of the reciprocal of the long-run variance.

Since the first-order bias of (12) is given by Theorems 2 and 2', the bias-corrected estimator of $1/\omega_{AR}$ is

$$\left(\frac{1}{\hat{\omega}_{AR}}\right)_{BC} = \frac{1}{\hat{\omega}_{AR}} - \hat{b}, \tag{14}$$

where

$$\hat{b} = \begin{cases} \dfrac{1}{T-p}\left[\dfrac{1}{\hat{\sigma}_\varepsilon^2}\left\{2(1-\iota'\hat{\phi})\iota'(K_p + B_p\hat{\phi}) + \hat{\sigma}_\varepsilon^2 \iota'\hat{R}^{-1}\iota \right.\right. \\ \qquad \left.\left. +(p+2)(1-\iota'\hat{\phi})^2\right\} + \dfrac{(1-\iota'\hat{\phi})^2}{\hat{\sigma}_\varepsilon^2}\left\{\dfrac{\widehat{E(\varepsilon_t^4)}}{\hat{\sigma}_\varepsilon^4} - 1\right\}\right] & \text{for } p \geq 1, \\ \dfrac{1}{T}\left[\dfrac{2}{\hat{\sigma}_\varepsilon^2} + \dfrac{1}{\hat{\sigma}_\varepsilon^2}\left\{\dfrac{\widehat{E(\varepsilon_t^4)}}{\hat{\sigma}_\varepsilon^4} - 1\right\}\right] & \text{for } p = 0, \end{cases}$$

and $\hat{\phi}$, $\hat{\sigma}_\varepsilon^2 = (T-p)^{-1}\sum_{t=p+1}^T \hat{\varepsilon}_t^2$, $\widehat{E(\varepsilon_t^4)} = (T-p)^{-1}\sum_{t=p+1}^T \hat{\varepsilon}_t^4$, and $\hat{\gamma}_{ij}$ for the (i,j) element of $\hat{R}$ are the least squares estimators of ϕ, σ_ε^2, $E(\varepsilon_t^4)$, and γ_{ij}, respectively.[1]

For example, when $p = 1$, the correcting term is given by

$$\hat{b} = \frac{1}{T-1}\cdot\frac{1}{\hat{\sigma}_\varepsilon^2}\left[(1-\hat{\phi})(8+6\hat{\phi}) + (1-\hat{\phi})^2\left\{\frac{\widehat{E(\varepsilon_t^4)}}{\hat{\sigma}_\varepsilon^4} - 1\right\}\right].$$

4.2 Tests based on the bias-corrected long-run variance estimator

The bias-corrected test statistic can be obtained by using the bias-corrected estimator (14). For example, the bias-corrected sup-Wald test statistic is given by

$$\text{sup-}W_{BC} = \max_{T_b \in [\varepsilon T, (1-\varepsilon)T]} W_{BC}(T_b), \tag{15}$$

[1] Other consistent estimators can also be plugged in.

where

$$W_{BC}(T_b) = \left(\frac{1}{\hat{\omega}_{AR}}\right)_{BC} \cdot (SSR_0 - SSR(T_b)).$$

Similarly, the bias-corrected CUSUM test statistic[2] is given by

$$CUSUM_{H_1,BC} = \max_{T_b \in [\varepsilon T, (1-\varepsilon)T]} \left| \sqrt{\left(\frac{1}{\hat{\omega}_{AR}}\right)_{BC}} \cdot T^{-1/2} \sum_{t=1}^{T_b} \tilde{u}_t \right|. \tag{16}$$

Since the correcting terms are $O_p(T^{-1})$, the asymptotic distribution of the test statistic under the null hypothesis is exactly the same as that of the original test, and thus we do not have to modify critical values in order to apply the bias-corrected test. Moreover, even under the alternative, it can be shown that the first-order bias is asymptotically negligible, so that there is no asymptotic power loss.

5 Extension to the Model with General Error Processes

In this section, we consider the case where the error term u_t is generated by a stationary AR(∞) process. In this case, we make the following assumption:

Assumption 1" *$u_t = \sum_{j=1}^{\infty} \phi_j u_{t-j} + \varepsilon_t$, where $1 - \sum_{j=1}^{\infty} \phi_j z^j \neq 0$ for $|z| \leq 1$, $\sum_{j=1}^{\infty} |\phi_j| < \infty$, and $\{\varepsilon_t\}$ is a martingale difference sequence with a finite 4th moment, which satisfies $E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma_\varepsilon^2$ and $E(\varepsilon_t^3 | \mathcal{F}_{t-1}) = \kappa_3$.*

Although only the absolute summability of $\{\phi_j\}$ is assumed in Assumption 1", we may require the higher order summability of $\{\phi_j\}$, as explained below.

Since the error term is an infinite order AR process, we need to truncate the lag order at some point p_T and consider estimating the AR(p_T) model. The following assumption is concerned with the lag truncation point p_T.

Assumption L

$$(a) \quad p_T \to \infty \quad and \quad \tfrac{p_T^4}{T} \to 0 \quad as \quad T \to \infty. \tag{17}$$

$$(b) \quad \textstyle\sum_{j=p_T+1}^{\infty} |\phi_j| = o(p_T/T) \quad as \quad T \to \infty.$$

[2] We use a trimming for the CUSUM test so that Assumption 2 is satisfied.

Assumption L(a) gives the upper bound of the divergence rate of p_T. This rate guarantees the consistency of the autoregressive spectral density estimator as proved by Berk (1974) and den Haan and Levin (1998), although condition (17) is stronger than theirs. Assumption L(b) not only imposes the lower bound of p_T but is also related with the higher order summability of $\{\phi_j\}$. For example, when $\sum_{j=0}^{\infty} j^{3+\alpha}|\phi_j| < \infty$ holds and p_T is greater than $O(T^{1/(4+\alpha)})$ for some $\alpha > 0$, Assumption L(b) is satisfied. Note that this assumption is satisfied if u_t is generated by a finite-order ARMA process and $p_T = O(T^{\delta})$ for some $\delta > 0$, because $|\phi_j|$ declines geometrically to zero.

The next theorem gives the bias of the reciprocal of the autoregressive spectral density estimator up to $O(p_T/T)$:

Theorem 2" *Under Assumptions 1", 2, and L, the expectation of* $1/\hat{\omega}_{AR}$ *up to* $O(p_T/T)$ *is given by*

$$E\left[\frac{1}{\hat{\omega}_{AR}}\right] = \frac{(1-\iota'\phi)^2}{\sigma_\varepsilon^2} + \frac{1}{T-p_T}\left[\frac{1}{\sigma_\varepsilon^2}\left\{2(1-\iota'\phi)\iota'(K_{p_T}+B_{p_T}\phi) + \sigma_\varepsilon^2\iota' R^{-1}\iota + (p_T+2)(1-\iota'\phi)^2\right\} + \frac{(1-\iota'\phi)^2}{\sigma_\varepsilon^2}\left\{\frac{E(\varepsilon_t^4)}{\sigma_\varepsilon^4} - 1\right\}\right] + o\left(\frac{p_T}{T}\right),$$

where

$$\phi = \begin{bmatrix} \phi_{p_T,1} \\ \phi_{p_T,2} \\ \vdots \\ \phi_{p_T,p_T} \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p_T-1} \\ \gamma_1 & \gamma_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma_1 \\ \gamma_{p_T-1} & \cdots & \gamma_1 & \gamma_0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_{p_T} \end{bmatrix}. \quad (18)$$

This first-order bias is exactly the same as the one in Theorem 2. Therefore, we can implement the bias correction as explained in Section 4.

6 Simulation Results

In this section, we investigate the finite sample performance of the tests through a Monte Carlo experiment. The data generating process is as follows:

$$y_t = \mu + \delta \cdot DU_t(T_b^0) + u_t, \quad \mu = 0, \quad \delta = \frac{c}{\sqrt{T}}, \quad T_b^0 = 0.5T.$$

We consider the following three processes of u_t:

$$\begin{cases} \text{AR}(1) : u_t = \phi u_{t-1} + \varepsilon_t, & \varepsilon_t \sim i.i.d.\ N\left(0, 1-\phi^2\right), \\ \text{AR}(2) : u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \varepsilon_t, & \varepsilon_t \sim i.i.d.\ N\left(0, \frac{(1+\phi_2)\{(1-\phi_2)^2-\phi_1^2\}}{1-\phi_2}\right), \\ \text{MA}(1) : u_t = \varepsilon_t + \theta\varepsilon_{t-1}, & \varepsilon_t \sim i.i.d.\ N\left(0, \frac{1}{1+\theta^2}\right), \end{cases}$$

where the variance of ε_t is selected so that $Var(u_t) = 1$.

In subsections 6.1 and 6.2, we compare the sizes and powers of the following tests:

(sup-Wald test)

(i): sup-W: the sup-Wald test (6) with the long-run variance estimator given by (3).

(ii): sup-W_{AR}: the sup-Wald test (6) with the long-run variance estimator given by (5).

(iii): sup-W_{BC}: the bias-corrected sup-Wald test (15).

(iv): sup-W_{kej}: the sup-Wald test (6) with the hybrid long-run variance estimator by Kejriwal (2009).

(v): fixed-b sup-W: the fixed-b sup-Wald test based on Sayginsoy and Vogelsang (2011), where we use the J statistic as a scaling factor. We use the Daniell kernel with the feasible integrated power optimal data-dependent bandwidth as described in Sayginsoy and Vogelsang (2011), and a 10% trimming for this test.

(CUSUM test)

(i): $CUSUM_{H_1}$: the CUSUM test with a 15% trimming, which is given by

$$CUSUM_{H_1} = \max_{T_b \in [0.15T, 0.85T]} \left| \frac{T^{-1/2} \sum_{t=1}^{T_b} \tilde{u}_t}{\sqrt{\hat{\omega}^*(T_b)}} \right|. \tag{19}$$

We use the long-run variance estimator given by (3).

(ii): $CUSUM_{H_1,AR}$: the CUSUM test (19) with the long-run variance estimator given by (5).

(iii): $CUSUM_{H_1,BC}$: the bias-corrected CUSUM test (16).

(iv): SN: the self-normalizing method by Shao and Zhang (2010).

For the kernel estimator (3), we use the quadratic spectral kernel with the bandwidth parameter selected by Andrews' (1991) rule to estimate the long-run variance, except for the fixed-b sup-Wald test. When we implement the AR(p) regression to obtain the autoregressive spectral density estimator, we select the lag length p by the Bayesian Information Criterion (BIC), where the maximum lag length is 5. For the sup-Wald and CUSUM tests, we use a 15% trimming, except for the fixed-b sup-Wald test. The number of replications is 2,000, and the nominal size is 0.05.

6.1 Empirical sizes of the tests

Tables 2-5 show the empirical sizes of the tests. When the error follows an AR(1) process, we can see from Table 2 that the original sup-Wald test tends to over-reject the null hypothesis as ϕ gets larger. By using the autoregressive spectral density estimator, we can mitigate the over-rejection problem, except the case where $\phi = 0.2$, but the test still has size distortion. We need to note that, when $\phi = 0.2$, the sup-W_{AR} test has a larger size distortion than the original sup-Wald test because the lag length selected by the BIC is sometimes too short in finite samples. The bias-corrected sup-Wald test performs much better than the bias-uncorrected tests, in particular when u_t is strongly serially correlated. The empirical sizes of the sup-Wald test based on Kejriwal (2009) and the fixed-b sup-Wald test are relatively close to the nominal one, although the fixed-b test is rather conservative. We observe similar results for the CUSUM test. The bias-corrected CUSUM test ($CUSUM_{H_1,BC}$) has much less size distortion than the bias-uncorrected CUSUM tests ($CUSUM_{H_1}$ and $CUSUM_{H_1,AR}$), unless $\phi = 0.2$. Moreover, the $CUSUM_{H_1,BC}$ test performs better than the self-normalization based test when ϕ is large. As the sample size increases, the sizes of all tests get closer to the nominal one.

Tables 3 and 4 show the empirical sizes with AR(2) errors. We can see that the relative performance holds when $\phi_2 = -0.3$, compared to the case with AR(1), whereas when $\phi_2 = 0.3$ and $T = 100$, all the tests tend to over-reject the null hypothesis, including the bias-corrected tests. In this case, only the fixed-b sup-Wald test has relatively good size. However, as the sample size increases, the performance of the bias-corrected tests greatly improves, and it is superior to that of the other tests. When the error follows an MA(1) process, we can see from Table 5 that the bias-corrected tests have good finite sample properties.

6.2 Size-adjusted power of the tests

In this subsection, we compare the size-adjusted power of the tests.[3][4] Figure 1 shows the size-adjusted powers with AR(1) errors and $T = 100$. We can see from Figure 1 that, when $\phi = 0.6$, the bias-corrected sup-Wald test is more powerful than the sup-W_{kej} and fixed-b sup-Wald tests, while for the CUSUM test, the bias-corrected test performs much better than the self-normalization based test. We can see that the power difference between the bias-corrected and bias-uncorrected tests is relatively small. Similar results are obtained when $\phi = 0.8$. Although the power loss due to bias correction is slightly larger than that of the case with $\phi = 0.6$, the bias-corrected test has higher power than the other tests.

As in Figure 2, when $T = 200$, the power difference between the bias-corrected and bias-uncorrected tests is much smaller than the case when $T = 100$. In this case, the bias-corrected test still outperforms the other tests.

6.3 Comparison of the finite sample performance of bias-corrected tests

In this subsection, we focus on only the bias-corrected versions of the tests commonly used in the literature and compare their sizes and powers. We consider the sup-Wald test (sup-W_{BC}) by Andrews (1993), the mean-Wald test (mean-W_{BC}) and the exponential-Wald test (exp-W_{BC}) by Andrews, Lee, and Ploberger (1996), the locally best invariant test against the random walk alternative by Nabeya and Tanaka (1988) (which we denote as $LM_{H_1,BC}$), the asymptotically point optimal test against the random walk alternative by Elliott and Müller (2006) (which we denote as $qLL_{H_1,BC}$), and the $CUSUM_{H_1,BC}$ test given by (19). Since the original LM, qLL, and CUSUM tests use the long-run variance estimator under the null hypothesis and they have non-monotonic power, we consider estimating the long-run variance under the alternative of a one-time break. For the LM and qLL tests, we use the residuals under the alternative with break date $\hat{T}_b = \arg\min_{T_b \in [0.15T, 0.85T]} SSR(T_b)$. For the CUSUM test, we use the bias-corrected test statistic (16).

The empirical sizes with AR(1) errors are given in Table 6 (we omit the other cases to

[3]Because the critical value of the fixed-b sup-Wald test is data-dependent, we adjust the size of the other tests to the empirical size of the fixed-b sup-Wald test with nominal one 0.05.

[4]Since the size-adjusted powers of the sup-W_{AR} and $CUSUM_{H_1,AR}$ tests are almost the same as those of the sup-W and $CUSUM_{H_1}$ tests, respectively, we omit the results of sup-W_{AR} and $CUSUM_{H_1,AR}$ tests.

save space). We observe that the bias-corrected mean-Wald test has relatively good size, while the other tests are slightly over-sized.

The size-adjusted powers of the tests are given in Figure 3. We observe that the bias-corrected CUSUM test performs best, while the mean-Wald, LM, and qLL tests suffer from power loss, in particular when the errors are strongly serially correlated, or when the sample size is small.

Overall, we can see that the bias-corrected CUSUM test with the long-run variance estimated under the alternative has the best finite sample properties, against the alternative of a one-time break. However, it is not clear whether this bias-corrected CUSUM test outperforms other tests against various kinds of the alternative, such as multiple breaks or time-varying parameter models. This is our further work.

7 Conclusion

We have proposed a bias correction to the long-run variance estimator, which is estimated under the alternative hypothesis of a one-time break. We have derived the first-order bias of the reciprocal of the long-run variance estimator, taking a structural break into account. By Monte Carlo simulations, we have found that our bias-corrected tests have better finite sample properties than the existing tests.

So far, we have considered tests for a mean shift, but it is also in our interest to consider bias correction to test for structural change in general regression models. We wish to investigate such topics in future studies.

Appendix A: Proofs of Theorem 1 and Some Related Lemmas

Lemma 2 *Under Assumptions 1 and 2,*

$$E(\hat{\phi}) = \phi + R^{-1}E\left[\hat{r} - r - (\hat{R} - R)\phi\right] - R^{-1}E\left[(\hat{R} - R)R^{-1}\left\{\hat{r} - r - (\hat{R} - R)\phi\right\}\right] + o(T^{-1}), \tag{20}$$

where $\hat{R}$ *and* R *are* $p \times p$ *matrices such that* $(\hat{R})_{ij} = \hat{r}_{ij}$, $(R)_{ij} = r_{ij}$, $\hat{r} = [\hat{r}_{01}, \cdots, \hat{r}_{0p}]'$, $r = [r_{01}, \cdots, r_{0p}]'$, $\hat{r}_{ij} = (T-p)^{-1}\sum_{t=p+1}^{T} \hat{u}_{t-i}\hat{u}_{t-j}$, *and* $r_{ij} = E(u_{t-i}u_{t-j})$.

Proof of Lemma 2

Since $\hat{R}^{-1}$ can be expressed as

$$\hat{R}^{-1} = R^{-1} - R^{-1}(\hat{R} - R)\hat{R}^{-1}, \tag{21}$$

we obtain

$$\begin{aligned} \hat{R}^{-1} &= R^{-1} - R^{-1}(\hat{R} - R)R^{-1} + R^{-1}(\hat{R} - R)R^{-1}(\hat{R} - R)R^{-1} \\ &\quad -R^{-1}(\hat{R} - R)R^{-1}(\hat{R} - R)R^{-1}(\hat{R} - R)\hat{R}^{-1}, \end{aligned} \tag{22}$$

by recursively using relation (21). Therefore, since $\phi = R^{-1}r$, $\hat{r} - r = O_p(T^{-1/2})$, and $\hat{R} - R = O_p(T^{-1/2})$, we have

$$\begin{aligned} \hat{\phi} &= \hat{R}^{-1}\hat{r} \\ &= R^{-1}r + R^{-1}(\hat{r} - r) - R^{-1}(\hat{R} - R)R^{-1}r - R^{-1}(\hat{R} - R)R^{-1}(\hat{r} - r) \\ &\quad +R^{-1}(\hat{R} - R)R^{-1}(\hat{R} - R)R^{-1}r + o_p(T^{-1}) \\ &= \phi + R^{-1}\left\{\hat{r} - r - (\hat{R} - R)\phi\right\} \\ &= -R^{-1}(\hat{R} - R)R^{-1}\left\{\hat{r} - r - (\hat{R} - R)\phi\right\} + o_p(T^{-1}). \end{aligned} \tag{23}$$

By ignoring the $o_p(T^{-1})$ term and taking expectation of the rest of (23), we obtain (20).
■

Lemma 3 *Under Assumptions 1 and 2,*

$$E(\hat{r}_{ij} - r_{ij}) = -\frac{2}{T-p}\omega + o(T^{-1}),$$

where $\omega = \sigma_\varepsilon^2/(1 - \sum_{j=1}^{p}\phi_j)^2$.

Proof of Lemma 3

Without loss of generality, we assume $i \le j$. From (8), we have

$$\begin{aligned}
\hat{r}_{ij} &= \frac{1}{T-p}\sum_{t=p+1}^{T}\hat{u}_{t-i}\hat{u}_{t-j} \\
&= \frac{T_b-p+i}{T-p}\left\{\frac{1}{T_b-p+i}\sum_{t=p+1}^{T_b+i}(u_{t-i}-\bar{u}_1)(u_{t-j}-\bar{u}_1)\right\} \\
&\quad +\frac{1}{T-p}\sum_{t=T_b+i+1}^{T_b+j}(u_{t-i}-\bar{u}_2)(u_{t-j}-\bar{u}_1) \\
&\quad +\frac{T-T_b-j}{T-p}\left\{\frac{1}{T-T_b-j}\sum_{t=T_b+j+1}^{T}(u_{t-i}-\bar{u}_2)(u_{t-j}-\bar{u}_2)\right\}.
\end{aligned}$$

Note that the second term in the last equation does not appear when $i = j$. Therefore,

$$\begin{aligned}
E(\hat{r}_{ij}) &= \frac{T_b-p+i}{T-p}\left\{r_{ij}-\frac{1}{\lambda(T-p)}\omega+o(T^{-1})\right\}+\left\{\frac{j-i}{T-p}r_{ij}+o(T^{-1})\right\} \\
&\quad +\frac{T-T_b-j}{T-p}\left\{r_{ij}-\frac{1}{(1-\lambda)(T-p)}\omega+o(T^{-1})\right\} \\
&= r_{ij}-\lambda\cdot\frac{\omega}{\lambda(T-p)}-(1-\lambda)\cdot\frac{\omega}{(1-\lambda)(T-p)}+o(T^{-1}) \\
&= r_{ij}-\frac{2}{T-p}\omega+o(T^{-1}).\blacksquare
\end{aligned}$$

Lemma 4 *Under Assumptions 1 and 2,*

$$Cov(\hat{r}_{ij},\hat{r}_{k\ell}) = Cov(\tilde{r}_{ij},\tilde{r}_{k\ell}) + O(T^{-3/2}),$$

where $\tilde{r}_{ij} = (T-p)^{-1}\sum_{t=p+1}^{T}u_{t-i}u_{t-j}$.

Proof of Lemma 4

Without loss of generality, we assume $i \le j$ and $k \le \ell$. We can see that $\hat{r}_{ij}$ can be expressed as

$$\begin{aligned}
\hat{r}_{ij} &= \frac{1}{T-p}\left\{\sum_{t=p+1}^{T_b+i}(u_{t-i}-\bar{u}_1)(u_{t-j}-\bar{u}_1)+\sum_{t=T_b+i+1}^{T_b+j}(u_{t-i}-\bar{u}_2)(u_{t-j}-\bar{u}_1)\right. \\
&\qquad \left. +\sum_{t=T_b+j+1}^{T}(u_{t-i}-\bar{u}_2)(u_{t-j}-\bar{u}_2)\right\} \\
&= \frac{1}{T-p}\left\{\sum_{t=p+1}^{T_b+i}u_{t-i}u_{t-j}-\bar{u}_1\left(\sum_{t=p+1}^{T_b+i}u_{t-i}\right)-\bar{u}_1\left(\sum_{t=p+1}^{T_b+i}u_{t-j}\right)+(T_b-p+i)\,\bar{u}_1^2\right. \\
&\quad +\sum_{t=T_b+i+1}^{T_b+j}u_{t-i}u_{t-j}-\bar{u}_1\left(\sum_{t=T_b+i+1}^{T_b+j}u_{t-i}\right)-\bar{u}_2\left(\sum_{t=T_b+i+1}^{T_b+j}u_{t-j}\right)+(j-i)\,\bar{u}_1\bar{u}_2 \\
&\quad \left. +\sum_{t=T_b+j+1}^{T}u_{t-i}u_{t-j}-\bar{u}_2\left(\sum_{t=T_b+j+1}^{T}u_{t-i}\right)-\bar{u}_2\left(\sum_{t=T_b+j+1}^{T}u_{t-j}\right)+(T-T_b-j)\,\bar{u}_2^2\right\} \\
&= (\tilde{r}_{ij,1}-c_{ij,1}-c_{ij,2}+c_{ij,3})+(\tilde{r}_{ij,2}-c_{ij,4}-c_{ij,5}+c_{ij,6})+(\tilde{r}_{ij,3}-c_{ij,7}-c_{ij,8}+c_{ij,9}), \quad \text{say}, \\
&= \tilde{r}_{ij}+c_{ij},
\end{aligned}$$

where $c_{ij} = \sum_{n=1}^{9} c_{ij,n}$. Note that $\tilde{r}_{ij,2}$, $c_{ij,4}$, $c_{ij,5}$, and $c_{ij,6}$ do not appear when $i = j$.

Therefore,

$$\begin{aligned}
Cov(\hat{r}_{ij},\hat{r}_{k\ell}) &= Cov(\tilde{r}_{ij},\tilde{r}_{k\ell})+Cov(c_{ij},\tilde{r}_{k\ell})+Cov(\tilde{r}_{ij},c_{k\ell})+Cov(c_{ij},c_{k\ell}) \\
&= Cov(\tilde{r}_{ij},\tilde{r}_{k\ell})+d_1+d_2+d_3, \quad \text{say}.
\end{aligned}$$

First, let us consider d_1, which can be expressed as $d_1 = Cov(c_{ij},\tilde{r}_{k\ell}) = \sum_{n=1}^{9} Cov(c_{ij,n},\tilde{r}_{k\ell})$. For $n = 1$, by Cauchy-Schwarz inequality,

$$\begin{aligned}
|Cov(c_{ij,1},\tilde{r}_{k\ell})| &\le \left(Var\left[\left(\frac{1}{T-p}\sum_{t=p+1}^{T_b+i}u_{t-i}\right)\bar{u}_1\right]Var\left[\frac{1}{T-p}\sum_{t=p+1}^{T}u_{t-k}u_{t-\ell}\right]\right)^{1/2} \\
&= d_{11}^{1/2}d_{12}^{1/2}, \quad \text{say}.
\end{aligned}$$

Since

$$\begin{aligned} d_{11} &= Var\left[\frac{T_b}{T-p}\left(\bar{u}_1 - \frac{1}{T_b}\sum_{t=1}^{p-i} u_t\right)\bar{u}_1\right] \\ &\leq 2\left[Var\left(\frac{T_b}{T-p}\bar{u}_1^2\right) + Var\left(\left(\frac{1}{T-p}\sum_{t=1}^{p-i} u_t\right)\bar{u}_1\right)\right] = O(T^{-2}) \end{aligned}$$

and $d_{12} = O(T^{-1})$, we obtain $Cov(c_{ij,1}, \tilde{r}_{k\ell}) = O(T^{-3/2})$. Similarly, for $n = 2, \cdots, 9$, $Cov(c_{ij,n}, \tilde{r}_{k\ell})$ can be shown to be $O(T^{-3/2})$. Therefore, we have $d_1 = O(T^{-3/2})$. In the same way, $d_2 = O(T^{-3/2})$ can be proved.

Then, consider the term d_3. Since $d_3 = Cov(c_{ij}, c_{k\ell}) = \sum_{n_1=1}^{9}\sum_{n_2=1}^{9} Cov(c_{ij,n_1}, c_{k\ell,n_2})$ and

$$\begin{aligned} |Cov(c_{ij,n_1}, c_{k,\ell,n_2})| &\leq (Var(c_{ij,n_1})Var(c_{k\ell,n_2}))^{1/2} \\ &= \left(O(T^{-2})O(T^{-2})\right)^{1/2} = O(T^{-2}), \end{aligned}$$

d_3 is of order T^{-2}. Therefore, we conclude that $Cov(\hat{r}_{ij}, \hat{r}_{k\ell}) = Cov(\tilde{r}_{ij}, \tilde{r}_{k\ell}) + O(T^{-3/2})$. ■

Proof of Theorem 1

By Lemma 2,

$$\begin{aligned} E(\hat{\phi}) &= \phi + R^{-1}E\left[\hat{r} - r - (\hat{R} - R)\phi\right] - R^{-1}E\left[(\hat{R} - R)R^{-1}\left\{\hat{r} - r - (\hat{R} - R)\phi\right\}\right] + o(T^{-1}) \\ &= \phi + (A) - (B) + o(T^{-1}), \quad \text{say.} \end{aligned}$$

Since $E(\hat{r} - r) = -(T-p)^{-1} \cdot 2\omega\iota + o(T^{-1})$ and $E(\hat{R} - R) = -(T-p)^{-1} \cdot 2\omega\iota\iota' + o(T^{-1})$ by Lemma 3, where ι is a $p \times 1$ vector of ones, we obtain

$$\begin{aligned} (A) &= R^{-1}E\left[\hat{r} - r - (\hat{R} - R)\phi\right] && (24) \\ &= R^{-1}\left[-\frac{2}{T-p}\omega\iota + \frac{2}{T-p}\omega\iota\iota'\phi + o(T^{-1})\right] \\ &= R^{-1}\left[-\frac{2}{T-p} \cdot \frac{\sigma_\varepsilon^2}{\left(1 - \sum_{j=1}^{p}\phi_j\right)^2} \cdot \iota\left(1 - \sum_{j=1}^{p}\phi_j\right) + o(T^{-1})\right] \\ &= -\frac{2}{T-p} \cdot \frac{\sigma_\varepsilon^2}{1 - \sum_{j=1}^{p}\phi_j} R^{-1}\iota + o(T^{-1}). && (25) \end{aligned}$$

Note that the first-order bias of (A) is equal to (-2) times equation (3.6) in Shaman and Stine (1988).[5] Therefore, from (5.4) in Shaman and Stine (1988), the j-th element of (A) is given by $(T-p)^{-1} \cdot 2\sum_{\ell=0}^{j-1}(\phi_\ell - \phi_{p-\ell})$, where $\phi_0 = -1$, so that

$$(A) = -\frac{1}{T-p} F_p \cdot (2B_{3p})\phi^* + o(T^{-1}), \tag{26}$$

where $F_p = [0_{p\times 1}, I_p]$, $\phi^* = [-1, \phi']'$ and B_{3p} is defined in Section 3 and Patterson (2000).

For (B), we can see from Lemma 4 that

$$(B) = R^{-1}E\left[(\tilde{R}-R)R^{-1}\left\{\tilde{r} - r - (\tilde{R}-R)\phi\right\}\right] + o(T^{-1}), \tag{27}$$

where $(\tilde{R})_{ij} = \tilde{r}_{ij}$ and $\tilde{r} = [\tilde{r}_{01}, \cdots, \tilde{r}_{0p}]'$. Hence, the first-order bias (B) is the same as the one in Shaman and Stine (1988), which is given by the sum of (5.1) and (5.3) in Shaman and Stine (1988). Therefore, we have

$$(B) = \frac{1}{T-p} F_p(B_{1p} + B_{2p})\phi^* + o(T^{-1}), \tag{28}$$

where B_{1p} and B_{2p} are defined in Section 3 and Patterson (2000).

From (26) and (28), we obtain

$$\begin{aligned} E(\hat{\phi}) &= \phi + (A) - (B) + o(T^{-1}) \\ &= \phi - \frac{1}{T-p} F_p(B_{1p} + B_{2p} + 2B_{3p})\phi^* + o(T^{-1}) \\ &= \phi - \frac{1}{T-p}(K_p + B_p\phi) + o(T^{-1}). \blacksquare \end{aligned}$$

Appendix B: Proofs of Theorem 2" and Some Related Lemmas

Because the AR(p) model is a special case of the AR(∞) model, we only prove the results for AR(∞) errors. Lemmas 1 and 1', and Theorems 2 and 2' can be proved similarly. Note that p_T becomes a fixed number for the finite order AR model and thus, for example, the order given by $o(p_T/T)$ in the following lemmas becomes $o(1/T)$ in the AR(p) case.

In this appendix, we use the vector norm $\|x\|_\infty = \max_{1\le i\le n}|x_i|$ for an $n\times 1$ vector $x = [x_1, \cdots, x_n]'$, and a matrix norm $\|A\|_\infty = \max_{1\le i\le n}\left(\sum_{j=1}^n |a_{ij}|\right)$ for an $n \times n$ matrix

[5] Note that the notation in our paper is different from that in Shaman and Stine (1988). For example, ϕ_j $(j = 1, \cdots, p)$ corresponds to $-\alpha_j$ $(j = 1, \cdots, p)$ in Shaman and Stine (1988).

$A = (a_{ij})$. This matrix norm is sub-multiplicative, that is, $\|AB\|_\infty \leq \|A\|_\infty \cdot \|B\|_\infty$ holds for $n \times n$ matrices A and B (cf. Hannan and Deistler, 1988, p.266). Moreover, $|x'Ay| \leq n \cdot \|x\|_\infty \cdot \|A\|_\infty \cdot \|y\|_\infty$ holds for $n \times 1$ vectors x, y, and an $n \times n$ matrix A.

Lemma 1" *Under Assumptions 1", 2, and L, the following relations hold:*

$$\begin{aligned} &(a) \quad E\left[\iota'\hat{\phi}\right] = \iota'\phi - \frac{1}{T-p_T}\iota'(K_{p_T} + B_{p_T}\phi) + o\left(\frac{p_T}{T}\right), \\ &(b) \quad E\left[\iota'(\hat{\phi}-\phi)(\hat{\phi}-\phi)'\iota\right] = \frac{1}{T-p_T}\sigma_\varepsilon^2\iota' R^{-1}\iota + o\left(\frac{p_T}{T}\right), \end{aligned}$$

where ϕ is defined by (18)*, ι is a $p_T \times 1$ vector of ones and*

$$\hat{\phi} = \begin{bmatrix} \hat{\phi}_{p_T,1} \\ \hat{\phi}_{p_T,2} \\ \vdots \\ \hat{\phi}_{p_T,p_T} \end{bmatrix} = \begin{bmatrix} \hat{r}_{11} & \hat{r}_{12} & \cdots & \hat{r}_{1,p_T} \\ \hat{r}_{21} & \hat{r}_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \hat{r}_{p_T-1,p_T} \\ \hat{r}_{p_T,1} & \cdots & \hat{r}_{p_T,p_T-1} & \hat{r}_{p_T,p_T} \end{bmatrix}^{-1} \begin{bmatrix} \hat{r}_{10} \\ \hat{r}_{20} \\ \vdots \\ \hat{r}_{p_T,0} \end{bmatrix}.$$

Proof of Lemma 1"

Proof of (a). Using (22) and the relation $\phi = R^{-1}r$, we have,

$$\begin{aligned} \iota'\hat{\phi} &= \iota'\hat{R}^{-1}\hat{r} \\ &= \iota'\phi + \iota'R^{-1}(\hat{r}-r) - \iota'R^{-1}(\hat{R}-R)\phi - \iota'R^{-1}(\hat{R}-R)R^{-1}(\hat{r}-r) \\ &\quad + \iota'R^{-1}(\hat{R}-R)R^{-1}(\hat{R}-R)\phi + \iota'R^{-1}(\hat{R}-R)R^{-1}(\hat{R}-R)R^{-1}(\hat{r}-r) \\ &\quad - \iota'R^{-1}(\hat{R}-R)R^{-1}(\hat{R}-R)R^{-1}(\hat{R}-R)\hat{R}^{-1}\hat{r} \\ &= \iota'\phi + (a) - (b) - (c) + (d) + (e) - (f), \quad \text{say.} \end{aligned}$$

First, let us consider (a). Since Lemma 3 holds uniformly in $0 \leq i \leq p_T$ and $0 \leq j \leq p_T$, we have

$$E(\hat{r}_{ij}) = r_{ij} - \frac{2}{T-p_T}\omega + \xi_{ij}, \tag{29}$$

where $\xi_{ij} = o(T^{-1})$ uniformly in $0 \leq i \leq p_T$ and $0 \leq j \leq p_T$. Therefore,

$$\begin{aligned} E[(a)] &= \iota'R^{-1}E(\hat{r}-r) \\ &= -\iota'R^{-1} \cdot \frac{2}{T-p_T}\omega\iota + \iota'R^{-1}\xi \\ &= (a_1) + (a_2), \quad \text{say,} \end{aligned}$$

where $\xi = [\xi_{10}, \cdots, \xi_{p_T,0}]'$. Since $\|R^{-1}\|_\infty = O(1)$ (cf. den Haan and Levin, 1998), we obtain

$$\begin{aligned} |(a_2)| &\leq p_T \cdot \|\iota\|_\infty \cdot \|R^{-1}\|_\infty \cdot \|\xi\|_\infty \\ &= p_T \cdot O(1) \cdot O(1) \cdot o(T^{-1}) = o(p_T/T). \end{aligned}$$

For (b), we have

$$\begin{aligned} E[(b)] &= \iota' R^{-1} E(\hat{R} - R)\phi \\ &= \iota' R^{-1} \cdot \frac{2}{T - p_T} \omega \iota \iota' \phi + \iota' R^{-1} \Xi \phi \\ &= (b_1) + (b_2), \quad \text{say}, \end{aligned}$$

where Ξ is a $p_T \times p_T$ matrix whose (i,j) element is ξ_{ij}. Since $\Xi\phi = \left[\sum_{k=1}^{p_T} \xi_{1k}\phi_{p_T,k}, \cdots, \sum_{k=1}^{p_T} \xi_{p_T,k}\phi_{p_T,k}\right]'$, we have

$$\begin{aligned} \|\Xi\phi\|_\infty &= \max_{1 \leq j \leq p_T} \left| \sum_{k=1}^{p_T} \xi_{jk}\phi_{p_T,k} \right| \\ &\leq \left(\sum_{k=1}^{p_T} |\phi_{p_T,k}| \right) \cdot \max_{1 \leq j,k \leq p_T} |\xi_{jk}| \\ &= O(1) \cdot o(T^{-1}) = o(T^{-1}). \end{aligned}$$

Therefore,

$$\begin{aligned} |(b_2)| &\leq p_T \cdot \|\iota\|_\infty \cdot \|R^{-1}\|_\infty \cdot \|\Xi\phi\|_\infty \\ &= p_T \cdot O(1) \cdot O(1) \cdot o(T^{-1}) = o(p_T/T). \end{aligned}$$

Combining these results, we have

$$E[(a) - (b)] = (a_1) - (b_1) + o(p_T/T), \tag{30}$$

where $(a_1) - (b_1)$ corresponds to the first order bias of (A) given in (25) in the proof of Theorem 1.

We next consider (c). Since the result of Lemma 4 holds uniformly in $0 \leq i \leq p_T$, $0 \leq j \leq p_T$, $0 \leq k \leq p_T$, and $0 \leq \ell \leq p_T$, we have

$$E\left[(\hat{r}_{ij} - r_{ij})(\hat{r}_{k\ell} - r_{k\ell})\right] = \frac{1}{T - p_T} b_{ij,k\ell} + \xi_{ij,k\ell}, \tag{31}$$

where $b_{ij,k\ell}$ is the first-order bias of $(\hat{r}_{ij} - r_{ij})(\hat{r}_{k\ell} - r_{k\ell})$, and $\xi_{ij,k\ell} = O(T^{-3/2})$ uniformly in $0 \le i \le p_T$, $0 \le j \le p_T$, $0 \le k \le p_T$, and $0 \le \ell \le p_T$. Now, we have

$$(\hat{R} - R)R^{-1}(\hat{r} - r) = \begin{bmatrix} \sum_{\ell=1}^{p_T} \sum_{k=1}^{p_T} r^{k\ell}(\hat{r}_{1k} - r_{1k})(\hat{r}_{\ell 0} - r_{\ell 0}) \\ \vdots \\ \sum_{\ell=1}^{p_T} \sum_{k=1}^{p_T} r^{k\ell}(\hat{r}_{p_T,k} - r_{p_T,k})(\hat{r}_{\ell 0} - r_{\ell 0}) \end{bmatrix},$$

so that

$$\begin{aligned} E[(\hat{R} - R)R^{-1}(\hat{r} - r)] &= \begin{bmatrix} \sum_{\ell=1}^{p_T} \sum_{k=1}^{p_T} r^{k\ell} \cdot \frac{1}{T-p_T} b_{1k,\ell 0} + \sum_{\ell=1}^{p_T} \sum_{k=1}^{p_T} r^{k\ell}\xi_{1k,\ell 0} \\ \vdots \\ \sum_{\ell=1}^{p_T} \sum_{k=1}^{p_T} r^{k\ell} \cdot \frac{1}{T-p_T} b_{p_T k,\ell 0} + \sum_{\ell=1}^{p_T} \sum_{k=1}^{p_T} r^{k\ell}\xi_{p_T k,\ell 0} \end{bmatrix} \\ &= B_1 + \tilde{\xi}, \quad \text{say,} \end{aligned}$$

where r^{ij} is the (i, j) element of R^{-1}. By (31), we obtain

$$\begin{aligned} \|\tilde{\xi}\|_\infty &= \max_{1 \le j \le p_T} \left| \sum_{\ell=1}^{p_T} \sum_{k=1}^{p_T} r^{k\ell}\xi_{jk,\ell 0} \right| \\ &\le \left(\sum_{\ell=1}^{p_T} \sum_{k=1}^{p_T} |r^{k\ell}| \right) \cdot \max_{1 \le j,k,\ell \le p_T} |\xi_{jk,\ell 0}| \\ &= O(p_T) \cdot O(T^{-3/2}) = O(p_T/T^{3/2}), \end{aligned}$$

where $\sum_{\ell=1}^{p_T} \sum_{k=1}^{p_T} |r^{k\ell}| = O(p_T)$ holds because $\|R^{-1}\|_\infty = O(1)$. Therefore,

$$\begin{aligned} E[(c)] &= \iota' R^{-1} E[(\hat{R} - R)R^{-1}(\hat{r} - r)] \\ &= \iota' R^{-1} B_1 + \iota' R^{-1}\tilde{\xi} \\ &= (c_1) + (c_2), \quad \text{say} \end{aligned}$$

Note that

$$\begin{aligned} |(c_2)| &\le p_T \cdot \|\iota\|_\infty \cdot \|R^{-1}\|_\infty \cdot \|\tilde{\xi}\|_\infty \\ &= p_T \cdot O(1) \cdot O(1) \cdot O(p_T/T^{3/2}) = O(p_T^2/T^{3/2}). \end{aligned}$$

For (d), because the (i, j) element of $(\hat{R} - R)R^{-1}(\hat{R} - R)$ is given by $\sum_{\ell=1}^{p_T} \sum_{k=1}^{p_T} r^{k\ell}(\hat{r}_{ik} -$

$r_{ik})(\hat{r}_{\ell j} - r_{\ell j})$, we have

$$
\begin{aligned}
& E[(\hat{R}-R)R^{-1}(\hat{R}-R)] \\
= & \begin{bmatrix} \sum_{\ell=1}^{p_T}\sum_{k=1}^{p_T} r^{k\ell}\left(\frac{1}{T-p_T}b_{1k,\ell 1}+\xi_{1k,\ell 1}\right) & \cdots & \sum_{\ell=1}^{p_T}\sum_{k=1}^{p_T} r^{k\ell}\left(\frac{1}{T-p_T}b_{1k,\ell p_T}+\xi_{1k,\ell p_T}\right) \\ \vdots & \ddots & \vdots \\ \sum_{\ell=1}^{p_T}\sum_{k=1}^{p_T} r^{k\ell}\left(\frac{1}{T-p_T}b_{p_T k,\ell 1}+\xi_{p_T k,\ell 1}\right) & \cdots & \sum_{\ell=1}^{p_T}\sum_{k=1}^{p_T} r^{k\ell}\left(\frac{1}{T-p_T}b_{p_T k,\ell p_T}+\xi_{p_T k,\ell p_T}\right) \end{bmatrix} \\
= & B_2 + \tilde{\Xi}, \quad \text{say,}
\end{aligned}
$$

and each element of $\tilde{\Xi}$ is uniformly $O(p_T/T^{3/2})$. Therefore, we have

$$
\begin{aligned}
E[(d)] &= \iota' R^{-1} E[(\hat{R}-R)R^{-1}(\hat{R}-R)]\phi \\
&= \iota' R^{-1} B_2 \phi + \iota' R^{-1}\tilde{\Xi}\phi \\
&= (d_1) + (d_2), \quad \text{say.}
\end{aligned}
$$

Note that

$$
\begin{aligned}
|(d_2)| &\leq p_T \cdot \|\iota\|_\infty \cdot \|R^{-1}\|_\infty \cdot \|\tilde{\Xi}\|_\infty \cdot \|\phi\|_\infty \\
&= p_T \cdot O(1) \cdot O(1) \cdot O(p_T^2/T^{3/2}) \cdot O(1) = O(p_T^3/T^{3/2}).
\end{aligned}
$$

Combining the above results, we have

$$
E[-(c) + (d)] = -(c_1) + (d_1) + o(p_T/T), \tag{32}
$$

where $-(c_1) + (d_2)$ corresponds to the first order bias of (B) given in (27) in the proof of Theorem 1.

For (e), because $\|\hat{R} - R\|_\infty = O_p(p_T/\sqrt{T})$ and $\|\hat{r} - r\|_\infty = O_p(T^{-1/2})$ (cf. den Haan and Levin, 1998), we have

$$
\begin{aligned}
|(e)| &\leq p_T \cdot \|\iota\|_\infty \left\{ \|R^{-1}\|_\infty \cdot \|\hat{R}-R\|_\infty \right\}^2 \|R^{-1}\|_\infty \cdot \|\hat{r}-r\|_\infty \\
&= p_T \cdot O(1) \cdot \left\{ O(1) \cdot O_p(p_T/\sqrt{T}) \right\}^2 \cdot O(1) \cdot O_p(T^{-1/2}) \\
&= O_p(p_T^3/T^{3/2}).
\end{aligned}
$$

Finally, let us consider (f), which can be expressed as $(f) = \iota' R^{-1}(\hat{R} - R)R^{-1}(\hat{R} - R)R^{-1}(\hat{R}-R)\left\{\phi + (\hat{\phi} - \phi)\right\}$. Since $(\hat{R}-R)\phi = \left[\sum_{\ell=1}^{p_T}(\hat{r}_{1\ell} - r_{1\ell})\phi_{p_T,\ell}, \cdots, \sum_{\ell=1}^{p_T}(\hat{r}_{p_T,\ell} - r_{p_T,\ell})\phi_{p_T,\ell}\right]'$, we have

$$
\begin{aligned}
\|(\hat{R}-R)\phi\|_\infty &= \max_{1\le j\le p_T}\left|\sum_{\ell=1}^{p_T}(\hat{r}_{j\ell}-r_{j\ell})\phi_{p_T,\ell}\right| \\
&\le \max_{1\le j,\ell\le p_T}|\hat{r}_{j\ell}-r_{j\ell}|\cdot\left(\sum_{\ell=1}^{p_T}|\phi_{p_T,\ell}|\right) \\
&= O_p(T^{-1/2})\cdot O(1) \\
&= O_p(T^{-1/2}).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
|(f)| &\le p_T\cdot\|\iota\|_\infty\left\{\|R^{-1}\|_\infty\cdot\|\hat{R}-R\|_\infty\right\}^2\|R^{-1}\|_\infty\left\{\|(\hat{R}-R)\phi\|_\infty+\|\hat{R}-R\|_\infty\cdot\|\hat{\phi}-\phi\|_\infty\right\} \\
&= p_T\cdot O(1)\cdot\left\{O(1)\cdot O_p(p_T/\sqrt{T})\right\}^2\cdot O(1)\cdot\left\{O_p(T^{-1/2})+O_p(p_T/\sqrt{T})\cdot O_p(T^{-1/2})\right\} \\
&= O_p(p_T^3/T^{3/2}).
\end{aligned}
$$

Therefore, we have $E(\iota'\hat{\phi})=\iota'\phi+(a_1)-(b_1)-(c_1)+(d_1)+o(p_T/T)$ because $p_T^4/T\to 0$. Since the first-order bias of $\iota'\hat{\phi}$ given by $(a_1)-(b_1)-(c_1)+(d_1)$ is exactly equal to the one derived in Appendix A, we obtain the desired result. ■

Proof of (b). By defining $\eta_{p_T,t}=\sum_{j=1}^{p_T}(\phi_j-\phi_{p_T,j})u_{t-j}+\sum_{j=p_T+1}^{\infty}\phi_j u_{t-j}$, we can see that

$$
u_t=\sum_{j=1}^{p_T}\phi_{p_T,j}u_{t-j}+\eta_{p_T,t}+\varepsilon_t. \tag{33}
$$

Therefore, we have

$$
\begin{aligned}
\bar{u}_1 &= \frac{1}{T_b}\sum_{t=1}^{T_b}u_t \\
&= \sum_{j=1}^{p_T}\phi_{p_T,j}\left(\frac{1}{T_b}\sum_{t=1}^{T_b}u_{t-j}\right)+\frac{1}{T_b}\sum_{t=1}^{T_b}\eta_{p_T,t}+\frac{1}{T_b}\sum_{t=1}^{T_b}\varepsilon_t \\
&= \sum_{j=1}^{p_T}\phi_{p_T,j}\bar{u}_1+\bar{\eta}_1+\bar{\varepsilon}_1+\sum_{j=1}^{p_T}\phi_{p_T,j}\left\{\frac{1}{T_b}\sum_{\ell=1}^{j}(u_{1-\ell}-u_{T_b+1-\ell})\right\},
\end{aligned} \tag{34}
$$

where $\bar{\eta}_1=T_b^{-1}\sum_{t=1}^{T_b}\eta_{p_T,t}$ and $\bar{\varepsilon}_1=T_b^{-1}\sum_{t=1}^{T_b}\varepsilon_t$. From (33) and (34), we have, for $t=p_T+1,\cdots,T_b$,

$$
\begin{aligned}
\hat{u}_t &= u_t-\bar{u}_1 \\
&= \sum_{j=1}^{p_T}\phi_{p_T,j}\hat{u}_{t-j}+(\eta_{p_T,t}-\bar{\eta}_1)+(\varepsilon_t-\bar{\varepsilon}_1)+h_1,
\end{aligned} \tag{35}
$$

where $h_1 = -\sum_{j=1}^{p_T} \phi_{p_T,j} \left\{ T_b^{-1} \sum_{\ell=1}^{j} (u_{1-\ell} - u_{T_b+1-\ell}) \right\}$.

Similarly, for $t = T_b + 1, \cdots, T_b + p_T$, we have

$$\begin{aligned} \hat{u}_t &= u_t - \bar{u}_1 \cdot 1\{t \le T_b\} - \bar{u}_2 \cdot 1\{t > T_b\} \\ &= \sum_{j=1}^{p_T} \phi_{p_T,j} u_{t-j} + \eta_{p_T,t} + \varepsilon_t - \bar{u}_1 \cdot 1\{t \le T_b\} - \bar{u}_2 \cdot 1\{t > T_b\} \\ &= \sum_{j=1}^{p_T} \phi_{p_T,j} \left[\hat{u}_{t-j} + \bar{u}_1 \cdot 1\{t - j \le T_b\} + \bar{u}_2 \cdot 1\{t - j > T_b\} \right] + \eta_{p_T,t} + \varepsilon_t \\ &\qquad - \bar{u}_1 \cdot 1\{t \le T_b\} - \bar{u}_2 \cdot 1\{t > T_b\} \\ &= \sum_{j=1}^{p_T} \phi_{p_T,j} \hat{u}_{t-j} + \eta_{p_T,t} + \varepsilon_t + \tilde{h}_t, \end{aligned} \tag{36}$$

where $\tilde{h}_t = \sum_{j=1}^{p_T} \phi_{p_T,j} \left[\bar{u}_1 \cdot 1\{t - j \le T_b\} + \bar{u}_2 \cdot 1\{t - j > T_b\} \right] - \bar{u}_1 \cdot 1\{t \le T_b\} - \bar{u}_2 \cdot 1\{t > T_b\}$.

For $t = T_b + p_T + 1, \cdots, T$, we have

$$\hat{u}_t = \sum_{j=1}^{p_T} \phi_{p_T,j} \hat{u}_{t-j} + (\eta_{p_T,t} - \bar{\eta}_2) + (\varepsilon_t - \bar{\varepsilon}_2) + h_2, \tag{37}$$

where $h_2 = -\sum_{j=1}^{p_T} \phi_{p_T,j} \left\{ (T - T_b)^{-1} \sum_{\ell=1}^{j} (u_{T_b+1-\ell} - u_{T+1-\ell}) \right\}$, $\bar{\varepsilon}_2 = (T - T_b)^{-1} \sum_{t=T_b+1}^{T} \varepsilon_t$, and $\bar{\eta}_2 = (T - T_b)^{-1} \sum_{t=T_b+1}^{T} \eta_{p_T,t}$.

From (35)–(37), we obtain

$$\hat{u}_t = \begin{cases} \underline{\hat{u}}_{t-1}' \phi + (\eta_{p_T,t} - \bar{\eta}_1) + (\varepsilon_t - \bar{\varepsilon}_1) + h_1 & \text{for } t = p_T + 1, \cdots, T_b, \\ \underline{\hat{u}}_{t-1}' \phi + \eta_{p_T,t} + \varepsilon_t + \tilde{h}_t & \text{for } t = T_b + 1, \cdots, T_b + p_T, \\ \underline{\hat{u}}_{t-1}' \phi + (\eta_{p_T,t} - \bar{\eta}_2) + (\varepsilon_t - \bar{\varepsilon}_2) + h_2 & \text{for } t = T_b + p_T + 1, \cdots, T, \end{cases} \tag{38}$$

where $\underline{\hat{u}}_t = [\hat{u}_t, \cdots, \hat{u}_{t-p_T+1}]'$.

Since $\hat{\phi} = \left(\sum_{t=p_T+1}^{T} \underline{\hat{u}}_{t-1}\underline{\hat{u}}'_{t-1}\right)^{-1}\left(\sum_{t=p_T+1}^{T} \underline{\hat{u}}_{t-1}\hat{u}_t\right)$, we obtain, using (38),

$$
\begin{aligned}
& \sqrt{T-p_T}\,(\hat{\phi}-\phi) \\
= & \left(\frac{1}{T-p_T}\sum_{t=p_T+1}^{T} \underline{\hat{u}}_{t-1}\underline{\hat{u}}'_{t-1}\right)^{-1} \\
& \times\left[\frac{1}{\sqrt{T-p_T}}\left\{\sum_{t=p_T+1}^{T_b}\underline{\hat{u}}_{t-1}(\eta_{p_T,t}-\bar{\eta}_1)+\sum_{t=T_b+1}^{T_b+p_T}\underline{\hat{u}}_{t-1}\eta_{p_T,t}+\sum_{t=T_b+p_T+1}^{T}\underline{\hat{u}}_{t-1}(\eta_{p_T,t}-\bar{\eta}_2)\right\}\right. \\
& \quad +\frac{1}{\sqrt{T-p_T}}\left\{\sum_{t=p_T+1}^{T_b}\underline{\hat{u}}_{t-1}(\varepsilon_t-\bar{\varepsilon}_1)+\sum_{t=T_b+1}^{T_b+p_T}\underline{\hat{u}}_{t-1}\varepsilon_t+\sum_{t=T_b+p_T+1}^{T}\underline{\hat{u}}_{t-1}(\varepsilon_t-\bar{\varepsilon}_2)\right\} \\
& \quad \left.+\frac{1}{\sqrt{T-p_T}}\left\{\sum_{t=p_T+1}^{T_b}\underline{\hat{u}}_{t-1}h_1+\sum_{t=T_b+1}^{T_b+p_T}\underline{\hat{u}}_{t-1}\tilde{h}_t+\sum_{t=T_b+p_T+1}^{T}\underline{\hat{u}}_{t-1}h_2\right\}\right] \\
= & \hat{R}^{-1}[(A)+(B)+(C)], \quad \text{say.}
\end{aligned}
$$

First, let us consider (A), which can be expressed as

$$
\begin{aligned}
(A) = & \frac{1}{\sqrt{T-p_T}}\left[\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\eta_{p_T,t}-\sum_{t=p_T+1}^{T_b}\underline{u}_{t-1}\bar{\eta}_1-\bar{u}_1\iota\sum_{t=p_T+1}^{T_b}\eta_{p_T,t}+(T_b-p_T)\bar{u}_1\iota\bar{\eta}_1\right. \\
& \left.+\sum_{t=T_b+1}^{T_b+p_T}(\underline{\hat{u}}_{t-1}-\underline{u}_{t-1})\eta_{p_T,t}-\sum_{t=T_b+p_T+1}^{T}\underline{u}_{t-1}\bar{\eta}_2-\bar{u}_2\iota\sum_{t=T_b+p_T+1}^{T}\eta_{p_T,t}+(T-T_b-p_T)\bar{u}_2\iota\bar{\eta}_2\right] \\
= & (A_1)-(A_2)-(A_3)+(A_4)+(A_5)-(A_6)-(A_7)+(A_8), \quad \text{say,}
\end{aligned}
$$

where $\underline{u}_t = [u_t, \cdots, u_{t-p_T+1}]'$.

By den Haan and Levin (1998), Assumption L(b) implies

$$\sum_{j=1}^{p_T}|\phi_j-\phi_{p_T,j}| = o\left(\frac{p_T}{T}\right). \tag{39}$$

Therefore, from Assumption L(b) and (39), we obtain

$$\begin{aligned}
& E\left|\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T} u_{t-\ell}\eta_{p_T,t}\right| \\
= \; & E\left|\sum_{j=1}^{p_T}(\phi_j-\phi_{p_T,j})\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}u_{t-\ell}u_{t-j}\right)+\sum_{j=p_T+1}^{\infty}\phi_j\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}u_{t-\ell}u_{t-j}\right)\right| \\
\leq \; & \left(\sum_{j=1}^{p_T}|\phi_j-\phi_{p_T,j}|+\sum_{j=p_T+1}^{\infty}|\phi_j|\right)\cdot\sup_{j\geq 1}E\left|\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}u_{t-\ell}u_{t-j}\right| \\
= \; & o(p_T/T)\cdot O(\sqrt{T}) \\
= \; & o(p_T/\sqrt{T})
\end{aligned}$$

uniformly in $1\leq\ell\leq p_T$, so that $\|(A_1)\|_\infty = o_p(p_T/\sqrt{T})$.

Similarly, since $E\left|(T-p_T)^{-1}\sum_{t=p_T+1}^{T}\eta_{p_T,t}\right| = o(p_T/T)$, we have $\|(A_2)\|_\infty = o_p(p_T/T)$. In the same way, we obtain $\|(A_3)\|_\infty = o_p(p_T/T)$, $\|(A_4)\|_\infty = o_p(p_T/T)$, $\|(A_5)\|_\infty = O_p(p_T/\sqrt{T})$, $\|(A_6)\|_\infty = o_p(p_T/T)$, $\|(A_7)\|_\infty = o_p(p_T/T)$, and $\|(A_8)\|_\infty = o_p(p_T/T)$. Therefore, $\|(A)\|_\infty = O_p(p_T/\sqrt{T})$.

For (B), since $(T-p_T)^{-1/2}\sum_{t=p_T+1}^{T}u_{t-\ell} = O_p(1)$ uniformly in $1\leq\ell\leq p_T$, $\bar{\varepsilon}_1 = O_p(T^{-1/2})$ and $\bar{\varepsilon}_2 = O_p(T^{-1/2})$, we have

$$\left\|(B)-\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t\right\|_\infty = O_p(T^{-1/2}).$$

Now let us consider (C). Since

$$\begin{aligned}
|h_1| \;\leq\; & \left(\sum_{j=1}^{p_T}|\phi_{p_T,j}|\right)\cdot\frac{1}{T_b}\sum_{\ell=1}^{p_T}(|u_{t-\ell}|+|u_{T_b+1-\ell}|) \\
= \; & O(1)\cdot O_p(p_T/T) \\
= \; & O_p(p_T/T)
\end{aligned}$$

and similarly $|h_2| = O_p(p_T/T)$, we have

$$\begin{aligned}
\|(C)\|_\infty &\leq \left\| \frac{1}{\sqrt{T-p_T}} \sum_{t=p_T+1}^{T_b} \underline{\hat{u}}_{t-1} \right\|_\infty \cdot |h_1| + \left\| \frac{1}{\sqrt{T-p_T}} \sum_{t=T_b+1}^{T_b+p_T} \underline{\hat{u}}_{t-1} \tilde{h}_t \right\|_\infty \\
&\quad + \left\| \frac{1}{\sqrt{T-p_T}} \sum_{t=T_b+p_T+1}^{T} \underline{\hat{u}}_{t-1} \right\|_\infty \cdot |h_2| \\
&= O_p(1) \cdot O_p(p_T/T) + O_p(p_T/\sqrt{T}) + O_p(1) \cdot O_p(p_T/T) \\
&= O_p(p_T/\sqrt{T}).
\end{aligned}$$

Therefore, we obtain

$$\sqrt{T-p_T}\,(\hat{\phi} - \phi) = \hat{R}^{-1} \left(\frac{1}{\sqrt{T-p_T}} \sum_{t=p_T+1}^{T} \underline{u}_{t-1} \varepsilon_t + \zeta_T \right), \tag{40}$$

where $\|\zeta_T\|_\infty = O_p(p_T/\sqrt{T})$.

Then we evaluate the expectation of $\iota'(\hat{\phi} - \phi)(\hat{\phi} - \phi)'\iota$ up to $O(p_T/T)$. Using (40), this can be expressed as

$$\begin{aligned}
&\iota'(\hat{\phi} - \phi)(\hat{\phi} - \phi)'\iota \\
=\; & \frac{1}{T-p_T} \iota' \left\{ R^{-1} + (\hat{R}^{-1} - R^{-1}) \right\} \left(\frac{1}{\sqrt{T-p_T}} \sum_{t=p_T+1}^{T} \underline{u}_{t-1} \varepsilon_t + \zeta_T \right) \\
& \times \left(\frac{1}{\sqrt{T-p_T}} \sum_{t=p_T+1}^{T} \underline{u}'_{t-1} \varepsilon_t + \zeta'_T \right) \left\{ R^{-1} + (\hat{R}^{-1} - R^{-1}) \right\} \iota \\
=\; & \frac{1}{T-p_T} \iota' R^{-1} \left(\frac{1}{\sqrt{T-p_T}} \sum_{t=p_T+1}^{T} \underline{u}_{t-1} \varepsilon_t \right) \left(\frac{1}{\sqrt{T-p_T}} \sum_{t=p_T+1}^{T} \underline{u}'_{t-1} \varepsilon_t \right) R^{-1} \iota \\
& + o_p(p_T/T),
\end{aligned} \tag{41}$$

because $\|R^{-1}\|_\infty = O(1)$, $\|\hat{R}^{-1} - R^{-1}\|_\infty = O_p(p_T/\sqrt{T})$, $\|(T-p_T)^{-1/2} \sum_{t=p_T+1}^{T} \underline{u}_{t-1}\varepsilon_t\|_\infty = O_p(1)$, and $\|\zeta_T\|_\infty = O_p(p_T/\sqrt{T})$.

Since $E(\varepsilon_t | \mathcal{F}_{t-1}) = 0$ and $E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma_\varepsilon^2$, we have

$$\begin{aligned}
E\left[\frac{1}{T-p_T} \left(\sum_{t=p_T+1}^{T} \underline{u}_{t-1} \varepsilon_t \right) \left(\sum_{t=p_T+1}^{T} \underline{u}'_{t-1} \varepsilon_t \right) \right] &= E\left[\frac{1}{T-p_T} \sum_{t=p_T+1}^{T} \underline{u}_{t-1} \underline{u}'_{t-1} \varepsilon_t^2 \right] \\
&= \sigma_\varepsilon^2 R.
\end{aligned} \tag{42}$$

Therefore, from (41) and (42), we obtain

$$E\left[\iota'(\hat{\phi}-\phi)(\hat{\phi}-\phi)'\iota\right] = \frac{1}{T-p_T}\sigma_\varepsilon^2\iota' R^{-1}\iota + o(p_T/T). \blacksquare$$

Lemma 2” *Under Assumptions 1”, 2, and L, the following relations hold:*

(a) $E\left[\left(1-\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j}\right)^2\right] = (1-\iota'\phi)^2 + \frac{1}{T-p_T}\left\{2(1-\iota'\phi)\iota'(K_{p_T}+B_{p_T}\phi) + \sigma_\varepsilon^2\iota' R^{-1}\iota\right\} + o\left(\frac{p_T}{T}\right)$,

(b) $E\left[\hat{\sigma}_\varepsilon^2\right] = \sigma_\varepsilon^2 - \frac{p_T+2}{T-p_T}\sigma_\varepsilon^2 + o\left(\frac{p_T}{T}\right)$,

(c) $Var\left[\hat{\sigma}_\varepsilon^2\right] = \frac{1}{T-p_T}\left\{E(\varepsilon_t^4)-\sigma_\varepsilon^4\right\} + o(T^{-1})$,

(d) $Cov\left[\hat{\sigma}_\varepsilon^2, \left(1-\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j}\right)^2\right] = o\left(\frac{p_T}{T}\right)$.

Proof of Lemma 2”

Proof of (a). Here we define $\psi = -\hat{\phi}+\phi-(T-p_T)^{-1}(K_{p_T}+B_{p_T}\phi)$. Then, from Lemma 1”, we obtain

$$\begin{aligned} E(\iota'\psi) &= o(p_T/T), && (43)\\ Var(\iota'\psi) &= \frac{1}{T-p}\sigma_\varepsilon^2\iota' R^{-1}\iota + o(p_T/T). && (44)\end{aligned}$$

Since $1-\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j} = 1-\iota'\hat{\phi} = 1-\iota'\phi+(T-p_T)^{-1}\iota'(K_{p_T}+B_{p_T}\phi)+\iota'\psi$, we have

$$\begin{aligned} E\left[\left(1-\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j}\right)^2\right] &= E\left[\left\{1-\iota'\phi+\frac{1}{T-p_T}\iota'(K_{p_T}+B_{p_T}\phi)+\iota'\psi\right\}^2\right]\\ &= \left\{1-\iota'\phi+\frac{1}{T-p_T}\iota'(K_{p_T}+B_{p_T}\phi)\right\}^2\\ &\quad +2\left\{1-\iota'\phi+\frac{1}{T-p_T}\iota'(K_{p_T}+B_{p_T}\phi)\right\}E\left[\iota'\psi\right]+E\left[(\iota'\psi)^2\right]\\ &= (a)+2\cdot(b)+(c), \quad \text{say.} && (45)\end{aligned}$$

By (43) and (44), we have

$$\begin{aligned} (a) &= (1-\iota'\phi)^2+\frac{2}{T-p_T}(1-\iota'\phi)\iota'(K_{p_T}+B_{p_T}\phi)+o(p_T/T),\\ (b) &= o(p_T/T),\\ (c) &= \frac{1}{T-p_T}\sigma_\varepsilon^2\iota' R^{-1}\iota+o(p_T/T).\end{aligned}$$

Therefore, the expectation up to $O(p_T/T)$ is given by

$$E\left[\left(1-\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j}\right)^2\right] = (1-\iota'\phi)^2+\frac{2}{T-p_T}(1-\iota'\phi)\iota'(K_{p_T}+B_{p_T}\phi) + \frac{1}{T-p_T}\sigma_\varepsilon^2\iota' R^{-1}\iota + o(p_T/T). \blacksquare$$

Proof of (b). For $t = p_T+1, \cdots, T_b$, $\hat{\varepsilon}_t$ can be expressed as

$$\begin{aligned}
\hat{\varepsilon}_t &= (u_t-\bar{u}_1)-\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j}(u_{t-j}-\bar{u}_1) \\
&= (u_t-\bar{u}_1)-\sum_{j=1}^{p_T}\phi_{p_T,j}(u_{t-j}-\bar{u}_1)-\sum_{j=1}^{p_T}(\hat{\phi}_{p_T,j}-\phi_{p_T,j})(u_{t-j}-\bar{u}_1) \\
&= (\varepsilon_t-\bar{\varepsilon}_1)-(\hat{\phi}-\phi)'\underline{u}_{t-1}+(\eta_{p_T,t}-\bar{\eta}_1)+h_1, \qquad (46)
\end{aligned}$$

where the last equality holds because $\varepsilon_t = u_t - \sum_{j=1}^{p_T}\phi_{p_T,j}u_{t-j}-\eta_{p_T,t}$ and $\bar{\varepsilon}_1 = \bar{u}_1 - \sum_{j=1}^{p_T}\phi_{p_T,j}\bar{u}_1 - \bar{\eta}_1 + h_1$.

For $t = T_b+1, \cdots, T_b+p_T$, we have

$$\begin{aligned}
\hat{\varepsilon}_t &= \hat{u}_t-\sum_{j=1}^{p_T}\phi_{p_T,j}\hat{u}_{t-j}-\sum_{j=1}^{p_T}(\hat{\phi}_{p_T,j}-\phi_{p_T,j})\hat{u}_{t-j} \\
&= \varepsilon_t+\eta_{p_T,t}+O_p(p_T/\sqrt{T}). \qquad (47)
\end{aligned}$$

Similarly, for $t = T_b+p_T+1, \cdots, T$, we obtain

$$\hat{\varepsilon}_t = (\varepsilon_t-\bar{\varepsilon}_2)-(\hat{\phi}-\phi)'\underline{\hat{u}}_{t-1}+(\eta_{p_T,t}-\bar{\eta}_2)+h_2. \qquad (48)$$

Using (46)-(48) and noting that $|h_1| = |h_2| = O_p(p_T/T)$, we have

$$\begin{aligned}
\hat{\sigma}_\varepsilon^2 &= \frac{1}{T-p_T}\left[\left\{\sum_{t=p_T+1}^{T_b}(\varepsilon_t-\bar{\varepsilon}_1)^2+\sum_{t=T_b+1}^{T_b+p_T}\varepsilon_t^2+\sum_{t=T_b+p_T+1}^{T}(\varepsilon_t-\bar{\varepsilon}_2)^2\right\}\right.\\
&\quad -2\left\{\sum_{t=p_T+1}^{T_b}(\hat{\phi}-\phi)'\underline{\hat{u}}_{t-1}(\varepsilon_t-\bar{\varepsilon}_1)+\sum_{t=T_b+p_T+1}^{T}(\hat{\phi}-\phi)'\underline{\hat{u}}_{t-1}(\varepsilon_t-\bar{\varepsilon}_2)\right\}\\
&\quad +\left\{\sum_{t=p_T+1}^{T_b}\left((\hat{\phi}-\phi)'\underline{\hat{u}}_{t-1}\right)^2+\sum_{t=T_b+p_T+1}^{T}\left((\hat{\phi}-\phi)'\underline{\hat{u}}_{t-1}\right)^2\right\}\\
&\quad +\left\{2\sum_{t=p_T+1}^{T_b}(\eta_{p_T,t}-\bar{\eta}_1)\left((\varepsilon_t-\bar{\varepsilon}_1)-(\hat{\phi}-\phi)'\underline{\hat{u}}_{t-1}\right)+2\sum_{t=T_b+1}^{T_b+p_T}\eta_{p_T,t}\varepsilon_t\right.\\
&\quad +2\sum_{t=T_b+p_T+1}^{T}(\eta_{p_T,t}-\bar{\eta}_2)\left((\varepsilon_t-\bar{\varepsilon}_2)-(\hat{\phi}-\phi)'\underline{\hat{u}}_{t-1}\right)\\
&\quad \left.\left.+\sum_{t=p_T+1}^{T_b}(\eta_{p_T,t}-\bar{\eta}_1)^2+\sum_{t=T_b+1}^{T_b+p_T}\eta_{p_T,t}^2+\sum_{t=T_b+p_T+1}^{T}(\eta_{p_T,t}-\bar{\eta}_2)^2\right\}\right]+o_p\left(\frac{p_T}{T}\right)\\
&= (A)-2\cdot(B)+(C)+(D)+o_p\left(\frac{p_T}{T}\right), \quad \text{say.}
\end{aligned}$$

First, consider the term (A). Since

$$\begin{aligned}
(A) &= \frac{T_b-p_T}{T-p_T}\left\{\frac{1}{T_b-p_T}\sum_{t=p_T+1}^{T_b}(\varepsilon_t-\bar{\varepsilon}_1)^2\right\}+\frac{1}{T-p_T}\sum_{t=T_b+1}^{T_b+p_T}\varepsilon_t^2\\
&\quad +\frac{T-T_b-p_T}{T-p_T}\left\{\frac{1}{T-T_b-p_T}\sum_{t=T_b+p_T+1}^{T}(\varepsilon_t-\bar{\varepsilon}_2)^2\right\},
\end{aligned}$$

we have

$$\begin{aligned}
E[(A)] &= \sigma_\varepsilon^2-\left\{\lambda\cdot\frac{1}{\lambda(T-p_T)}\sigma_\varepsilon^2+(1-\lambda)\cdot\frac{1}{(1-\lambda)(T-p_T)}\sigma_\varepsilon^2\right\}+o(T^{-1})\\
&= \sigma_\varepsilon^2-\frac{2}{T-p_T}\sigma_\varepsilon^2+o(T^{-1}).
\end{aligned}$$

Next, let us consider (B). Since

$$\frac{1}{\sqrt{T-p_T}}\left\{\sum_{t=p_T+1}^{T_b}\underline{\hat{u}}_{t-1}(\varepsilon_t-\bar{\varepsilon}_1)+\sum_{t=T_b+p_T+1}^{T}\underline{\hat{u}}_{t-1}(\varepsilon_t-\bar{\varepsilon}_2)\right\}=\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t+\tilde{\zeta}_T,$$

where $\|\tilde{\zeta}_T\|_\infty = O_p(\sqrt{p_T/T})$, we have

$$
\begin{aligned}
(B) &= \frac{1}{T-p_T}\cdot\sqrt{T-p_T}(\hat{\phi}-\phi)'\left[\frac{1}{\sqrt{T-p_T}}\left\{\sum_{t=p_T+1}^{T_b}\underline{\hat{u}}_{t-1}(\varepsilon_t-\bar{\varepsilon}_1)+\sum_{t=T_b+p_T+1}^{T}\underline{\hat{u}}_{t-1}(\varepsilon_t-\bar{\varepsilon}_2)\right\}\right] \\
&= \frac{1}{T-p_T}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}'_{t-1}\varepsilon_t+\zeta'_T\right)\left\{R^{-1}+(\hat{R}^{-1}-R^{-1})\right\} \\
&\quad\times\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t+\tilde{\zeta}_T\right) \\
&= \frac{1}{T-p_T}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}'_{t-1}\varepsilon_t\right)R^{-1}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t\right)+o_p\left(\frac{p_T}{T}\right),
\end{aligned}
$$

because $\|\zeta_T\|_\infty = O_p(p_T/\sqrt{T})$.

From (42), we obtain

$$
\begin{aligned}
& E\left[\frac{1}{T-p_T}\left(\sum_{t=p_T+1}^{T}\underline{u}'_{t-1}\varepsilon_t\right)R^{-1}\left(\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t\right)\right] \\
&= tr\left[R^{-1}E\left(\frac{1}{T-p_T}\left(\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t\right)\left(\sum_{t=p_T+1}^{T}\underline{u}'_{t-1}\varepsilon_t\right)\right)\right] \\
&= tr\left[R^{-1}\cdot\sigma_\varepsilon^2 R\right] \\
&= p_T\sigma_\varepsilon^2, \qquad (49)
\end{aligned}
$$

so that $E[(B)] = (T-p_T)^{-1}\cdot p_T\sigma_\varepsilon^2 + o(p_T/T)$.

Next, let us consider (C):

$$
\begin{aligned}
(C) &= (\hat{\phi}-\phi)'\left[\frac{1}{T-p_T}\left\{\sum_{t=p_T+1}^{T_b}\underline{\hat{u}}_{t-1}\underline{\hat{u}}'_{t-1}+\sum_{t=T_b+p_T+1}^{T}\underline{\hat{u}}_{t-1}\underline{\hat{u}}'_{t-1}\right\}\right](\hat{\phi}-\phi) \\
&= (\hat{\phi}-\phi)'\hat{\hat{R}}(\hat{\phi}-\phi),
\end{aligned}
$$

where $\hat{\hat{R}} = (T-p_T)^{-1}\left\{\sum_{t=p_T+1}^{T_b}\underline{\hat{u}}_{t-1}\underline{\hat{u}}'_{t-1}+\sum_{t=T_b+p_T+1}^{T}\underline{\hat{u}}_{t-1}\underline{\hat{u}}'_{t-1}\right\}$.

Since $\|\hat{\hat{R}} - R\|_\infty = O_p(p_T/\sqrt{T})$, we have

$$\begin{aligned}
(C) &= \frac{1}{T-p_T}\left\{\sqrt{T-p_T}\,(\hat{\phi}-\phi)'\right\}\hat{\hat{R}}\left\{\sqrt{T-p_T}\,(\hat{\phi}-\phi)\right\} \\
&= \frac{1}{T-p_T}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}'\varepsilon_t + \zeta_T'\right)\left\{R^{-1} + (\hat{R}^{-1} - R^{-1})\right\}\left\{R + (\hat{\hat{R}} - R)\right\} \\
&\qquad \times\left\{R^{-1} + (\hat{R}^{-1} - R^{-1})\right\}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t + \zeta_T\right) \\
&= \frac{1}{T-p_T}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}'\varepsilon_t\right)R^{-1}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t\right) + o_p\left(\frac{p_T}{T}\right),
\end{aligned}$$

so that we obtain $E[(C)] = (T-p_T)^{-1}\cdot p_T\sigma_\varepsilon^2 + o(p_T/T)$, using (49).

Finally, let us consider (D), which can be expressed as

$$\begin{aligned}
(D) &= \frac{1}{T-p_T}\left[2\left\{\sum_{t=p_T+1}^{T_b}(\eta_{p_T,t}-\bar{\eta}_1)(\varepsilon_t-\bar{\varepsilon}_1) + \sum_{t=T_b+1}^{T_b+p_T}\eta_{p_T,t}\varepsilon_t + \sum_{t=T_b+p_T+1}^{T}(\eta_{p_T,t}-\bar{\eta}_2)(\varepsilon_t-\bar{\varepsilon}_2)\right\}\right. \\
&\qquad -2(\hat{\phi}-\phi)'\left\{\sum_{t=p_T+1}^{T_b}\hat{\underline{u}}_{t-1}(\eta_{p_T,t}-\bar{\eta}_1) + \sum_{t=T_b+p_T+1}^{T}\hat{\underline{u}}_{t-1}(\eta_{p_T,t}-\bar{\eta}_2)\right\} \\
&\qquad \left. + \left\{\sum_{t=p_T+1}^{T_b}(\eta_{p_T,t}-\bar{\eta}_1)^2 + \sum_{t=T_b+1}^{T_b+p_T}\eta_{p_T,t}^2 + \sum_{t=T_b+p_T+1}^{T}(\eta_{p_T,t}-\bar{\eta}_2)^2\right\}\right] \\
&= 2\cdot(D_1) - 2\cdot(D_2) + (D_3), \quad \text{say.}
\end{aligned}$$

First, consider the term (D_1). By Assumption L(b) and (39), we have

$$\begin{aligned}
E\left|\frac{1}{T-p_T}\sum_{t=p_T+1}^{T}\eta_{p_T,t}\varepsilon_t\right| &\le \left(\sum_{j=1}^{p_T}|\phi_j - \phi_{p_T,j}| + \sum_{j=p_T+1}^{\infty}|\phi_j|\right)\cdot\sup_{j\ge1}E\left|\frac{1}{T-p_T}\sum_{t=p_T+1}^{T}u_{t-j}\varepsilon_t\right| \\
&= o(p_T/T)\cdot O(1) \\
&= o(p_T/T),
\end{aligned}$$

and thus $(D_1) = o_p(p_T/T)$.

Then, let us consider (D_2). Here we define

$$P = \frac{1}{T-p_T}\left\{\sum_{t=p_T+1}^{T_b}\hat{\underline{u}}_{t-1}(\eta_{p_T,t}-\bar{\eta}_1) + \sum_{t=T_b+p_T+1}^{T}\hat{\underline{u}}_{t-1}(\eta_{p_T,t}-\bar{\eta}_2)\right\}.$$

Then, $\|P\|_\infty = o_p(p_T/T)$ because $(T-p_T)^{-1}\sum_{t=p_T+1}^{T} u_{t-\ell}\eta_{p_T,t} = o_p(p_T/T)$ uniformly in $1 \le \ell \le p_T$.

Since $(D_2) = (T-p_T)^{-1/2} \cdot \left\{\sqrt{T-p_T}(\hat{\phi}-\phi)\right\} P$, we have

$$\begin{aligned}
\|(D_2)\|_\infty &\le \frac{1}{\sqrt{T-p_T}} \cdot p_T \cdot \left(\left\| \frac{1}{\sqrt{T-p_T}} \sum_{t=p_T+1}^{T} \underline{u}_{t-1}\varepsilon_t \right\|_\infty + \|\zeta_T\|_\infty \right) \cdot \|\hat{R}^{-1}\|_\infty \cdot \|P\|_\infty \\
&= O(T^{-1/2}) \cdot p_T \cdot \left\{ O_p(1) + O_p(p_T/\sqrt{T}) \right\} \cdot O_p(1) \cdot o_p(p_T/T) \\
&= o_p(p_T^2/T^{3/2}).
\end{aligned}$$

Then, let us consider (D_3). First, we have

$$\begin{aligned}
E\left|\frac{1}{T-p_T}\sum_{t=p_T+1}^{T}\eta_{p_T,t}^2\right| &= E\left|\frac{1}{T-p_T}\sum_{t=p_T+1}^{T}\left\{\sum_{j=1}^{p_T}(\phi_j-\phi_{p_T,j})u_{t-j} + \sum_{j=p_T+1}^{\infty}\phi_j u_{t-j}\right\}^2\right| \\
&\le E\left|\frac{1}{T-p_T}\sum_{t=p_T+1}^{T}\left\{\sum_{j=1}^{p_T}(\phi_j-\phi_{p_T,j})u_{t-j}\right\}^2\right| \\
&\quad +2E\left|\frac{1}{T-p_T}\sum_{t=p_T+1}^{T}\left\{\sum_{j=1}^{p_T}(\phi_j-\phi_{p_T,j})u_{t-j}\right\}\left\{\sum_{j=p_T+1}^{\infty}\phi_j u_{t-j}\right\}\right| \\
&\quad +E\left|\frac{1}{T-p_T}\sum_{t=p_T+1}^{T}\left\{\sum_{j=p_T+1}^{\infty}\phi_j u_{t-j}\right\}^2\right| \\
&= (D_3-1) + 2\cdot(D_3-2) + (D_3-3), \quad \text{say.}
\end{aligned}$$

Since

$$\begin{aligned}
(D_3-1) &\le \left(\sum_{j=1}^{p_T}|\phi_j-\phi_{p_T,j}|\right)^2 \cdot \sup_{s,t} E|u_s u_t| \\
&= o(p_T^2/T^2)\cdot O(1) \\
&= o(p_T^2/T^2),
\end{aligned}$$

and similarly $(D_3-2) = o(p_T^2/T^2)$ and $(D_3-3) = o(p_T^2/T^2)$, we have $E\left|(T-p_T)^{-1}\sum_{t=p_T+1}^{T}\eta_{p_T,t}^2\right| = o(p_T^2/T^2)$, so that $(D_3) = o_p(p_T^2/T^2)$. Thus we have $(D) = o_p(p_T/T)$.

Using the above results, we obtain

$$E(\hat{\sigma}_\varepsilon^2) = \sigma_\varepsilon^2 - \frac{p_T+2}{T-p_T}\sigma_\varepsilon^2 + o(p_T/T). \blacksquare$$

Proof of (c).

Since $\hat{\sigma}_\varepsilon^2 = (T-p_T)^{-1}\sum_{t=p_T+1}^{T}\varepsilon_t^2 + O_p(p_T/T)$, we obtain

$$\begin{aligned}\sqrt{T-p_T}\,\left(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2\right) &= \frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}(\varepsilon_t^2 - \sigma_\varepsilon^2) + O_p(p_T/\sqrt{T}) \\ &\xrightarrow{d} N\left(0, E(\varepsilon_t^4) - \sigma_\varepsilon^4\right),\end{aligned}$$

so that $Var(\hat{\sigma}_\varepsilon^2) = (T-p_T)^{-1}\left\{E(\varepsilon_t^4) - \sigma_\varepsilon^4\right\} + o(T^{-1})$. ■

Proof of (d). We only need to obtain $E\left[(1-\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j})^2\hat{\sigma}_\varepsilon^2\right]$ to prove (d).

From (45), we have

$$\begin{aligned}\left(1-\sum_{j=1}^{p_T}\hat{\phi}_{p_T,j}\right)^2\hat{\sigma}_\varepsilon^2 &= \left\{1-\iota'\phi+\frac{1}{T-p_T}\iota'(K_{p_T}+B_{p_T}\phi)\right\}^2\hat{\sigma}_\varepsilon^2 \\ &\quad +2\left\{1-\iota'\phi+\frac{1}{T-p_T}\iota'(K_{p_T}+B_{p_T}\phi)\right\}(\iota'\psi)\hat{\sigma}_\varepsilon^2 + (\iota'\psi)^2\hat{\sigma}_\varepsilon^2 \\ &= (a)+2\cdot(b)+(c), \quad \text{say.}\end{aligned}$$

For (a), we obtain

$$E[(a)] = (1-\iota'\phi)^2\sigma_\varepsilon^2 - \frac{p_T+2}{T-p_T}(1-\iota'\phi)^2\sigma_\varepsilon^2 + \frac{2}{T-p_T}\sigma_\varepsilon^2(1-\iota'\phi)\iota'(K_{p_T}+B_{p_T}\phi) + o(p_T/T),$$

using the result of Lemma 2" (b).

For (b), we need to calculate $E[(\iota'\psi)\hat{\sigma}_\varepsilon^2]$ up to $O(p_T/T)$. Since

$$\begin{aligned}\sqrt{T-p_T}\,\iota'\psi &= \iota'\hat{R}^{-1}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t + \zeta_T\right) + o_p(1), \\ \sqrt{T-p_T}\,(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2) &= \frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}(\varepsilon_t^2 - \sigma_\varepsilon^2) + o_p(1),\end{aligned}$$

we have

$$
\begin{aligned}
& \iota'\psi(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2) \\
= \ & \frac{1}{T-p_T}\iota'\left\{R^{-1} + (\hat{R}^{-1} - R^{-1})\right\}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t + \zeta_T\right) \\
& \quad \times \left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}(\varepsilon_t^2 - \sigma_\varepsilon^2)\right) + o_p(p_T/T) \\
= \ & \frac{1}{T-p_T}\iota' R^{-1}\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t\right)\left(\frac{1}{\sqrt{T-p_T}}\sum_{t=p_T+1}^{T}(\varepsilon_t^2 - \sigma_\varepsilon^2)\right) + o_p(p_T/T).
\end{aligned}
$$

Here we have

$$
E\left[\frac{1}{T-p_T}\left(\sum_{t=p_T+1}^{T}\underline{u}_{t-1}\varepsilon_t\right)\left(\sum_{t=p_T+1}^{T}(\varepsilon_t^2 - \sigma_\varepsilon^2)\right)\right] = 0
$$

because ε_t is a martingale difference sequence with a finite 4th moment and satisfies $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma_\varepsilon^2$ and $E(\varepsilon_t^3|\mathcal{F}_{t-1}) = \kappa_3$. Therefore, we have $E[\iota'\psi(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2)] = o(p_T/T)$, and thus $E[\iota'\psi\hat{\sigma}_\varepsilon^2] = E[\iota'\psi(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2)] + E[\iota'\psi]\sigma_\varepsilon^2 = o(p_T/T)$, and $E[(b)] = o(p_T/T)$.

For (c), since

$$
\begin{aligned}
(\iota'\psi)^2\hat{\sigma}_\varepsilon^2 &= (\iota'\psi)^2\left\{\sigma_\varepsilon^2 + O_p(T^{-1/2})\right\} \\
&= (\iota'\psi)^2\sigma_\varepsilon^2 + O_p(p_T/T^{3/2})
\end{aligned}
$$

and $E[(\iota'\psi)^2] = (T-p_T)^{-1}\sigma_\varepsilon^2\iota' R^{-1}\iota + o(p_T/T)$, we have $E[(c)] = (T-p_T)^{-1}\sigma_\varepsilon^2\iota' R^{-1}\iota + o(p_T/T)$.

Using the results above and Lemma 2" (a) and (b), we obtain the desired result. ■

Proof of Theorem 2"

Here we slightly modify the relation (13). When $X - E(X) = O_p(p_T/\sqrt{T})$, $Y - E(Y) = O_p(T^{-1/2})$, $E(X) \neq 0$, and $E(Y) \neq 0$, we have

$$
E\left(\frac{X}{Y}\right) = \frac{E(X)}{E(Y)}\left[1 - \frac{Cov(X,Y)}{E(X)E(Y)} + \frac{Var(Y)}{\{E(Y)\}^2}\right] + o(p_T/T),
$$

because $p_T^4/T \to 0$. Therefore, using the results of Lemma 2", we obtain the desired result. ■

References

[1] Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, 817-858.

[2] Andrews, D. W. K. (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica* 61, 821-856.

[3] Andrews, D. W. K., I. Lee and W. Ploberger (1996), "Optimal Changepoint Tests for Normal Linear Regression," *Journal of Econometrics* 70, 9-38.

[4] Andrews, D. W. K. and W. Ploberger (1994), "Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative," *Econometrica* 62, 1383-1414.

[5] Aue, A. and L. Horváth (2013), "Structural Breaks in Time Series," *Journal of Time Series Analysis* 34, 1-16.

[6] Berk, K. N. (1974), "Consistent Autoregressive Spectral Estimates," *Annals of Statistics* 2, 489-502.

[7] Brown, R. L., J. Durbin and J. M. Evans (1975), "Techniques for Testing the Constancy of Regression Relationships over Time," *Journal of the Royal Statistical Society Series B* 37, 149-192.

[8] Crainiceanu, C. M. and T. J. Vogelsang (2007), "Nonmonotonic Power for Tests for a Mean Shift in a Time Series," *Journal of Statistical Computation and Simulation* 77, 457-476.

[9] Csörgő, M. and L. Horváth (1997), *Limit Theorems in Change-Point Analysis.* Wiley, New York.

[10] Deng, A. and P. Perron (2008), "A Non-Local Perspective on the Power Properties of the CUSUM and CUSUM of Squares Tests for Structural Change," *Journal of Econometrics* 142, 212-240.

[11] den Haan, W. J. and A. T. Levin (1998), "Vector Autoregressive Covariance Matrix Estimation," working paper.

[12] Elliott, G. and U. K. Müller (2006), "Efficient Tests for General Persistent Time Variation in Regression Coefficients," *Review of Economic Studies* 73, 907-940.

[13] Hannan, E. J. and M. Deistler (1988), "The Statistical Theory of Linear Systems," John Wiley & Sons, New York.

[14] Juhl, T. and Z. Xiao (2009), "Tests for Changing Mean with Monotonic Power," *Journal of Econometrics* 148, 14-24.

[15] Kejriwal, M. (2009), "Tests for a Mean Shift with Good Size and Monotonic Power," *Economics Letters* 102, 78-82.

[16] Kiefer, N. M. and T. J. Vogelsang (2005), "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," *Econometric Theory* 21, 1130-1164.

[17] Mood, A. M., F. A. Graybill and D. C. Boes (1974), *Introduction to the Theory of Statistics*, Third Edition, McGraw-Hill, Singapore.

[18] Nabeya, S. and Tanaka, K. (1988), "Asymptotic Theory of a Test for the Constancy of Regression Coefficients against the Random Walk Alternative," *Annals of Statistics* 16, 218-235.

[19] Nyblom, J. (1986), "Testing for Deterministic Linear Trend in Time Series," *Journal of the American Statistical Association* 81, 545-549.

[20] Nyblom, J. (1989), "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association* 84, 223-230.

[21] Nyblom, J. and T. Mäkeläinen (1983), "Comparisons of Tests for the Presence of Random Walk Coefficients in a Simple Linear Model," *Journal of the American Statistical Association* 78, 856-864.

[22] Patterson, K. D. (2000), "Bias Reduction in Autoregressive Models," *Economics Letters* 68, 135-141.

[23] Perron, P. (2006), "Dealing with Structural Breaks," in T. C. Mills and K. Patterson, ed., *Palgrave Handbook of Econometrics, Vol.1: Econometric Theory*. Palgrave Macmillan, New York.

[24] Perron, P. and Y. Yamamoto (2014), "On the Usefulness or Lack Thereof of Optimality Criteria for Structural Change Tests," *Econometric Reviews*, forthcoming.

[25] Ploberger, W. and W. Krämer (1992), "The CUSUM Test with OLS Residuals," *Econometrica* 60, 271-285.

[26] Sayginsoy, Ö. and T. J. Vogelsang (2011), "Testing for a Shift in Trend at an Unknown Date: A Fixed-b Analysis of Heteroskedasticity Autocorrelation Robust OLS-Based Tests," *Econometric Theory* 27, 992-1025.

[27] Shaman, P. and R. A. Stine (1988), "The Bias of Autoregressive Coefficient Estimators," *Journal of the American Statistical Association* 83, 842-848.

[28] Shao, X. and X. Zhang (2010), "Testing for Change Points in Time Series," *Journal of American Statistical Association* 105, 1228-1240.

[29] Stine, R. A. and P. Shaman (1989), "A Fixed Point Characterization for Bias of Autoregressive Estimators," *Annals of Statistics* 17, 1275-1284.

[30] Vogelsang, T. J. (1999), "Sources of Nonmonotonic Power when Testing for a Shift in Mean of a Dynamic Time Series," *Journal of Econometrics* 88, 283-299.

[31] Yang, J. and T. J. Vogelsang (2011), "Fixed-b Analysis of LM-Type Tests for a Shift in Mean," *Econometrics Journal* 14, 438-456.

Table 1: Values of K_p and B_p for $p = 1, \cdots, 5$

p	K_p	B_p
1	2	4
2	$\begin{bmatrix} 2 \\ 3 \end{bmatrix}$	$\begin{bmatrix} 1 & 2 \\ 0 & 5 \end{bmatrix}$
3	$\begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 3 \\ -2 & 5 & 2 \\ 0 & 0 & 6 \end{bmatrix}$
4	$\begin{bmatrix} 2 \\ 3 \\ 2 \\ 3 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & 2 \\ -2 & 2 & 2 & 3 \\ -3 & 0 & 6 & 2 \\ 0 & 0 & 0 & 7 \end{bmatrix}$
5	$\begin{bmatrix} 2 \\ 3 \\ 2 \\ 3 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & 0 & 3 \\ -2 & 2 & 0 & 3 & 2 \\ -3 & -2 & 6 & 2 & 3 \\ -2 & 0 & 0 & 7 & 2 \\ 0 & 0 & 0 & 0 & 8 \end{bmatrix}$

Table 2: Empirical size of the tests with AR(1) errors: $u_t = \phi u_{t-1} + \varepsilon_t$

	$T=100$					$T=200$				
	$\phi=0$	$\phi=0.2$	$\phi=0.4$	$\phi=0.6$	$\phi=0.8$	$\phi=0$	$\phi=0.2$	$\phi=0.4$	$\phi=0.6$	$\phi=0.8$
sup-W	0.075	0.107	0.145	0.207	0.335	0.065	0.088	0.111	0.143	0.227
sup-W_{AR}	0.072	0.140	0.128	0.132	0.214	0.064	0.105	0.079	0.085	0.125
sup-W_{BC}	0.061	0.126	0.101	0.078	0.102	0.058	0.096	0.066	0.062	0.069
sup-W_{kej}	0.064	0.077	0.069	0.060	0.045	0.060	0.069	0.064	0.058	0.043
fixed-b sup-W	0.014	0.019	0.026	0.027	0.036	0.032	0.029	0.028	0.031	0.037
CUSUM_{H_1}	0.067	0.092	0.132	0.188	0.312	0.055	0.077	0.093	0.123	0.195
$\text{CUSUM}_{H_1,AR}$	0.064	0.127	0.115	0.123	0.192	0.054	0.087	0.064	0.074	0.112
$\text{CUSUM}_{H_1,BC}$	0.057	0.114	0.086	0.073	0.094	0.048	0.083	0.053	0.053	0.063
SN	0.057	0.065	0.072	0.090	0.141	0.050	0.055	0.060	0.067	0.088

Table 3: Empirical size of the tests with AR(2) errors: $u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \varepsilon_t, \;\; \phi_2 = -0.3$

	$T=100$					$T=200$				
	$\phi_1=0.3$	$\phi_1=0.5$	$\phi_1=0.7$	$\phi_1=0.9$	$\phi_1=1.1$	$\phi_1=0.3$	$\phi_1=0.5$	$\phi_1=0.7$	$\phi_1=0.9$	$\phi_1=1.1$
sup-W	0.032	0.077	0.118	0.200	0.346	0.034	0.071	0.094	0.136	0.224
sup-W_{AR}	0.114	0.118	0.127	0.147	0.206	0.086	0.087	0.086	0.096	0.124
sup-W_{BC}	0.089	0.085	0.080	0.085	0.098	0.067	0.065	0.064	0.066	0.068
sup-W_{kej}	0.011	0.027	0.029	0.025	0.012	0.018	0.036	0.037	0.032	0.015
fixed-b sup-W	0.001	0.003	0.009	0.012	0.017	0.015	0.010	0.008	0.009	0.016
CUSUM_{H_1}	0.028	0.067	0.111	0.171	0.319	0.031	0.059	0.079	0.121	0.195
$\text{CUSUM}_{H_1,AR}$	0.104	0.104	0.109	0.129	0.181	0.073	0.073	0.076	0.081	0.107
$\text{CUSUM}_{H_1,BC}$	0.076	0.075	0.070	0.073	0.095	0.059	0.060	0.059	0.060	0.060
SN	0.041	0.052	0.059	0.070	0.103	0.043	0.045	0.052	0.059	0.072

Table 4: Empirical size of the tests with AR(2) errors: $u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \varepsilon_t,\ \ \phi_2 = 0.3$

	$T = 100$					$T = 200$				
	$\phi_1 = -0.3$	$\phi_1 = -0.1$	$\phi_1 = 0.1$	$\phi_1 = 0.3$	$\phi_1 = 0.5$	$\phi_1 = -0.3$	$\phi_1 = -0.1$	$\phi_1 = 0.1$	$\phi_1 = 0.3$	$\phi_1 = 0.5$
sup-W	0.161	0.221	0.308	0.328	0.403	0.139	0.205	0.277	0.245	0.292
sup-W_{AR}	0.182	0.176	0.267	0.276	0.335	0.106	0.108	0.139	0.124	0.174
sup-W_{BC}	0.155	0.148	0.246	0.212	0.204	0.085	0.085	0.111	0.082	0.097
sup-W_{kej}	0.117	0.177	0.272	0.202	0.163	0.106	0.180	0.251	0.162	0.125
fixed-b sup-W	0.064	0.070	0.078	0.078	0.064	0.058	0.057	0.059	0.065	0.074
CUSUM$_{H_1}$	0.148	0.195	0.282	0.293	0.374	0.119	0.180	0.248	0.221	0.261
CUSUM$_{H_1,AR}$	0.161	0.161	0.248	0.252	0.309	0.088	0.092	0.119	0.113	0.163
CUSUM$_{H_1,BC}$	0.135	0.134	0.227	0.192	0.192	0.071	0.069	0.102	0.075	0.089
SN	0.059	0.073	0.088	0.114	0.172	0.052	0.060	0.064	0.079	0.103

Table 5: Empirical size of the tests with MA(1) errors: $u_t = \varepsilon_t + \theta\varepsilon_{t-1}$

	$T = 100$					$T = 200$				
	$\theta = -0.8$	$\theta = -0.4$	$\theta = 0$	$\theta = 0.4$	$\theta = 0.8$	$\theta = -0.8$	$\theta = -0.4$	$\theta = 0$	$\theta = 0.4$	$\theta = 0.8$
sup-W	0.000	0.025	0.075	0.106	0.129	0.000	0.024	0.065	0.086	0.099
sup-W_{AR}	0.059	0.076	0.072	0.126	0.212	0.035	0.055	0.064	0.091	0.139
sup-W_{BC}	0.043	0.064	0.061	0.093	0.140	0.024	0.046	0.058	0.068	0.098
sup-W_{kej}	0.000	0.014	0.064	0.056	0.045	0.000	0.016	0.060	0.055	0.045
fixed-b sup-W	0.000	0.002	0.014	0.012	0.014	0.001	0.027	0.032	0.017	0.015
CUSUM$_{H_1}$	0.000	0.023	0.067	0.093	0.119	0.000	0.020	0.055	0.075	0.084
CUSUM$_{H_1,AR}$	0.044	0.067	0.064	0.108	0.190	0.021	0.047	0.054	0.079	0.120
CUSUM$_{H_1,BC}$	0.028	0.057	0.057	0.083	0.127	0.016	0.041	0.048	0.060	0.081
SN	0.000	0.026	0.057	0.065	0.067	0.000	0.033	0.050	0.055	0.057

Table 6: Empirical size of the bias-corrected tests with AR(1) errors: $u_t = \phi u_{t-1} + \varepsilon_t$

	$T = 100$					$T = 200$				
	$\phi = 0$	$\phi = 0.2$	$\phi = 0.4$	$\phi = 0.6$	$\phi = 0.8$	$\phi = 0$	$\phi = 0.2$	$\phi = 0.4$	$\phi = 0.6$	$\phi = 0.8$
sup-W_{BC}	0.061	0.126	0.101	0.078	0.102	0.058	0.096	0.066	0.062	0.069
mean-W_{BC}	0.066	0.101	0.074	0.065	0.068	0.050	0.070	0.051	0.053	0.055
exp-W_{BC}	0.068	0.125	0.098	0.085	0.098	0.051	0.081	0.059	0.059	0.071
$LM_{H_1,BC}$	0.071	0.124	0.107	0.100	0.123	0.053	0.082	0.065	0.067	0.087
$qLL_{H_1,BC}$	0.067	0.167	0.099	0.070	0.088	0.054	0.102	0.058	0.054	0.064
$\text{CUSUM}_{H_1,BC}$	0.057	0.114	0.086	0.073	0.094	0.048	0.083	0.053	0.053	0.063

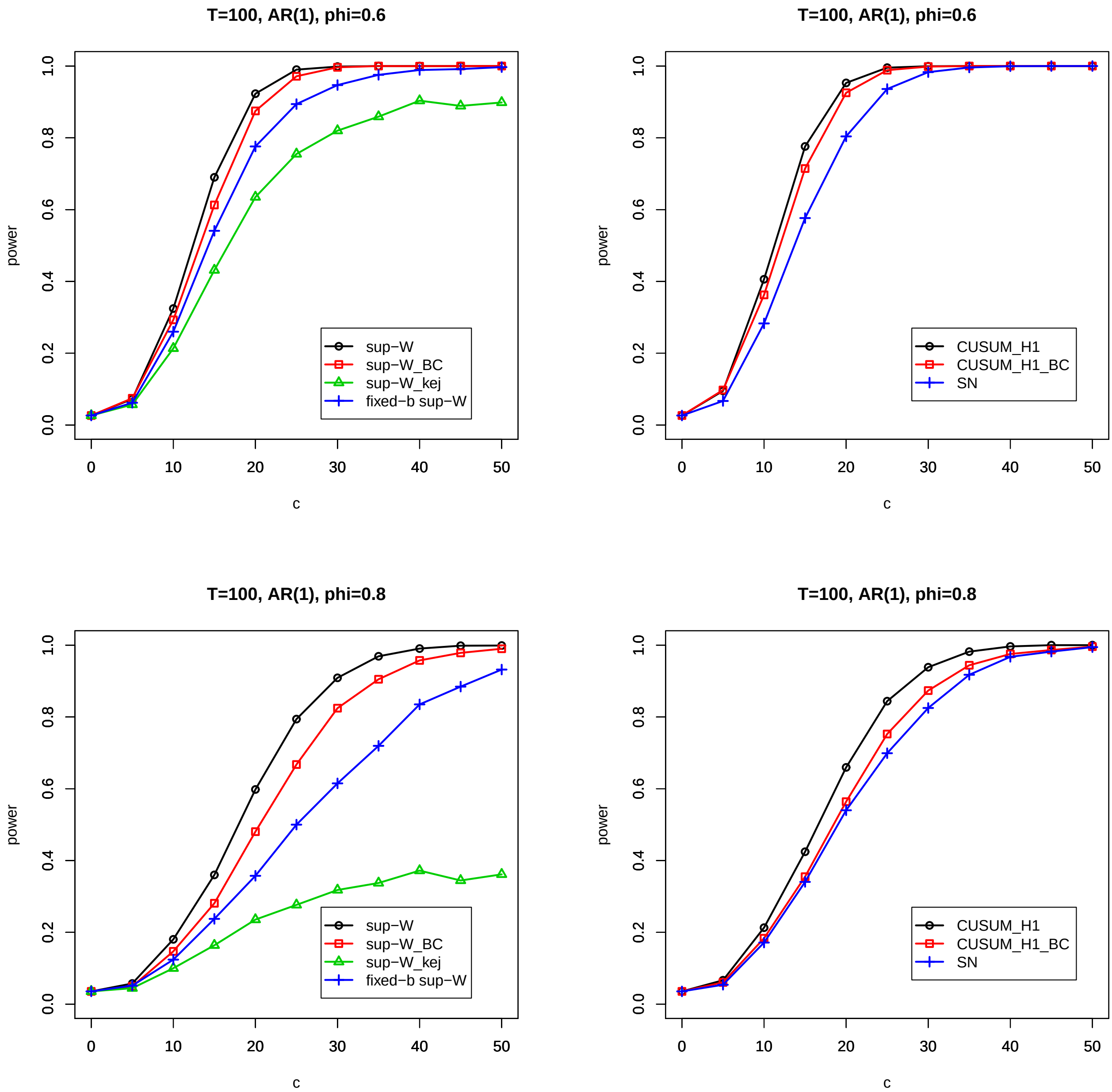

Figure 1: Size-adjusted power of the tests with AR(1) errors and $T = 100$

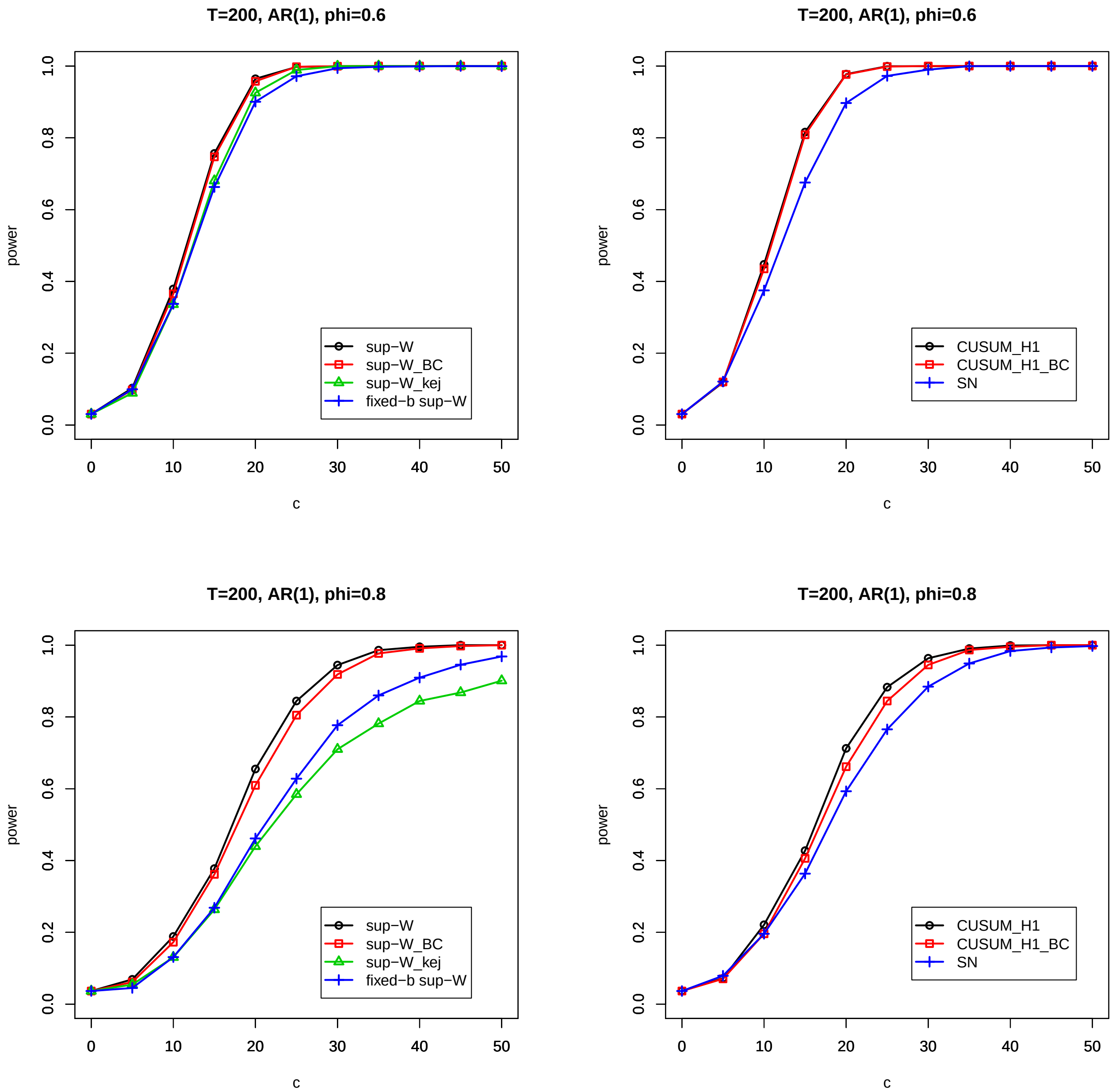

Figure 2: Size-adjusted power of the tests with AR(1) errors and $T = 200$

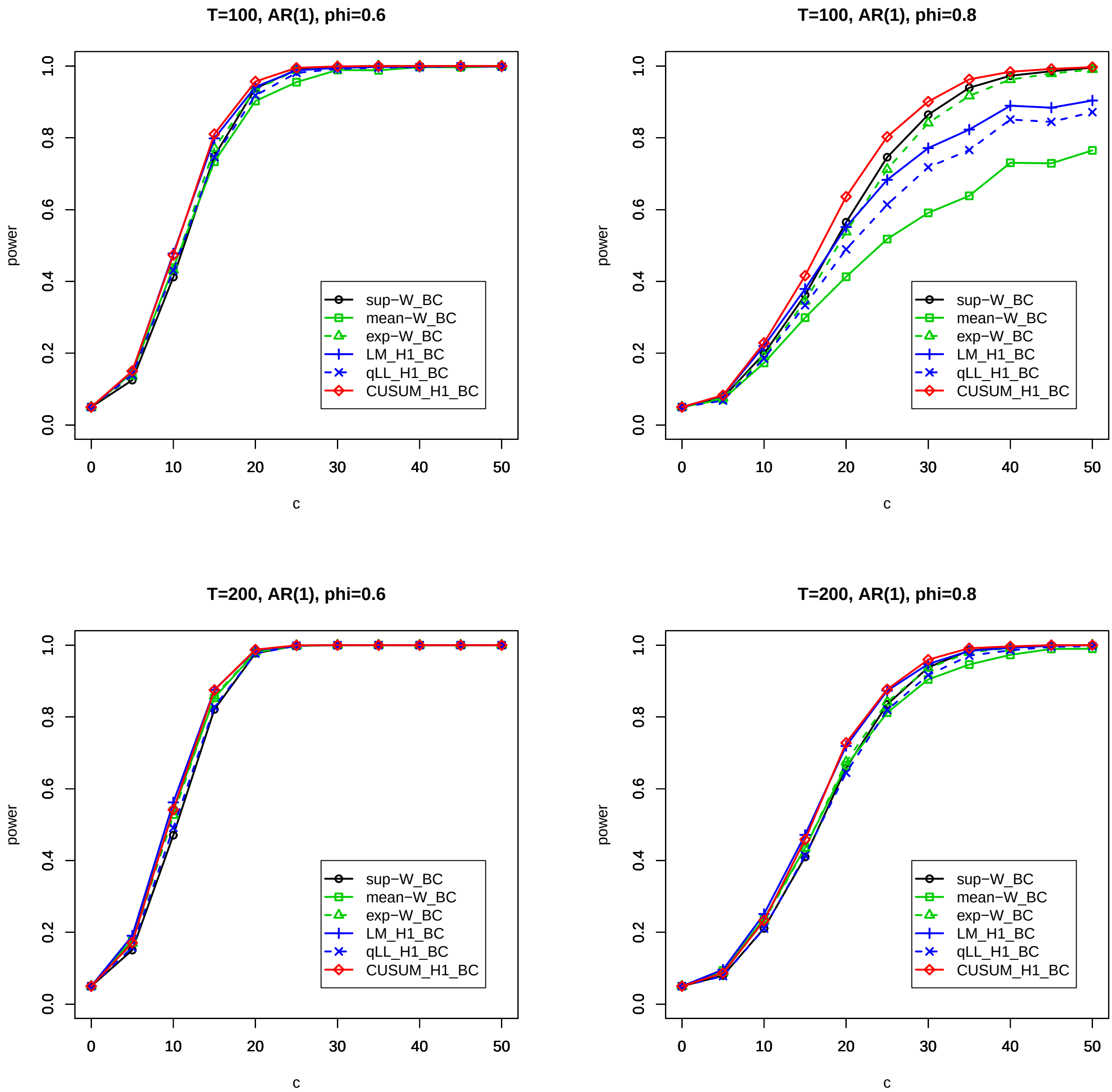

Figure 3: Size-adjusted power of the bias-corrected tests with AR(1) errors