HITOTSUBASHI UNIVERSITY

Global COE Hi-Stat Discussion Paper Series 085

Research Unit for Statistical and Empirical Analysis in Social Sciences (Hi-Stat)

Reducing the Size Distortion of the KPSS Test

Eiji Kurozumi
Shinya Tanaka

September 2009

Hi-Stat
Institute of Economic Research
Hitotsubashi University
2-1 Naka, Kunitatchi Tokyo, 186-8601 Japan
http://gcoe.ier.hit-u.ac.jp

Hi-Stat Discussion Paper

Reducing the Size Distortion of the KPSS Test [1]

Eiji Kurozumi[2] Shinya Tanaka[3]

Department of Economics

Hitotsubashi University

September 2, 2009

Abstract

This paper proposes a new stationarity test based on the KPSS test with less size distortion. We extend the boundary rule proposed by Sul, Phillips and Choi (2005) to the autoregressive spectral density estimator and parametrically estimate the long-run variance. We also derive the finite sample bias of the numerator of the test statistic up to the $1/T$ order and propose a correction to the bias term in the numerator. Finite sample simulations show that the correction term effectively reduces the bias in the numerator and that the finite sample size of our test is close to the nominal one as long as the long-run parameter in the model satisfies the boundary condition.

JEL classification: C12; C22

Key words: Stationary test; size distortion; boundary rule; bias correction

---

[1]We are grateful to Professor Katsumi Shimotsu for his helpful comment. All errors are our responsibility. This research is supported by the Global COE program, the Research Unit for Statistical and Empirical Analysis in Social Sciences at Hitotsubashi University.

[2]Correspondence: Eiji Kurozumi, Department of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo, 186-8601, Japan. E-mail: kurozumi@stat.hit-u.ac.jp

[3]E-mail: stanaka@wd5.so-net.ne.jp

1. Introduction

Following longstanding and well-known studies of the unit root problem, it has been common practice to test for a unit root in time series analysis. In conjunction with testing for a unit root, the null hypothesis of stationarity has often been investigated in practical analysis, and one of the widely applied stationarity tests is the one in Kwiatkowski et al. (1992; hereafter, referred to as KPSS (1992)), also known as the KPSS test.

Although the KPSS test is asymptotically free from nuisance parameter and hence we can asymptotically control the size of the test, it is also known that the test suffers from considerable size distortion in finite samples when the series tested is strongly serially correlated. See, for example, Caner and Killian (2001) and Müller (2005). In order to mitigate the size distortion problem, Rothman (1997) considers the use of size-adjusted critical values, but the KPSS test with size-adjusted critical values loses its power, as pointed out by Rothman (1997) and Caner and Killian (2001). Since one of the reasons for size distortion is the bias in the log-run variance estimator, Sul, Phillips and Choi (2005, SPC hereafter) propose a modified KPSS test by estimating the long-run variance using the prewhitening method proposed by Andrews and Monahan (1992) with the data-dependent boundary rule. Carrion-i-Silvestre and Sansó (2006) investigate the finite sample properties of several stationarity tests and conclude that the SPC test with first order autoregressive (AR(1)) prewhitening is preferable to others in terms of size control. However, their simulations also show that the SPC test with AR(1) prewhitening suffers from size distortion when the data generating process (DGP) is an AR(2) process. Harris, Leybourne and McCabe (2007) focus on the local-to-unity model and propose the GLS-type transformation of data before constructing the test statistic. The size of their test is close to the nominal one when the pre-specified localizing parameter is close to the true one but the test is undersized when the true process is moderately serially correlated, as is shown by Aznar and Ayuda (2008).

Since the size distortion problem is not specific to the KPSS test but a general problem for other stationarity tests, several methods for reducing the size distortion of stationarity tests have been proposed in the literature. For example, Cheung and Chinn (1997) and Kuo and Mikkola (1999) use size-adjusted critical values for Leybourne and McCabe (1994, 1999) tests and Saikkonen and Luukkonen (1993a, b) tests, respectively, but these methods, as in the case of the KPSS test, are unable to correct the loss in the power of the tests. Lanne and Saikkonen (2003) and Kurozumi (2009) propose to modify Leybourne and McCabe (1994, 1999) tests; note that though the sizes of these tests are closer to those of the original tests, they still suffer from size distortion in some cases. Aznar and Ayuda (2008) develop a new test for stationarity using a local-to-unity model but this test is undersized for a process with moderate serial correlation because their test is not designed for the null of stationarity. Unfortunately, all of the above methods seem to have a problem with controlling the size of stationarity tests.

In this paper, we propose a new KPSS-type test for (trend) stationarity with less size distortion. We extend the boundary rule proposed by SPC (2005) to the autoregressive spectral density estimator; the long-run variance is estimated based on the AR approximation. Although it is known that the long-run variance estimator based on the least squares method results in the inconsistency of the test as pointed out by Leybourne and McCabe (1994), we show that this problem can be avoided by applying the boundary rule of SPC (2005). This autoregressive spectral density estimator works relatively well but we still have another problem—the numerator of the KPSS test statistic has a downward bias, and as such, the KPSS test statistic corrected by the new long-run variance estimator becomes undersized. In order to correct the size of the test, we derive the finite sample bias of the numerator of the test statistic and propose the bias-corrected version of the KPSS test. It is shown that the empirical size of our modified test can be well controlled as compared to the other tests.

The paper is organized as follows. Section 2 introduces a model and briefly reviews the KPSS test and the SPC boundary rule. We consider the application of the boundary rule to the long-run variance estimator in Section 3. We also derive the finite sample bias of the numerator of the KPSS test statistic and propose the bias corrected test statistic. Section 4 investigates the finite sample properties of our test. Section 5 gives the concluding remarks.

2. Model and Review of KPSS Test

Let us consider the following model:

$$y_t = d_t'\beta + x_t \quad \text{for} \quad t = 1, 2, \cdots, T, \tag{1}$$

where d_t is deterministic and x_t is a stochastic component. As in the literature, we consider two cases: $d_t = 1$ (constant case) and $d_t = [1, t]'$ (trend case). The integrated order of y_t is determined by the behavior of x_t. We consider the following assumption in this paper.

Assumption 1. *(a) Under the null hypothesis, x_t is covariance stationary with 1-summable autocovariances; the spectral density function of x_t, given by $f(\lambda)$, is bounded and does not equal zero for $-\infty < \lambda < \infty$; the functional central limit theorem (FCLT) can be applied to the partial sum process of x_t.*
(b) Under the alternative hypothesis, Δx_t satisfies condition (a) where $\Delta = 1 - L$ with L being the lag operator.

According to Assumption 1, y_t is covariance stationary (trend stationary) under the null hypothesis while it is a unit root process under the alternative.

KPSS (1992) propose to test for the null of (trend) stationarity against the alternative of a unit root. The KPSS test statistic is defined as

$$KPSS = \frac{\frac{1}{T^2}\sum_{t=1}^{T}\left(\sum_{s=1}^{t}\hat{x}_s\right)^2}{\hat{\omega}}, \tag{2}$$

where $\hat{x}_t$ is the regression residual of y_t on d_t and $\hat{\omega}$ is a consistent estimator of the long-run variance ω defined by $\omega = \lim_{T\to\infty} \mathrm{Var}(T^{-1/2}\sum_{t=1}^{T} x_t)$. KPSS (1992) originally proposed to estimate ω by the nonparametric method using the Bartlett kernel, but other kernels such as the quadratic spectral kennel are also applicable.

By applying the FCLT to the partial sum process of $\hat{x}_t$, it is shown that $T^{-1/2}\sum_{t=1}^{[Tr]}\hat{x}_t$ weakly converges to $\sqrt{\omega}V_1(r)$ for the constant case and to $\sqrt{\omega}V_2(r)$ for the trend case where $[a]$ denotes the largest integer less than a, $V_1(r) = B(r) - rB(1)$ and $V_2(r) = B(r) + \left(2r - 3r^2\right)B(1) + \left(-6r + 6r^2\right)\int_0^1 B(s)\,ds$ with $B(r)$ being a standard Brownian motion.

Using these results, it is shown that $KPSS \xrightarrow{d} \int_0^1 \{V_i(r)\}^2 \, dr$ under the null hypothesis with $i = 1$ for the constant case and $i = 2$ for the trend case where $\xrightarrow{d}$ signifies convergence in distribution.

As is seen from the limiting null distribution, the KPSS test is free from nuisance parameter and we can asymptotically control the size of the test. However, as explained in the introduction section, it suffers from considerable size distortion in finite samples. In order to mitigate size distortion, SPC (2005) propose to estimate the long-run variance using the prewhitening method with a boundary rule. According to SPC (2005), we first estimate an AR(p) model for $\hat{x}_t$ as $\hat{x}_t = \hat{\rho}_1 \hat{x}_{t-1} + \cdots + \hat{\rho}_p \hat{x}_{t-p} + \hat{e}_t$ and then define the new long-run variance estimator as

$$\tilde{\omega} = \frac{\hat{\omega}_e}{(1-\tilde{\rho})^2} \quad \text{where} \quad \tilde{\rho} = \min\left(\hat{\rho}_1 + \cdots + \hat{\rho}_p, \ 1 - \frac{1}{\sqrt{T}}\right)$$

and $\hat{\omega}_e$ is the long-run variance estimator based on $\hat{e}_t$. SPC (2005) show that the KPSS test statistic (2) corrected by $\tilde{\omega}$ has the same limiting distribution under the null hypothesis while it diverges to infinity at rate T under the alternative.

In practice, it is often the case that the prewhitening method is implemented with an AR(1) approximation. Moreover, Carrion-i-Silvestre and Sansó (2006) show that the size of the SPC test with AR(1) prewhitening is close to the nominal one when the true DGP is an AR(1) process while it tends to be greater than the nominal size when the true DGP is an AR(2) process.

3. Bias Corrected KPSS Test

3.1. Estimation of the long-run variance

The boundary rule exploited by SPC (2005) is a clever tool to estimate the long-run variance with less bias. We apply this rule to the autoregressive spectral density estimator. As shown by Perron and Ng (1996), unit root tests corrected by the autoregressive spectral density estimator perform well and we expect that this would be the case for stationarity tests.

In order to see the AR expression of x_t, we first express x_t as an infinite order moving average (MA(∞)) process under the null hypothesis. Under Assumption 1(a), the original process x_t can be expressed, using the Wold representation and Theorem 3.8.4 of Brillinger (1981), as

$$x_t = \psi(L)\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \text{where} \quad \sum_{j=0}^{\infty} j|\psi_j| < \infty, \tag{3}$$

$\psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$ is a lag polynomial with $\psi_0 = 1$ and ε_t is a sequence of white noise with $E[\varepsilon_t^2] = \sigma_\varepsilon^2$. Further, from Theorem 3.8.2 of Brillinger (1981), the lag polynomial $\psi(L)$ is invertible and hence we have

$$\phi(L)x_t = \varepsilon_t, \quad \text{where} \quad \phi(L) = 1 - \sum_{j=1}^{\infty} \phi_j L^j \quad \text{and} \quad \sum_{j=1}^{\infty} j|\phi_j| < \infty. \tag{4}$$

Similarly, we can also see that x_t is expressed as $\phi_*(L)(1-L)x_t = \varepsilon_t$ under the alternative because Δx_t satisfies Assumption 1(a) under the alternative. By defining $\phi(L) = \phi_*(L)(1-L)$ under the alternative, we can see that the original process x_t has an AR(∞) representation given by $\phi(L)x_t = \varepsilon_t$ under both the null and the alternative hypotheses and thus the testing problem is given by

$$H_0: \ \phi(1) > 0 \quad \text{vs.} \quad H_1: \ \phi(1) = 0.$$

From (4), the long-run variance of x_t is given by $\sigma_\varepsilon^2/\phi^2(1)$ and the natural estimator is obtained by approximating the AR(∞) representation by the AR(p) model where p diverges to infinity at an appropriate rate as $T \to \infty$. However, as pointed out by Leybourne and McCabe (1994), the long-run variance estimator based on the least squares estimation of the AR(p) model results in the inconsistency of the test.

In order to avoid this problem, we make use of the boundary rule by SPC (2005). We first fit the AR(p) model to $\hat{x}_t$,

$$\hat{x}_t = \hat{\phi}_1 \hat{x}_{t-1} + \cdots + \hat{\phi}_p \hat{x}_{t-p} + \hat{\varepsilon}_t$$

and then estimate the long-run variance based on the autoregressive spectral density estimator as follows:

$$\tilde{\omega}_{AR} = \frac{\hat{\sigma}_\varepsilon^2}{(1-\tilde{\phi})^2} \quad \text{where} \quad \hat{\sigma}_\varepsilon^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2 \quad \text{and} \quad \tilde{\phi} = \min\left(\sum_{j=1}^{p}\hat{\phi}_j,\ 1-\frac{c}{\sqrt{T}}\right)$$

with c being some constant. We propose to test for the null of (trend) stationarity using the KPSS test statistic (2) with $\hat{\omega}$ replaced by $\tilde{\omega}_{AR}$. We call this test the modified KPSS test.

Note that under the null hypothesis, $\tilde{\phi}$ equals $\sum_{j=1}^{p}\hat{\phi}_j$ for large T; hence $\tilde{\omega}_{AR}$ converges in probability to ω. For details, see Berk (1974). We can see that the modified KPSS test statistic has the same limiting distribution as the original one. On the other hand, as in SPC (2005), $\hat{\sigma}_\varepsilon^2$ still converges in probability to σ_ε^2 under the alternative while $(1-\tilde{\phi})^2$ is shown to be of order $1/T$. As a result, $\tilde{\omega}_{AR}$ diverges to infinity at rate T. Since the numerator of the KPSS test statistic diverges to infinity at rate T^2, we can see that the KPSS test statistic corrected by $\tilde{\omega}_{AR}$ diverges to infinity at rate T under the alternative, and as such, the modified test is consistent.

3.2. Bias correction of the test statistic

As expected from the case of unit root tests by Perron and Ng (1996), the autoregressive spectral density estimator performs quite well in our preliminary simulation. However, once the long-run variance is well estimated, we encounter another problem—the numerator of the test statistic (2) is biased downward in finite samples. As we will see in the next section, the modified KPSS test tends to under-reject the null hypothesis because of this downward bias, and hence, it loses power considerably under the alternative.

In order to control the size of the test, we derive the bias in the numerator of the test statistic under the null hypothesis and consider the bias-corrected version of the modified KPSS test statistic. To calculate the bias, we first express x_t in (3) by the Beveridge-Nelson decomposition as

$$x_t = \psi(1)\varepsilon_t + v_{t-1} - v_t, \quad \text{where} \quad v_t = \sum_{j=0}^{\infty}\tilde{\psi}_j\varepsilon_{t-j} \quad \text{with} \quad \tilde{\psi}_j = \sum_{i=j+1}^{\infty}\psi_i. \tag{5}$$

Since $\hat{x}_t$ is obtained by regressing y_t on d_t, we can see that

$$\begin{aligned}
\hat{x}_t &= x_t - d_t'\left(\sum_{t=1}^{T} d_t d_t'\right)^{-1}\sum_{t=1}^{T} d_t x_t \\
&= \psi(1)\varepsilon_t + v_{t-1} - v_t - d_t'\left(\sum_{t=1}^{T} d_t d_t'\right)^{-1}\sum_{t=1}^{T} d_t(\psi(1)\varepsilon_t + v_{t-1} - v_t) \\
&= \psi(1)\hat{\varepsilon}_t - \widehat{\Delta v_t}
\end{aligned}$$

where $\hat{\varepsilon}_t$ and $\widehat{\Delta v_t}$ are the regression residuals of ε_t and Δv_t on d_t, respectively (note that $\widehat{\Delta v_t}$ is different from $\Delta \hat{v}_t$). Using this expression, the numerator of the KPSS test statistic (2) is decomposed into three terms:

$$\begin{aligned}
\frac{1}{T^2}\sum_{t=1}^{T}\left(\sum_{s=1}^{t}\hat{x}_s\right)^2 &= \frac{\psi^2(1)}{T^2}\sum_{t=1}^{T}\left(\sum_{s=1}^{t}\hat{\varepsilon}_s\right)^2 \\
&\quad + \frac{1}{T^2}\sum_{t=1}^{T}\left(\sum_{s=1}^{t}\widehat{\Delta v_s}\right)^2 - \frac{2\psi(1)}{T^2}\sum_{t=1}^{T}\left(\sum_{s=1}^{t}\hat{\varepsilon}_s\right)\left(\sum_{s=1}^{t}\widehat{\Delta v_s}\right). \\
&= \frac{\psi^2(1)}{T^2}\sum_{t=1}^{T}\left(\sum_{s=1}^{t}\hat{\varepsilon}_s\right)^2 + R_1 - R_2, \quad \text{say.} \qquad (6)
\end{aligned}$$

It can be shown that the first term on the right hand side of (6) is the leading term while the second and third terms are $o_p(1)$. From the simulation result in KPSS (1992), the finite sample distribution of the first term on the right hand side of (6), except for a scalar term $\psi^2(1)$, is well approximated by the limiting distribution and thus we expect that the downward bias in the numerator comes from R_1 and R_2. Therefore, we define the finite sample bias in the numerator as the expectation of $R_1 - R_2$ up to the $O(T^{-1})$ terms. This is denoted by b_T:

$$E[R_1 - R_2] = b_T + o\left(\frac{1}{T}\right).$$

The following theorem gives the expression of the bias term b_T.

Theorem 1. *Let $\gamma_0 = E[v_t^2]$ and $\phi(L)$ be the lag polynomial given in (4). Under Assumption 1(a), the bias term b_T in the numerator of the KPSS test statistic is expressed as*

$$b_T = \frac{b_0}{T}\left(\gamma_0 + \sigma_\varepsilon^2\frac{\phi'(1)}{\phi^3(1)}\right)$$

where $\phi'(1) = d\phi(z)/dz|_{z=1}$ *and* $b_0 = 5/3$ *for the constant case and* $b_0 = 19/15$ *for the trend case.*

The direction of the bias is not necessarily obvious because $\gamma_0 > 0$ while $\phi'(1)/\phi^3(1)$ is negative when x_t is positively serially correlated. However, as is shown in the following corollary, the bias turns out to be negative when x_t is an AR(1) process with positive serial correlation.

Corollary 1. *Assume that* x_t *is an AR(1) process given by* $x_t = \phi_1 x_{t-1} + \varepsilon_t$*. Then, when* $|\phi_1| < 1$*, the bias term* b_T *in the numerator of the KPSS test statistic is expressed as*

$$b_T = -\frac{b_0}{T}\frac{\sigma_\varepsilon^2 \phi_1}{(1-\phi_1)^2(1-\phi_1^2)}$$

where b_0 *is the same as in Theorem 1.*

This corollary is easily obtained by noting that

$$\gamma_0 = \frac{\sigma_\varepsilon^2 \phi_1^2}{(1-\phi_1)^2(1-\phi_1^2)}$$

for the AR(1) case and we omit the poof.

From corollary 1 we can see that the bias for an AR(1) process is always negative when $\phi_1 > 0$ and that the bias takes large negative values as ϕ_1 approaches 1. This explains why the modified KPSS test tends to be undersized when the process is strongly serially correlated. As we will see in the next section, the downward bias in the test statistic is serious when the process is strongly serially correlated and hence the power of the test can be below the significance level in some cases. We thus need to correct the bias in the numerator of the test statistic.

In practice, we need to estimate the bias based on the AR(p) approximation. Although we can easily estimate b_T for the AR(1) case because it is explicitly expressed as a function of the AR coefficient as given in Corollary 1, we have to estimate γ_0 in general. Since γ_0 cannot be expressed in the closed form using the AR coefficients for a general AR(p) model, we need to estimate it recursively like solving the Yule-Walker equations.

We first note that the bias b_T includes the reciprocal of $\phi(1)$ and the least squares estimator of this term might take large values because of the estimation error in the AR coefficients. As a result, the estimator of the bias might take explosively large negative values through the $\phi'(1)/\phi^3(1)$ term. In order to avoid the explosive behavior of the bias term, we estimate it based on the least squares method with the same inequality constraint as the boundary rule. That is, we estimate the AR(p) model by minimizing the sum of squared residuals with the inequality constraint given by $\sum_{j=1}^{p} \phi_j \leq 1 - c/\sqrt{T}$. We can see that the constrained estimator is consistent as long as the boundary rule is satisfied while it is not explosive even under the alternative because of the constraint.

In order to explain how to estimate γ_0, let us assume that $\phi(L)$ be the lag polynomial of order p, so that $x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t$. By inserting the MA($\infty$) expression (3) into both sides, we have

$$\sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} = \phi_1 \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i-1} + \cdots + \phi_p \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i-p} + \varepsilon_t.$$

By comparing the coefficients associated with ε_{t-i} for $i = 0, 1, 2, \cdots$, we can observe the following relations between $\{\psi_i\}$ and $\{\phi_i\}$: $\psi_0 = 1$, $\psi_1 = \phi_1 \psi_0$, $\psi_2 = \phi_1 \psi_1 + \phi_2 \psi_0$, $\cdots$. That is,

$$\psi_i = \sum_{k=1}^{i} \phi_k \psi_{i-k} \quad \text{for} \quad i = 1, \cdots, p-1 \tag{7}$$

$$\psi_i = \sum_{k=1}^{p} \phi_k \psi_{i-k} \quad \text{for} \quad i \geq p. \tag{8}$$

Using relation (7) and the constrained estimators of $\phi_1, \cdots, \phi_p$, we can get the estimators $\hat{\psi}_1, \cdots, \hat{\psi}_{p-1}$. In addition, we also obtain the estimators of $\tilde{\psi}_0, \cdots, \tilde{\psi}_{p-1}$ using the following relation:

$$\tilde{\psi}_i = \sum_{k=i+1}^{\infty} \psi_k = \sum_{k=0}^{\infty} \psi_k - \psi_0 - \cdots - \psi_i = \frac{1}{\phi(1)} - \psi_0 - \cdots - \psi_i,$$

for $i = 0, \cdots, p-1$, since $\psi(1) = 1/\phi(1)$.

We next make use of the relation among the autocovariances of v_t. By summing (8) over

$i \geq j+1$ with $j \geq p$, we have

$$\sum_{i=j+1}^{\infty} \psi_i = \phi_1 \sum_{i=j}^{\infty} \psi_i + \phi_2 \sum_{i=j-1}^{\infty} \psi_i + \cdots + \phi_p \sum_{i=j-p+1}^{\infty} \psi_i,$$

or equivalently,

$$\tilde{\psi}_j = \phi_1 \tilde{\psi}_{j-1} + \phi_2 \tilde{\psi}_{j-2} + \cdots + \phi_p \tilde{\psi}_{j-p} \quad \text{for} \quad j \geq p, \tag{9}$$

since $\tilde{\psi}_j = \sum_{i=j+1}^{\infty} \psi_i$. Multiplying both sides of (9) with $\tilde{\psi}_{j-k}$ and summing over $j \geq p$, we get

$$\sum_{j=p}^{\infty} \tilde{\psi}_j \tilde{\psi}_{j-k} = \phi_1 \sum_{j=p}^{\infty} \tilde{\psi}_{j-1} \tilde{\psi}_{j-k} + \phi_2 \sum_{j=p}^{\infty} \tilde{\psi}_{j-2} \tilde{\psi}_{j-k} + \cdots + \phi_p \sum_{j=p}^{\infty} \tilde{\psi}_{j-p} \tilde{\psi}_{j-k}. \tag{10}$$

Noting that $\gamma_k = E[v_t v_{t-k}] = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \tilde{\psi}_{j+k} \tilde{\psi}_j$ for $k = 0, 1, 2, \cdots$ because, as given in (5), $v_t = \sum_{j=0}^{\infty} \tilde{\psi}_j \varepsilon_{t-j}$, we can see that (10) can be expressed as

$$\gamma_k = \phi_1 \gamma_{k-1} + \cdots + \phi_{k-1} \gamma_1 + \phi_k \gamma_0 + \phi_{k+1} \gamma_1 + \cdots + \phi_p \gamma_{p-k} + a_k \tag{11}$$

for $k = 0, 1, \cdots, p$ where

$$\begin{aligned} a_k \;=\; & \sigma_\varepsilon^2 \left(\sum_{j=0}^{p-k-1} \tilde{\psi}_{j+k} \tilde{\psi}_j - \phi_1 \sum_{j=0}^{p-k-1} \tilde{\psi}_{j+k-1} \tilde{\psi}_j - \cdots - \phi_{k-1} \sum_{j=0}^{p-k-1} \tilde{\psi}_{j+1} \tilde{\psi}_j - \phi_k \sum_{j=0}^{p-k-1} \tilde{\psi}_j^2 \right. \\ & \left. -\phi_{k+1} \sum_{j=0}^{p-k-2} \tilde{\psi}_j \tilde{\psi}_{j+1} - \cdots - \phi_{p-1} \tilde{\psi}_0 \tilde{\psi}_{p-k-1} \right). \end{aligned}$$

Since we have already obtained the estimators of $\phi_1, \cdots, \phi_p$ and $\tilde{\psi}_0, \cdots, \tilde{\psi}_{p-1}$, we can also calculate a_k for $k = 0, \cdots, p$. Since (11) for $k = 0, \cdots, p$ can be seen as $p+1$ simultaneous equations with respect to $\gamma_0, \cdots, \gamma_p$, we can get the estimator of γ_0 by solving a set of these equations.

Once we get the estimator of γ_0, we can construct $\hat{b}_T$, the estimator of b_T. Finally, we construct the bias corrected version of the KPSS test statistic

$$KPSS_{BC} = \frac{\frac{1}{T^2} \sum_{t=1}^{T} \left(\sum_{s=1}^{t} \hat{x}_s \right)^2 - \hat{b}_T}{\tilde{\omega}_{AR}}.$$

Note that the bias corrected KPSS test statistic has the same limiting distribution as the original KPSS test statistic under the null hypothesis, and as such, we can use the critical values in the table given by KPSS (1992).

4. Simulation Study

In this section, we investigate the finite sample properties of the bias corrected version of the modified KPSS test statistic through Monte Carlo simulations. The DGP we considered is given as follows:

$$y_t = d_t'\beta + x_t, \quad x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \varepsilon_t$$

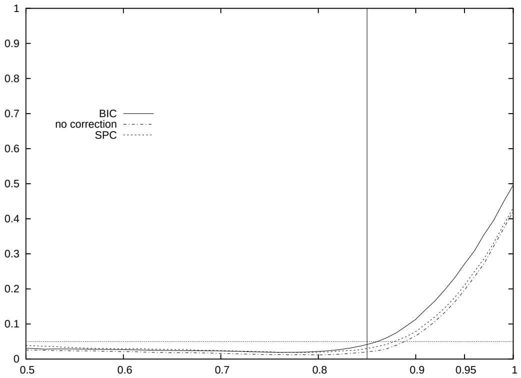
where $\varepsilon_t \sim i.i.d.N(0,1)$ and $\beta = 0$ throughout the simulations because the test statistic is invariant to the true values of β. For the AR(1) case, we set ϕ_1 to be from 0.5 to 1.0 in increments of 0.01 and take $\phi_2 = 0$. On the other hand, for the AR(2) case, we take $\phi_2 = 0.3$ or -0.3 and set ϕ_1 such that $\phi_1 + \phi_2$ ranges from 0.5 to 1.0. The significance level is 0.05 and the number of replications is 5,000.

To obtain the estimate of the long-run variance, $\tilde{\omega}_{AR}$, we need to determine the lag length in practice. For both of the AR(1) and AR(2) cases we choose the lag length using the Bayesian information criterion (BIC)[4]. In addition, in order to apply the boundary rule, we have to preset the value of c; however, the localizing parameter c is not necessarily interesting in practical analysis. The boundary value, $1 - c/\sqrt{T}$, is of greater importance in finite samples because we truncate the long-run parameter $\phi_1 + \cdots + \phi_p$ at the boundary value. For example in the AR(1) case, if we set the boundary value as 0.9, we expect that the size of the test would be close to the significance level when ϕ_1 is less than 0.9. On the other hand, when ϕ_1 is greater than 0.9, the null hypothesis would tend to be rejected. In our simulations, we choose c such that $1 - c/\sqrt{T}$ equals 0.85, 0.9 and 0.95 for $T = 50$, 100 and 300; further, the boundary value of 0.98 is considered for $T = 100$, 300 and 500.

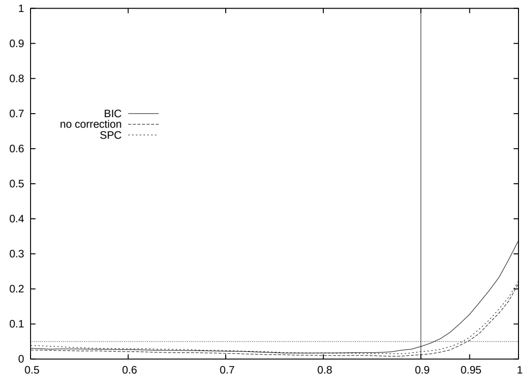
The above boundary value is also used to obtain the bias term $\hat{b}_T$. We used the GAUSS-CML routine to estimate the model by the least squares method with the inequality constraint given by $\phi_1 + \cdots + \phi_p \leq 1 - c/\sqrt{T}$.

Figure 1 provides the rejection frequencies of the tests for the constant AR(1) case where the horizontal axis corresponds to ϕ_1. In each figure, "BIC" denotes the rejection frequencies

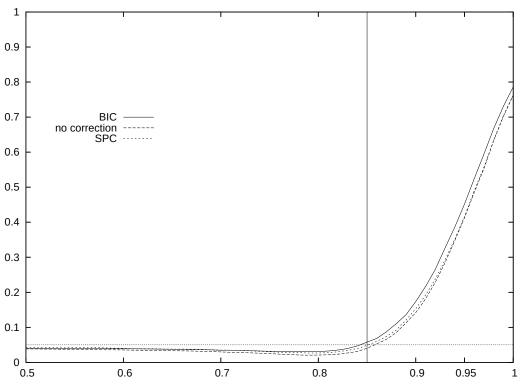
[4]We also used the Akaike information criterion to choose the lag length. The results are similar to those of the BIC.



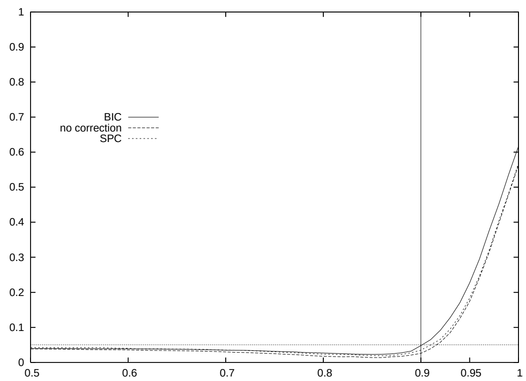
(i-a) $T = 50$, boundary$= 0.85$



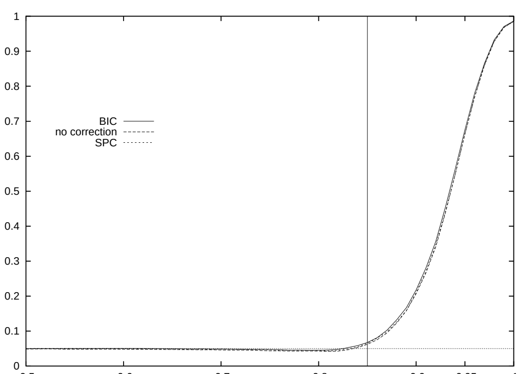
(ii-a) $T = 50$, boundary$= 0.90$



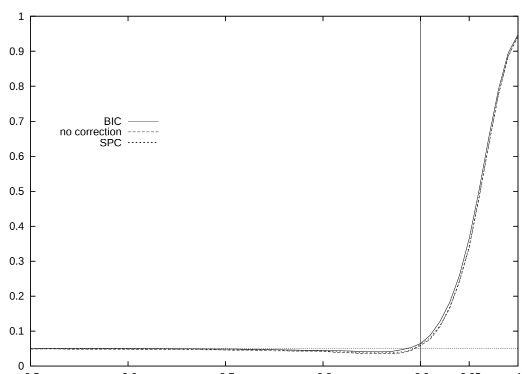
(i-a) $T = 100$, boundary$= 0.85$



(ii-a) $T = 100$, boundary$= 0.90$

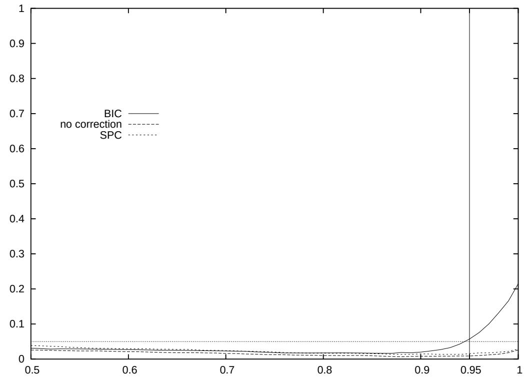


(i-a) $T = 300$, boundary$= 0.85$

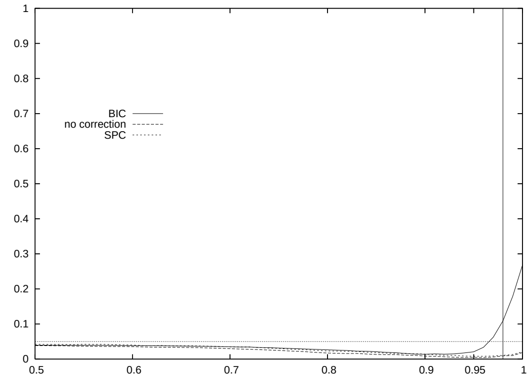


(ii-a) $T = 300$, boundary$= 0.90$

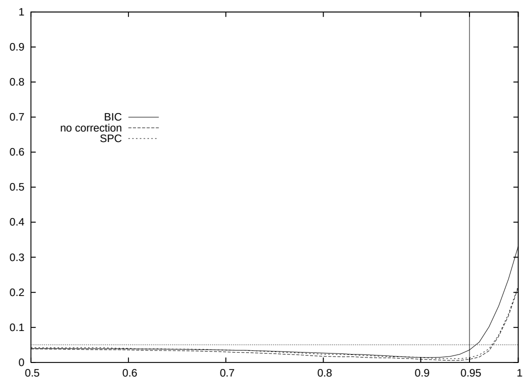
Figure 1: The finite sample performance; constant case (AR(1) model)



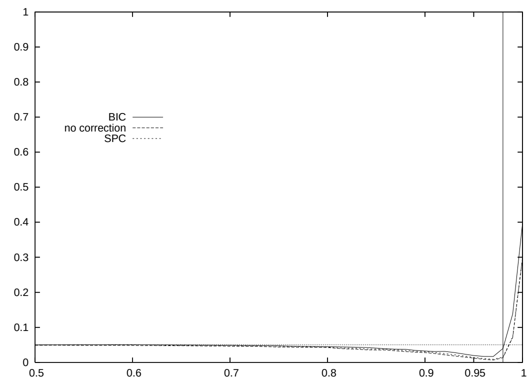
(i-a) $T = 50$, boundary$= 0.95$



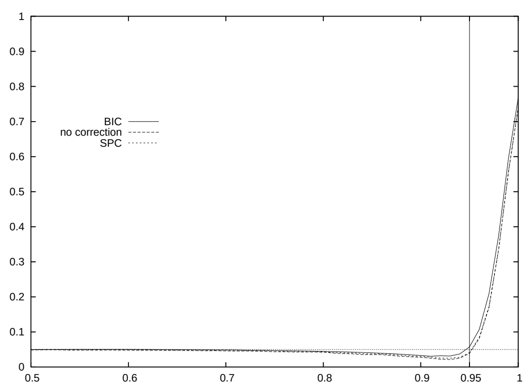
(ii-a) $T = 100$, boundary$= 0.98$



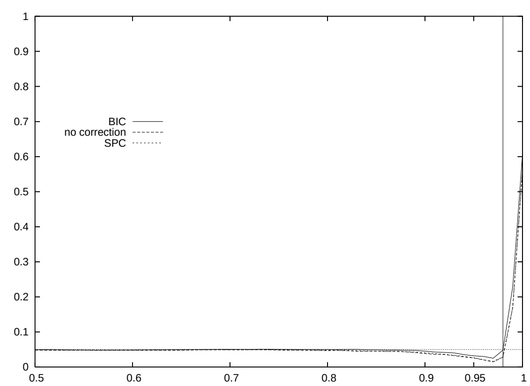
(i-a) $T = 100$, boundary$= 0.95$



(ii-a) $T = 300$, boundary$= 0.98$

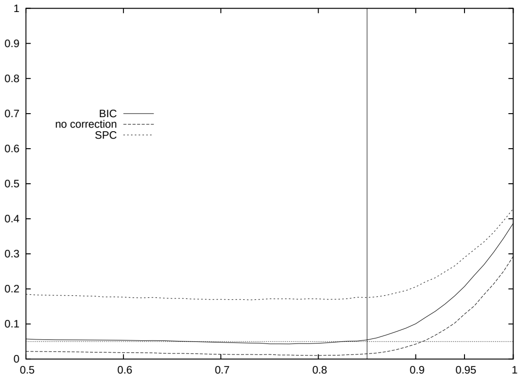


(i-a) $T = 300$, boundary$= 0.95$

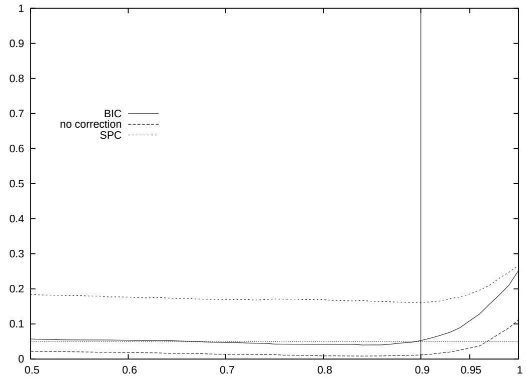


(ii-a) $T = 500$, boundary$= 0.98$

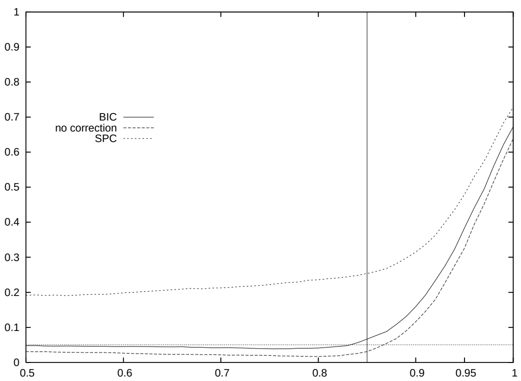
Figure 1: (continued)



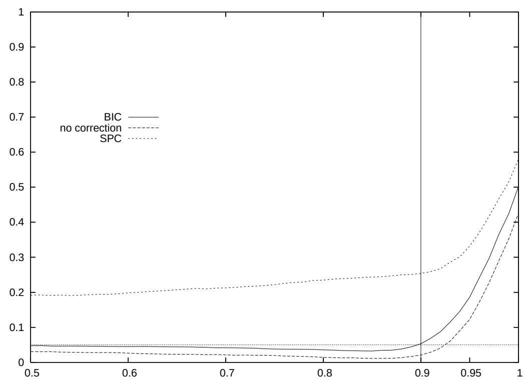
(i-a) $T = 50$, boundary$= 0.85$



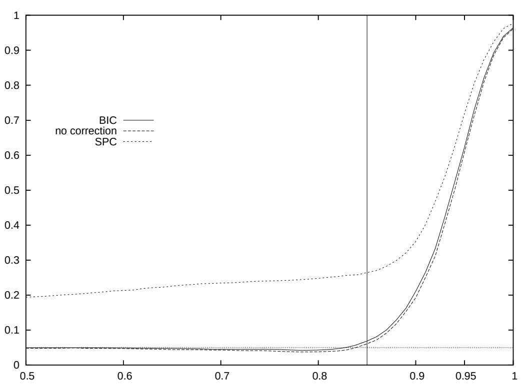
(ii-a) $T = 50$, boundary$= 0.90$



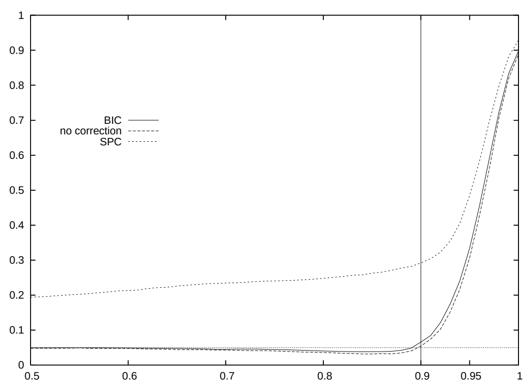
(i-a) $T = 100$, boundary$= 0.85$



(ii-a) $T = 100$, boundary$= 0.90$

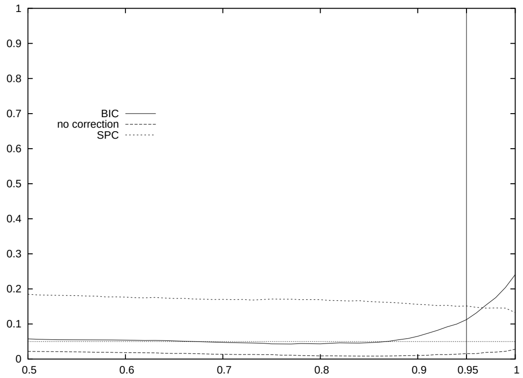


(i-a) $T = 300$, boundary$= 0.85$

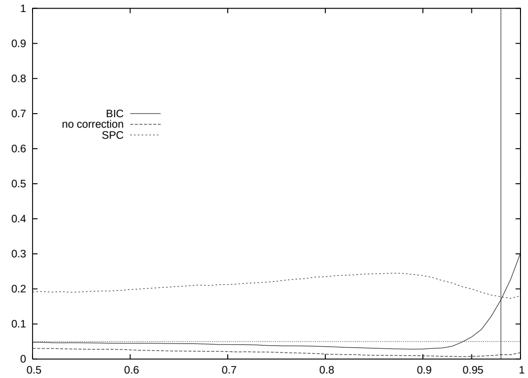


(ii-a) $T = 300$, boundary$= 0.90$

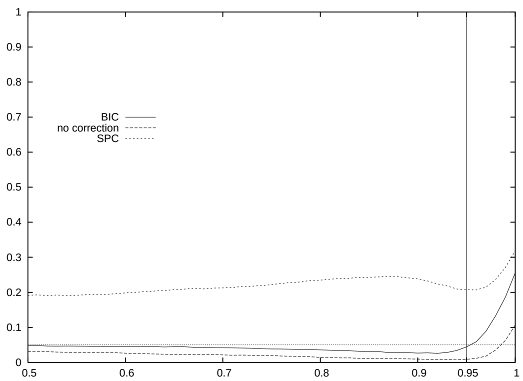
Figure 2: The finite sample performance; constant case (AR(2) with $\phi = 0.3$ model)



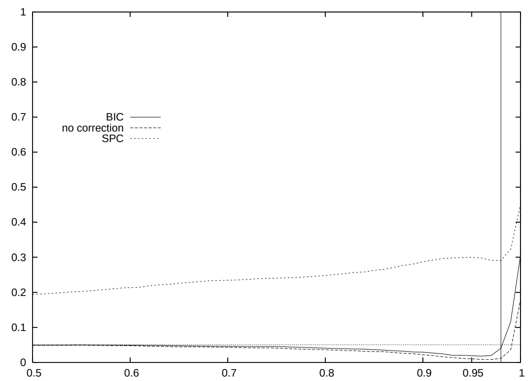
(i-a) $T = 50$, boundary$= 0.95$



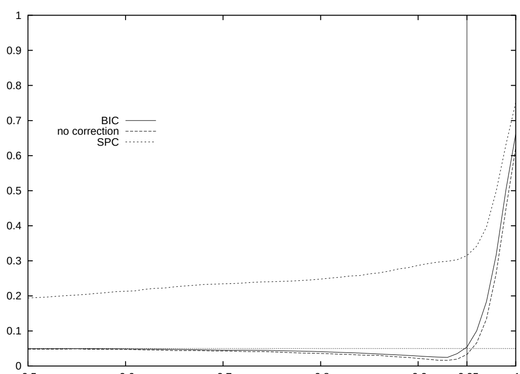
(ii-a) $T = 100$, boundary$= 0.98$



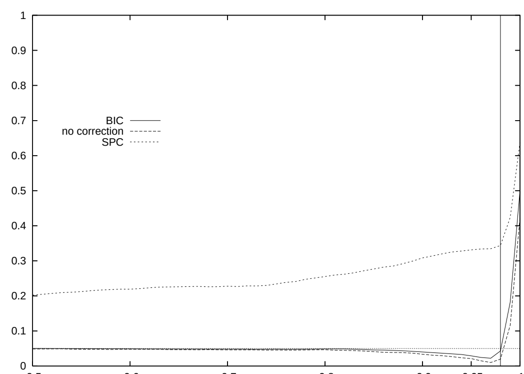
(i-a) $T = 100$, boundary$= 0.95$



(ii-a) $T = 300$, boundary$= 0.98$

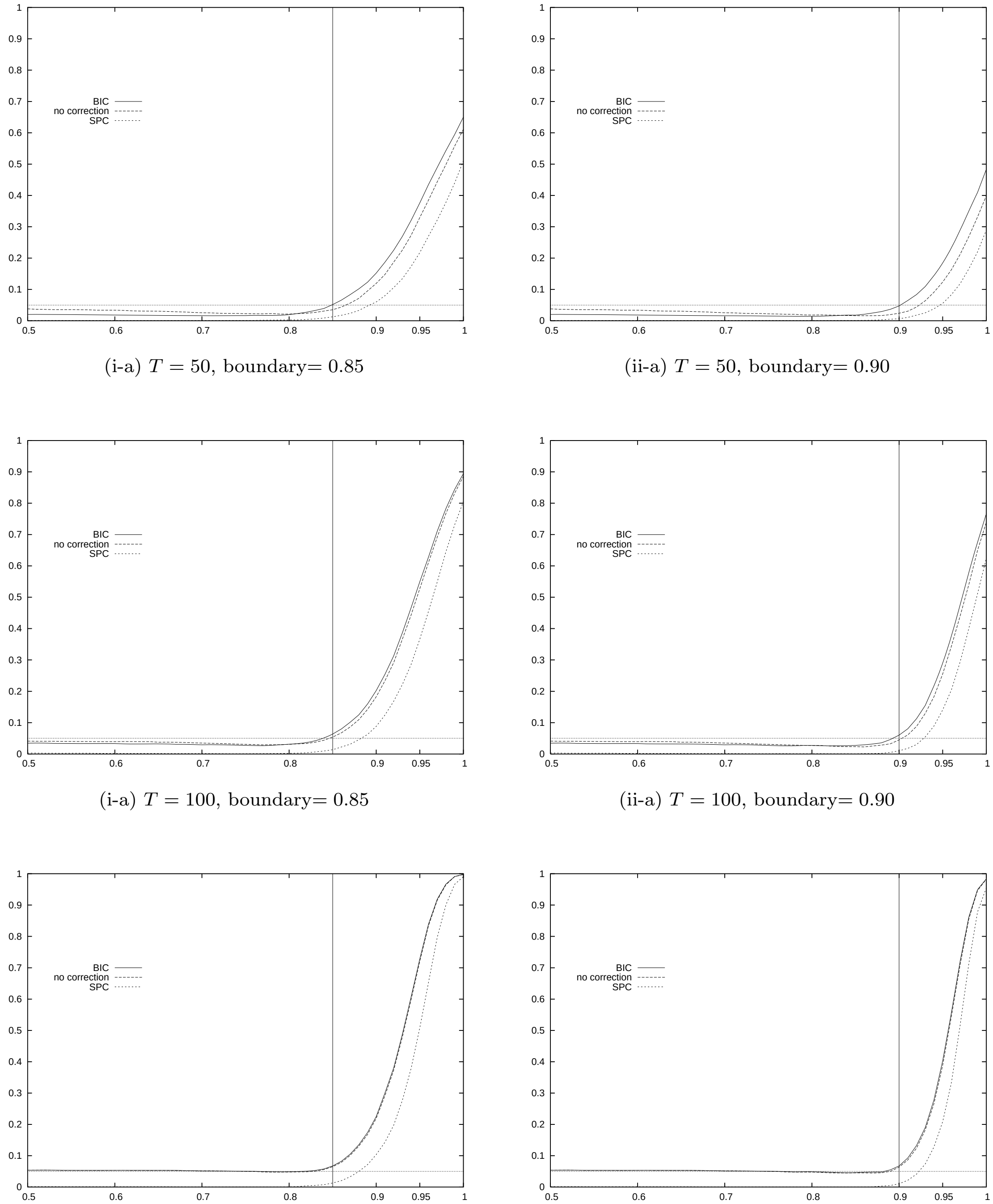


(i-a) $T = 300$, boundary$= 0.95$



(ii-a) $T = 500$, boundary$= 0.98$

Figure 2: (continued)



(i-a) $T = 50$, boundary$= 0.85$

(ii-a) $T = 50$, boundary$= 0.90$

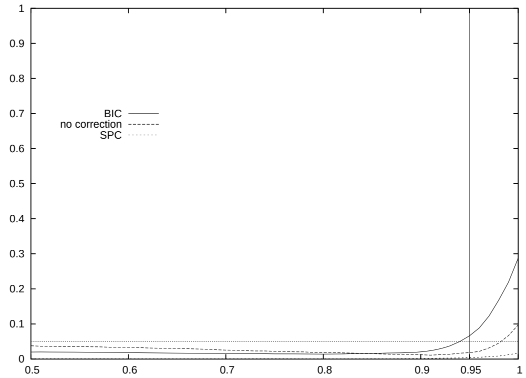
(i-a) $T = 100$, boundary$= 0.85$

(ii-a) $T = 100$, boundary$= 0.90$

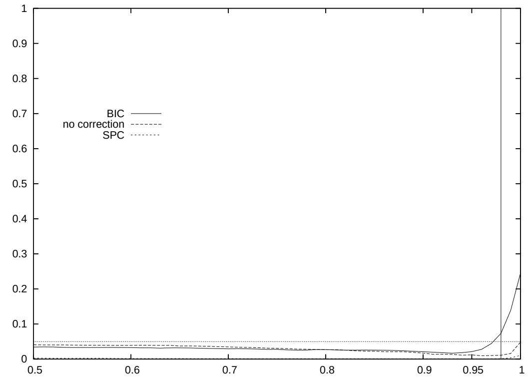
(i-a) $T = 300$, boundary$= 0.85$

(ii-a) $T = 300$, boundary$= 0.90$

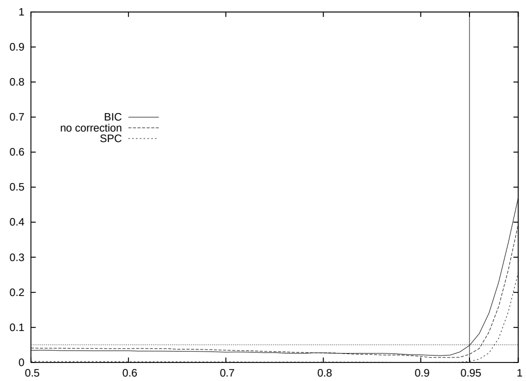
Figure 3: The finite sample performance; constant case (AR(2) with $\phi = -0.3$ model)



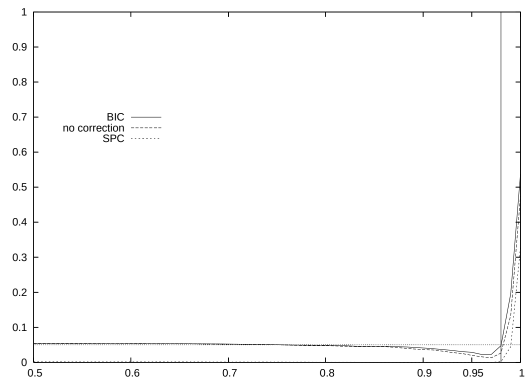
(i-a) $T = 50$, boundary$= 0.95$



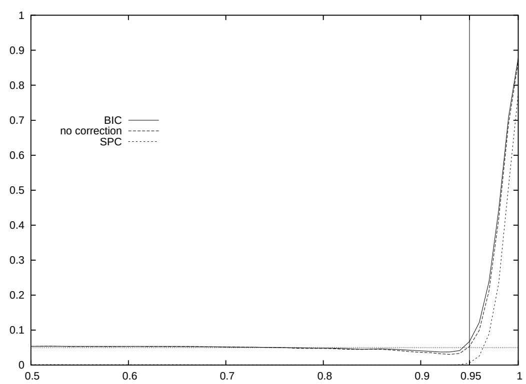
(ii-a) $T = 100$, boundary$= 0.98$



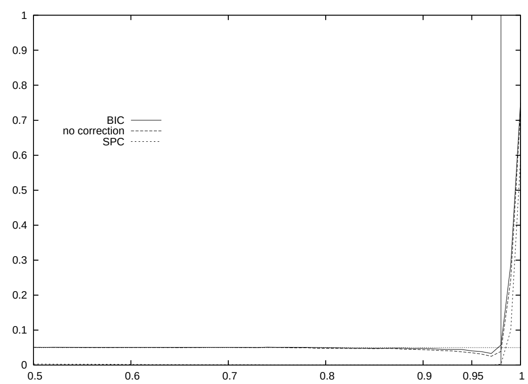
(i-a) $T = 100$, boundary$= 0.95$



(ii-a) $T = 300$, boundary$= 0.98$

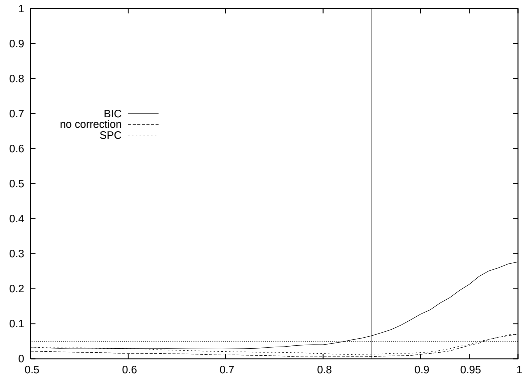


(i-a) $T = 300$, boundary$= 0.95$

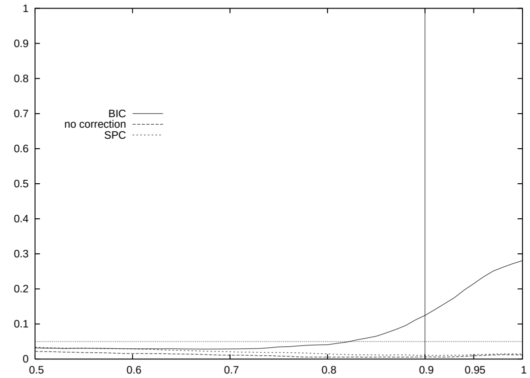


(ii-a) $T = 500$, boundary$= 0.98$

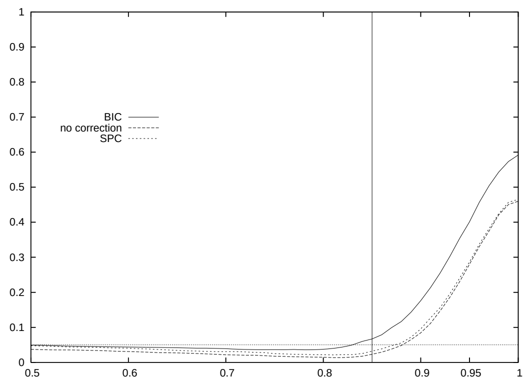
Figure 3: (continued)



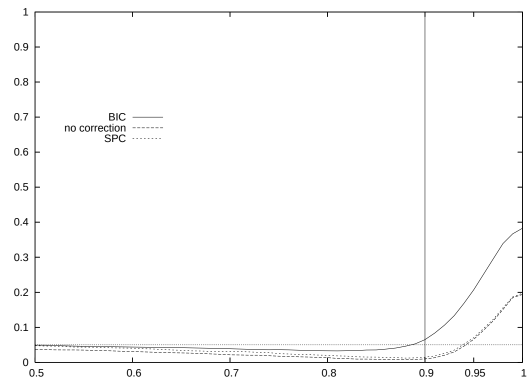
(i-a) $T = 50$, boundary= 0.85



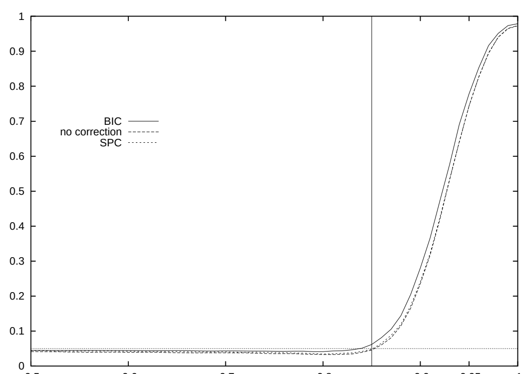
(ii-a) $T = 50$, boundary= 0.90



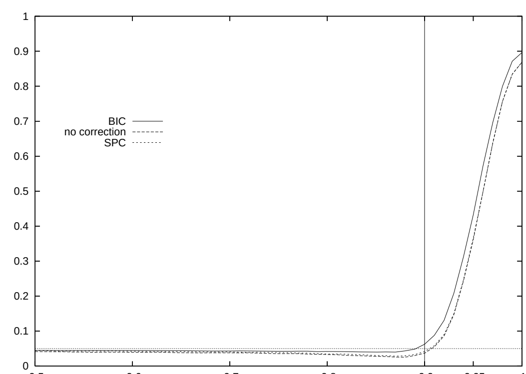
(i-a) $T = 100$, boundary= 0.85



(ii-a) $T = 100$, boundary= 0.90

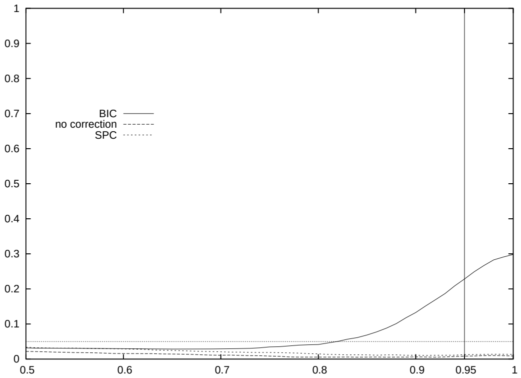


(i-a) $T = 300$, boundary= 0.85

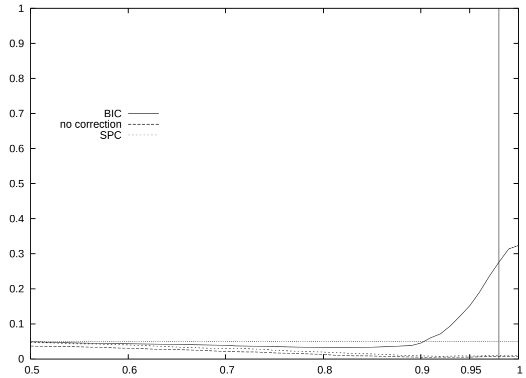


(ii-a) $T = 300$, boundary= 0.90

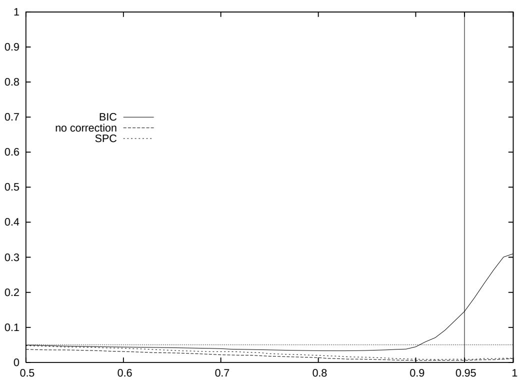
Figure 4: The finite sample performance; trend case (AR(1) model)



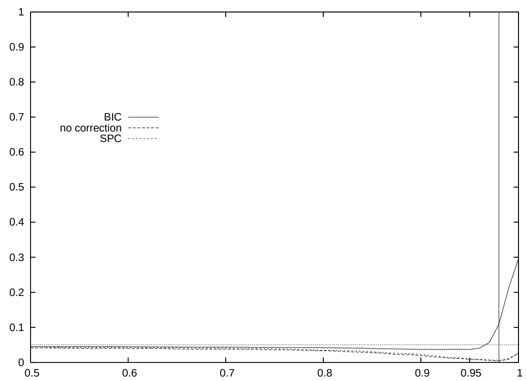
(i-a) $T = 50$, boundary= 0.95



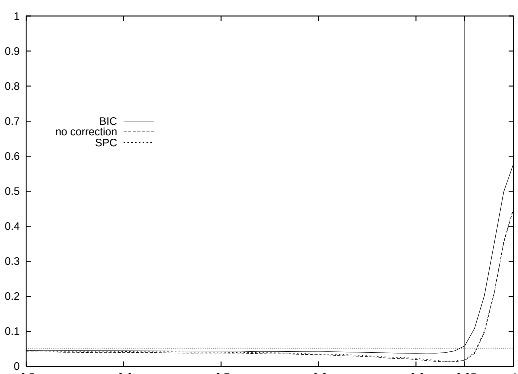
(ii-a) $T = 100$, boundary= 0.98



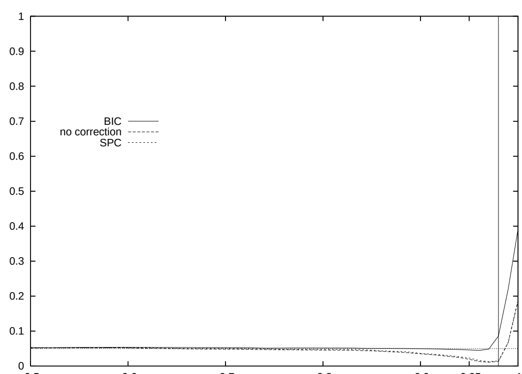
(i-a) $T = 100$, boundary= 0.95



(ii-a) $T = 300$, boundary= 0.98

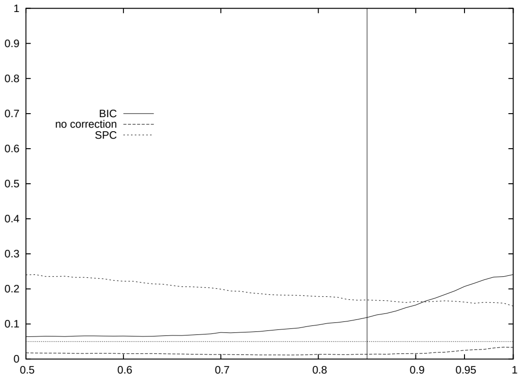


(i-a) $T = 300$, boundary= 0.95

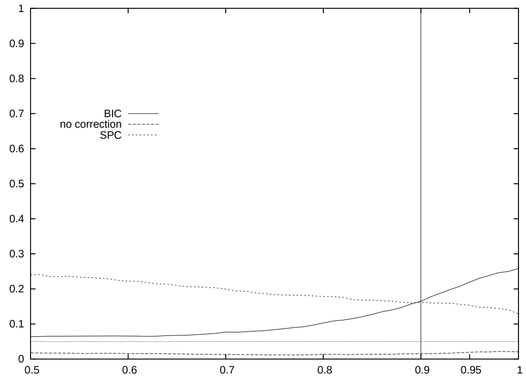


(ii-a) $T = 500$, boundary= 0.98

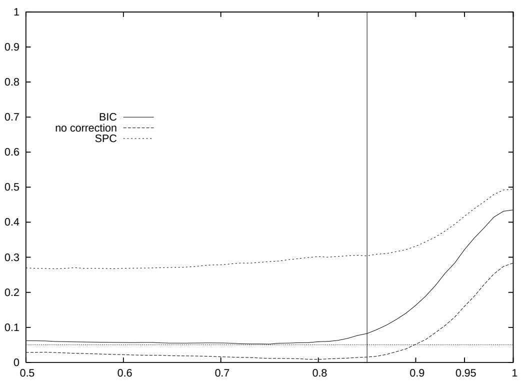
Figure 4: (continued)



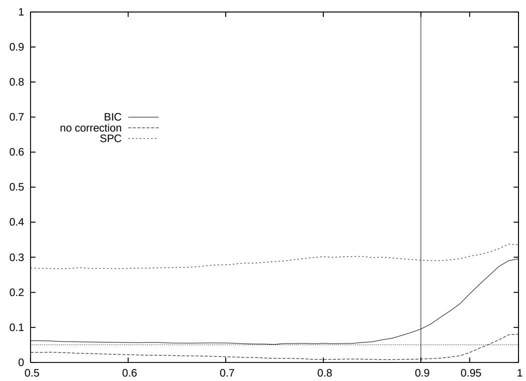
(i-a) $T = 50$, boundary$= 0.85$



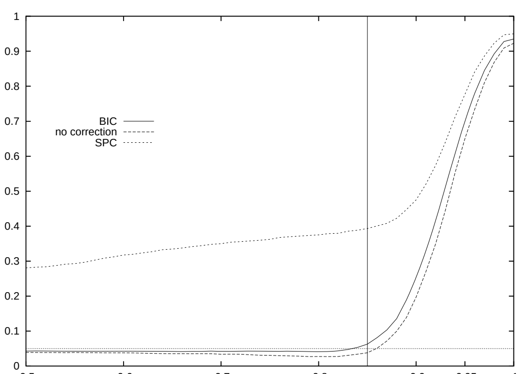
(ii-a) $T = 50$, boundary$= 0.90$



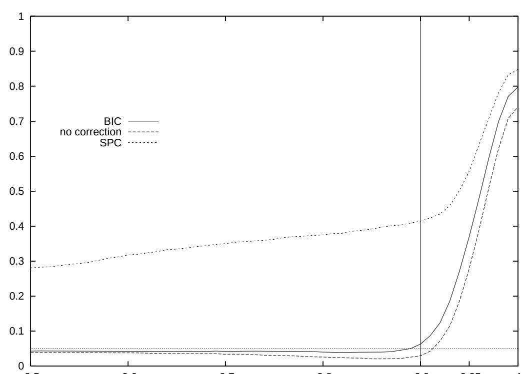
(i-a) $T = 100$, boundary$= 0.85$



(ii-a) $T = 100$, boundary$= 0.90$

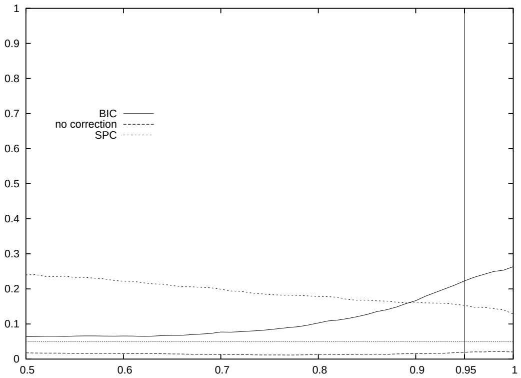


(i-a) $T = 300$, boundary$= 0.85$

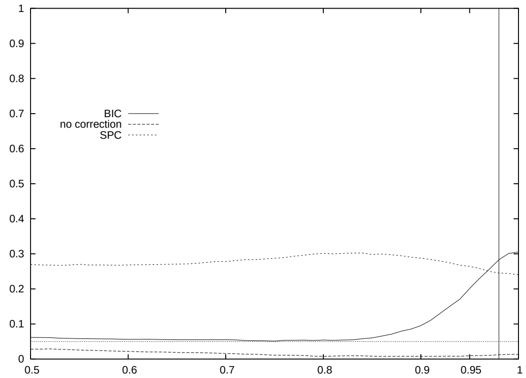


(ii-a) $T = 300$, boundary$= 0.90$

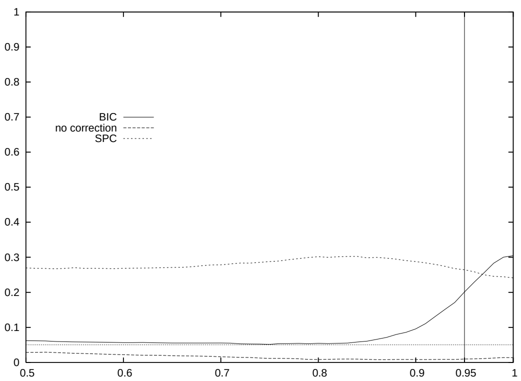
Figure 5: The finite sample performance; trend case (AR(2) with $\phi = 0.3$ model)



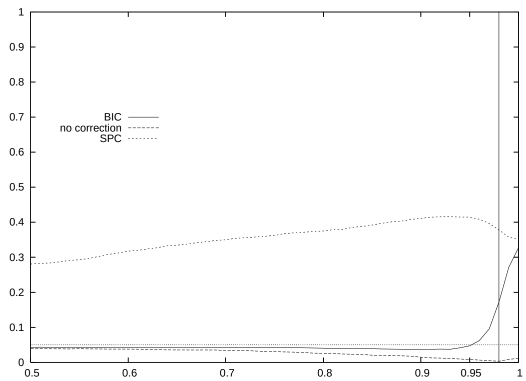
(i-a) $T = 50$, boundary$= 0.95$



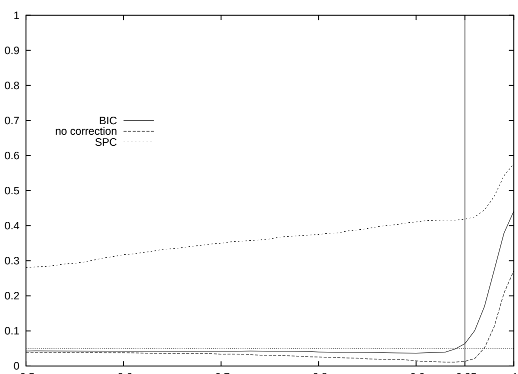
(ii-a) $T = 100$, boundary$= 0.98$



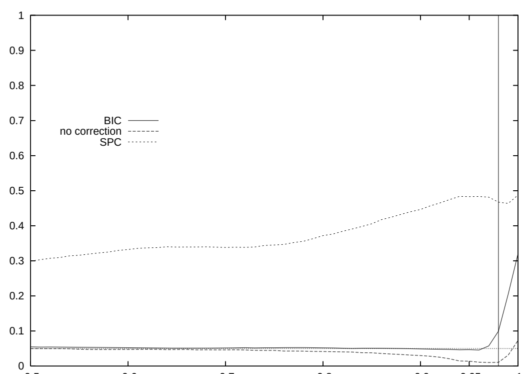
(i-a) $T = 100$, boundary$= 0.95$



(ii-a) $T = 300$, boundary$= 0.98$

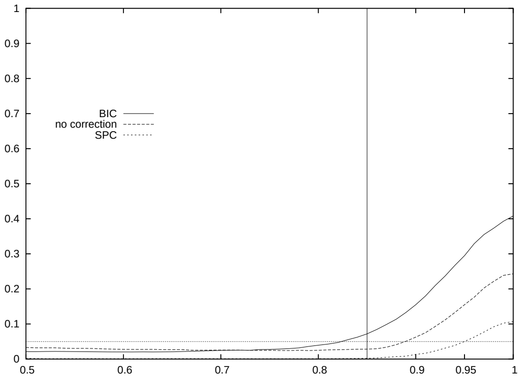


(i-a) $T = 300$, boundary$= 0.95$

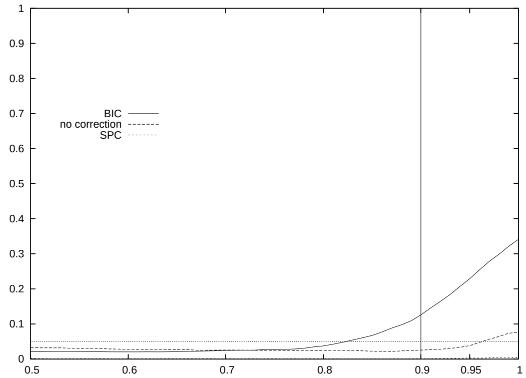


(ii-a) $T = 500$, boundary$= 0.98$

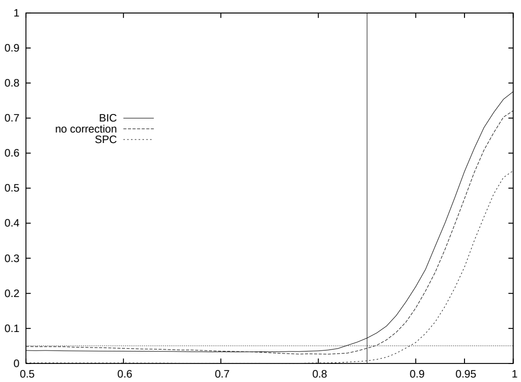
Figure 5: (continued)



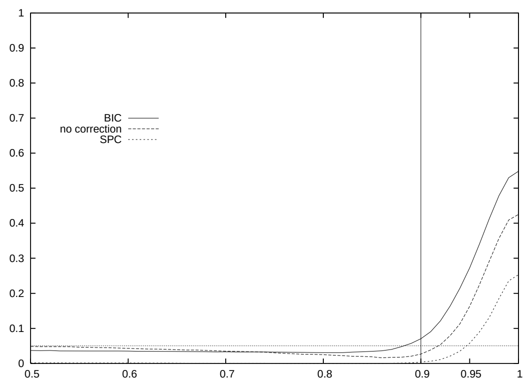
(i-a) $T = 50$, boundary$= 0.85$



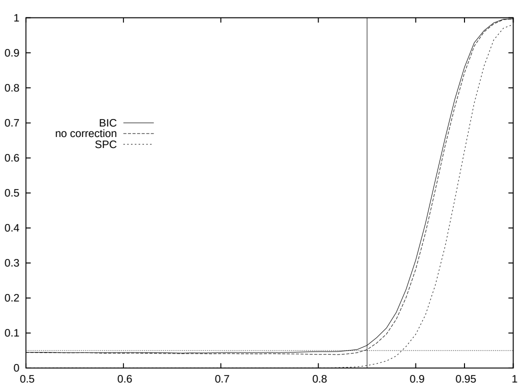
(ii-a) $T = 50$, boundary$= 0.90$



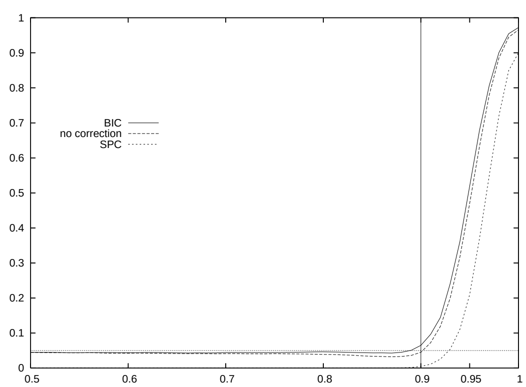
(i-a) $T = 100$, boundary$= 0.85$



(ii-a) $T = 100$, boundary$= 0.90$

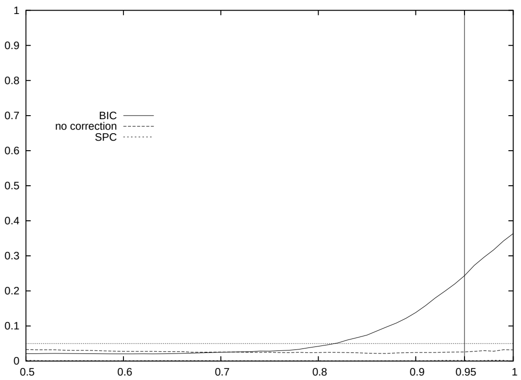


(i-a) $T = 300$, boundary$= 0.85$

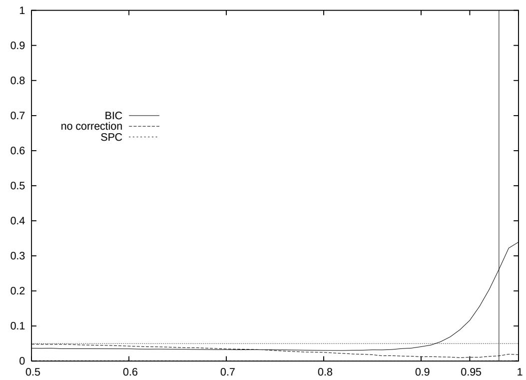


(ii-a) $T = 300$, boundary$= 0.90$

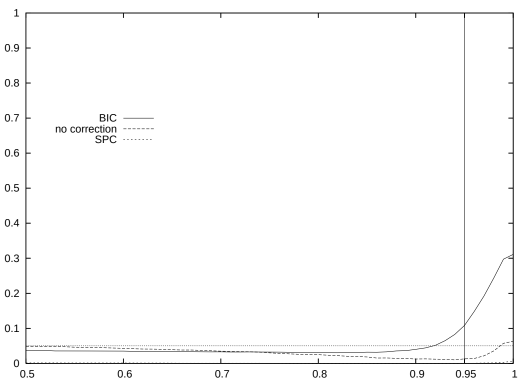
Figure 6: The finite sample performance; trend case (AR(2) with $\phi = -0.3$ model)



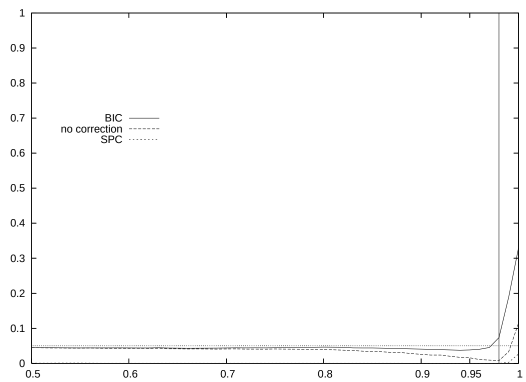
(i-a) $T = 50$, boundary$= 0.95$



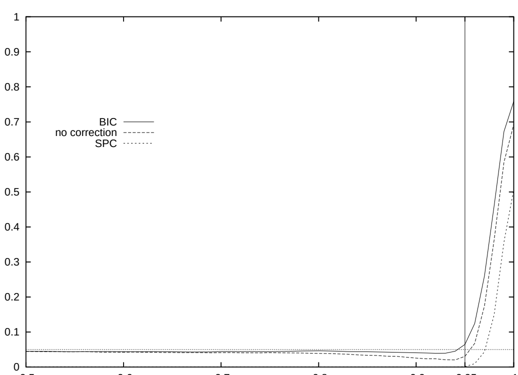
(ii-a) $T = 100$, boundary$= 0.98$



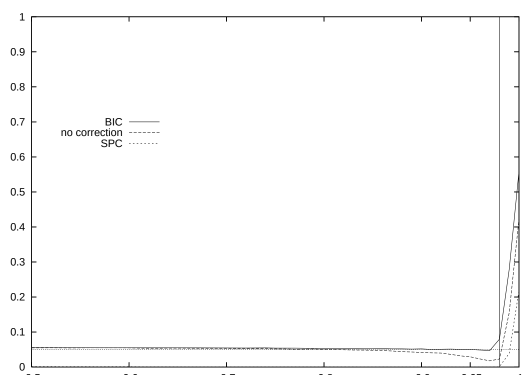
(i-a) $T = 100$, boundary$= 0.95$



(ii-a) $T = 300$, boundary$= 0.98$



(i-a) $T = 300$, boundary$= 0.95$



(ii-a) $T = 500$, boundary$= 0.98$

Figure 6: (continued)

of the bias corrected version of the modified KPSS test while "SPC" is the SPC test with AR(1) prewhitening, to which the above boundary rule is applied. To see the effect of the bias correction, we also show the result of the modified KPSS test in Subsection 3.1 with the true lag length, which is denoted as "no correction". From Figure 1, we can see that the bias correction term b_T effectively reduces the downward bias of the modified KPSS test statistic when the boundary value is 0.85, 0.9 and 0.95. For the boundary value of 0.98, our method overly corrects the modified KPSS test statistic when $T = 100$; however, as the sample size increases, the size of our test gets closer to the nominal size when the AR(1) parameter is less than the boundary value.

Figures 2 and 3 give the rejection frequencies of the tests for the constant AR(2) case. In this case, the horizontal axis corresponds to $\phi_1 + \phi_2$. Figure 2 shows that the SPC test with the AR(1) prewhitening suffers from size distortion when $\phi_2 = 0.3$ even if the process is not strongly serially correlated, while the modified KPSS test tends to under-reject the null hypothesis. Again, the size of the bias corrected version of the modified KPSS test is much closer to the nominal one than the other tests except when T is small and the boundary values are large. On the other hand, the SPC test tends to under-reject the null hypothesis when $\phi_2 = -0.3$, and as such, it loses power considerably.

Figures 4 to 6 give the results for the trend case. The relative performance of the tests is preserved compared to the constant case but it becomes more difficult to control the size of the tests for the trend case. When the sample size is small and the boundary is close to one, our method tends to correct the downward bias too much, so that our test also suffers from size distortion in some cases. However, this over-rejection is mitigated as the sample size increases.

5. Concluding Remarks

In this paper, we proposed a new KPSS-type test for stationarity with less size distortion. The distinctive features of our test are summarized in the following: First, we parametrically estimate the long-run variance imposing the boundary condition. Second, we correct the downward bias in the numerator of the KPSS test statistic. The simulation study showed that our method can mitigate the size distortion problem effectively. Our method could be extended to the panel stationarity test proposed by Hadri (2000) and it is our future research.

It is worth noting that we can control the empirical size only up to the boundary value and that the rejection frequencies of the test tend to be greater than the significance level when the long-run parameter is above the boundary value. This may be a natural result in view of Theorem 2 of Müller (2008), who pointed out (Discussion 4.3, Müller (2008)) that "Consistent stationarity tests should be thought of as testing jointly the I(0) property ... and additional restrictions on the behavior of the process." The boundary rule in this paper is one such restriction.

Appendix

Proof of Theorem 1: In the following, we will show that

$$
\begin{aligned}
E\left[R_1\right] &= \frac{1}{T} b_0 \gamma_0 + o\left(\frac{1}{T}\right) &(12)\\
E\left[R_2\right] &= \frac{1}{T} \frac{b_0 \sigma_\varepsilon^2 \phi'(1)}{\phi^3(1)}, &(13)
\end{aligned}
$$

where $b_0 = 5/3$ for the constant case and it is $19/15$ for the trend case.

Let us first consider the trend case. Since $\widehat{\Delta v_t}$ is the regression residual of Δv_t on d_t, we can expand R_1 as

$$
\begin{aligned}
R_1 &= \frac{1}{T^2} \sum_{t=1}^{T} \left\{ \sum_{s=1}^{t} \Delta v_s - \sum_{s=1}^{t} d_s' \left(\sum_{t=1}^{T} d_t d_t' \right)^{-1} \sum_{t=1}^{T} d_t \Delta v_t \right\}^2 \\
&= \frac{1}{T^2} \sum_{t=1}^{T} \left(\sum_{s=1}^{t} \Delta v_s \right)^2 - \frac{2}{T^2} \left(\sum_{t=1}^{T} \sum_{s=1}^{t} \Delta v_s \sum_{s=1}^{t} d_s' \right) \left(\sum_{t=1}^{T} d_t d_t' \right)^{-1} \sum_{t=1}^{T} d_t \Delta v_t \\
&\quad + \frac{1}{T^2} \sum_{t=1}^{T} \Delta v_t d_t' \left(\sum_{t=1}^{T} d_t d_t' \right)^{-1} \left(\sum_{t=1}^{T} \sum_{s=1}^{t} d_s \sum_{s=1}^{t} d_s' \right) \left(\sum_{t=1}^{T} d_t d_t' \right)^{-1} \sum_{t=1}^{T} d_t \Delta v_t \\
&= R_{11} - R_{12} + R_{13}, \quad \text{say.}
\end{aligned}
$$

Since $\sum_{s=1}^{t} \Delta v_s = v_t - v_0$, the expectation of R_{11} becomes

$$
\begin{aligned}
E[R_{11}] &= \frac{1}{T^2} \sum_{t=1}^{T} E\left[v_t^2 - 2 v_0 v_t + v_0^2 \right] \\
&= \frac{2}{T} \gamma_0 - \frac{2}{T^2} \sum_{t=1}^{T} \gamma_t = \frac{2}{T} \gamma_0 + O\left(\frac{1}{T^2} \right) &(14)
\end{aligned}
$$

because $|\sum_{t=1}^{T} \gamma_t| \leq \sum_{t=0}^{\infty} |\gamma_t| < \infty$.

In order to evaluate R_{12}, we express $E[R_{12}]$ as

$$
E[R_{12}] = \frac{2}{T^2} tr \left\{ \left(\sum_{t=1}^{T} d_t d_t' \right)^{-1} E\left[\left(\sum_{t=1}^{T} d_t \Delta v_t \right) \left(\sum_{t=1}^{T} \sum_{s=1}^{t} \Delta v_s \sum_{s=1}^{t} d_s' \right) \right] \right\}. \qquad (15)
$$

We evaluate each element of the expectation on the right hand side of (15). The (1,1) element becomes

$$
\begin{aligned}
E[\text{the (1,1) element}] &= E\left[(v_T - v_0)\sum_{t=1}^{T}(v_t - v_0)t\right] \\
&= \sum_{t=1}^{T} t\gamma_{T-t} - \gamma_T \sum_{t=1}^{T} t - \sum_{t=1}^{T} t\gamma_t + \gamma_0 \sum_{t=1}^{T} t \\
&= \frac{T^2}{2}\gamma_0 + o(T^2) \qquad (16)
\end{aligned}
$$

because γ_t is absolutely summable and $\gamma_T = o(1)$. In exactly the same way, we can see that

$$
\begin{aligned}
E[\text{the (1,2) element}] &= E\left[(v_T - v_0)\sum_{t=1}^{T}(v_t - v_0)\sum_{s=1}^{t} s\right] \\
&= \sum_{t=1}^{T}\gamma_{T-t}\sum_{s=1}^{t} s - \gamma_T \sum_{t=1}^{T}\sum_{s=1}^{t} s - \sum_{t=1}^{T}\gamma_t \sum_{s=1}^{t} s + \gamma_0 \sum_{t=1}^{T}\sum_{s=1}^{t} s \\
&= \frac{T^3}{6}\gamma_0 + o(T^3), \qquad (17)
\end{aligned}
$$

$$
\begin{aligned}
E[\text{the (2,1) element}] &= E\left[\left\{(T+1)v_T - \sum_{t=1}^{T} v_t - v_0\right\}\sum_{t=1}^{T}(v_t - v_0)t\right] \\
&= (T+1)\sum_{t=1}^{T} t\gamma_{T-t} - (T+1)\gamma_T \sum_{t=1}^{T} t - E\left[\sum_{t=1}^{T} v_t \sum_{t=1}^{T} tv_t\right] \\
&\quad + \sum_{t=1}^{T}\gamma_t \sum_{t=1}^{T} t - \sum_{t=1}^{T} t\gamma_t + \gamma_0 \sum_{t=1}^{T} t = o(T^3) \qquad (18)
\end{aligned}
$$

because $|E[\sum_{t=1}^{T} v_t \sum_{t=1}^{T} tv_t]| \le \sqrt{E[(\sum_t v_t)^2]E[(\sum_t tv_t)^2]} \le \sqrt{O(T)O(T^3)}$ by the Cauchy-Schwarz inequality, and

$$
\begin{aligned}
E[\text{the (2,2) element}] &= E\left[\left\{(T+1)v_T - \sum_{t=1}^{T} v_t - v_0\right\}\sum_{t=1}^{T}(v_t - v_0)\sum_{s=1}^{t} s\right] \\
&= (T+1)\sum_{t=1}^{T}\gamma_{T-t}\sum_{s=1}^{t} s - (T+1)\gamma_T \sum_{t=1}^{T}\sum_{s=1}^{t} s - E\left[\sum_{t=1}^{T} v_t \sum_{t=1}^{T} v_t \sum_{s=1}^{t} s\right] \\
&\quad + \sum_{t=1}^{T}\gamma_t \sum_{s=1}^{t} s - \sum_{t=1}^{T}\gamma_t \sum_{s=1}^{t} s + \gamma_0 \sum_{t=1}^{T}\sum_{s=1}^{t} s = o(T^4). \qquad (19)
\end{aligned}
$$

Since direct calculation yields

$$\left(\sum_{t=1}^{T} d_t d_t'\right)^{-1} = \begin{bmatrix} \dfrac{2(2T+1)}{T(T-1)} & -\dfrac{6}{T(T-1)} \\ -\dfrac{6}{T(T-1)} & \dfrac{12}{T(T^2-1)} \end{bmatrix}, \tag{20}$$

we have, using (15)–(19),

$$E[R_{12}] = \frac{2}{T}\gamma_0 + o\left(\frac{1}{T}\right). \tag{21}$$

The expectation of R_{13} is obtained in a similar manner. We first express $E[R_{13}]$ as

$$E[R_{13}] = \frac{1}{T^2} tr \left\{ \left(\sum_{t=1}^{T} d_t d_t'\right)^{-1} \left(\sum_{t=1}^{T}\sum_{s=1}^{t} d_s \sum_{s=1}^{t} d_s'\right) \left(\sum_{t=1}^{T} d_t d_t'\right)^{-1} E\left[\sum_{t=1}^{T} d_t \Delta v_t \sum_{t=1}^{T} \Delta v_t d_t'\right] \right\}.$$

Since it is shown that

$$\begin{aligned} &\left(\sum_{t=1}^{T} d_t d_t'\right)^{-1} \left(\sum_{t=1}^{T}\sum_{s=1}^{t} d_s \sum_{s=1}^{t} d_s'\right) \left(\sum_{t=1}^{T} d_t d_t'\right)^{-1} \\ &= \begin{bmatrix} \dfrac{(17T^2-10T+2)(T+1)}{15T(T-1)} & -\dfrac{11T^2-5T+6}{10T(T-1)} \\ -\dfrac{11T^2-5T+6}{10T(T-1)} & \dfrac{6(T^2+1)}{5(T^2-1)T} \end{bmatrix}, \end{aligned} \tag{22}$$

$$E\left[\sum_{t=1}^{T} d_t \Delta v_t \sum_{t=1}^{T} \Delta v_t d_t'\right] = \begin{bmatrix} 2\gamma_0 + o(1) & T\gamma_0 + o(T) \\ T\gamma_0 + o(T) & T^2\gamma_0 + O(T) \end{bmatrix},$$

we have

$$E[R_{13}] = \frac{19}{15T}\gamma_0 + o\left(\frac{1}{T}\right). \tag{23}$$

Hence, we obtain (12) from (14), (21) and (23).

The evaluation of the expectation of R_2 proceeds in the same way but is more complicated. We first expand R_2, except for the scalar, $2\psi(1)/T^2$, as

$$
\begin{aligned}
&\sum_{t=1}^{T}\left(\sum_{s=1}^{t}\hat{\varepsilon}_s\right)\left(\sum_{s=1}^{t}\widehat{\Delta v_s}\right) \\
= \;& \sum_{t=1}^{T}\left\{\sum_{s=1}^{t}\varepsilon_s - \sum_{s=1}^{t}d_s'\left(\sum_{t=1}^{T}d_t d_t'\right)^{-1}\sum_{t=1}^{T}d_t\varepsilon_t\right\} \\
& \left\{\sum_{s=1}^{t}\Delta v_s - \sum_{s=1}^{t}d_s'\left(\sum_{t=1}^{T}d_t d_t'\right)^{-1}\sum_{t=1}^{T}d_t\Delta v_t\right\} \\
= \;& \sum_{t=1}^{T}\sum_{s=1}^{t}\varepsilon_s\sum_{s=1}^{t}\Delta v_s - \sum_{t=1}^{T}\sum_{s=1}^{t}\Delta v_s\sum_{s=1}^{t}d_s'\left(\sum_{t=1}^{T}d_t d_t'\right)^{-1}\sum_{t=1}^{T}d_t\varepsilon_t \\
& -\sum_{t=1}^{T}\sum_{s=1}^{t}\varepsilon_s\sum_{s=1}^{t}d_s'\left(\sum_{t=1}^{T}d_t d_t'\right)^{-1}\sum_{t=1}^{T}d_t\Delta v_t \\
& +\sum_{t=1}^{T}\varepsilon_t d_t'\left(\sum_{t=1}^{T}d_t d_t'\right)^{-1}\sum_{t=1}^{T}\sum_{s=1}^{t}d_s\sum_{s=1}^{t}d_s'\left(\sum_{t=1}^{T}d_t d_t'\right)^{-1}\sum_{t=1}^{T}d_t\Delta v_t \\
= \;& R_{21} - R_{22} - R_{23} + R_{24}, \quad \text{say.}
\end{aligned}
$$

We evaluate each term. The expectation of R_{21} becomes

$$
E[R_{21}] = \sum_{t=1}^{T}E\left[\sum_{s=1}^{t}\varepsilon_s(v_t - v_0)\right] = \sigma_\varepsilon^2\sum_{t=1}^{T}\sum_{s=1}^{t}\tilde{\psi}_{t-s} = \sigma_\varepsilon^2 T\sum_{t=0}^{T-1}\left(1-\frac{t}{T}\right)\tilde{\psi}_t. \tag{24}
$$

In order to evaluate the expectations of R_{22}, R_{23} and R_{24}, we use the following lemma.

Lemma 1. *Let f_t and g_t be deterministic sequences for $t = 1, \cdots, T$. Then,*

$$
\begin{aligned}
E\left[\left(\sum_{t=1}^{T}f_t\varepsilon_t\right)\left(\sum_{t=1}^{T}g_t v_t\right)\right] &= \sigma_\varepsilon^2\sum_{t=0}^{T-1}\left(\sum_{s=1}^{T-t}f_s g_{s+t}\right)\tilde{\psi}_t, &(25)\\
\sum_{t=1}^{T}f_t\sum_{s=1}^{t}\varepsilon_s &= \sum_{t=1}^{T}\left(\sum_{s=t}^{T}f_s\right)\varepsilon_t. &(26)
\end{aligned}
$$

We omit the proof of Lemma 1 because it is directly obtained by noting that $E[\varepsilon_{t-s}v_t] = \sigma_\varepsilon^2\tilde{\psi}_s$ for $s \geq 0$ and $E[\varepsilon_{t+s}v_t] = 0$ for $s > 0$.

For $E[R_{22}]$, note that

$$
E[R_{22}] = tr\left\{\left(\sum_{t=1}^{T}d_t d_t'\right)^{-1}E\left[\sum_{t=1}^{T}d_t\varepsilon_t\sum_{t=1}^{T}\sum_{s=1}^{t}\Delta v_s\sum_{s=1}^{t}d_s'\right]\right\}. \tag{27}
$$

Each element of the expectation on the right hand side of (27) is evaluated as

$$\begin{aligned}
E[\text{the (1,1) element}] &= E\left[\sum_{t=1}^{T}\varepsilon_t\sum_{t=1}^{T}tv_t\right] = \sigma_\varepsilon^2\sum_{t=0}^{T-1}\frac{T^2-t^2}{2}\tilde{\psi}_t + O(T),\\
E[\text{the (2,1) element}] &= E\left[\sum_{t=1}^{T}t\varepsilon_t\sum_{t=1}^{T}tv_t\right] = \sigma_\varepsilon^2\sum_{t=0}^{T-1}\left(\frac{T^3}{3}-\frac{tT^2}{2}+\frac{t^3}{6}\right)\tilde{\psi}_t + O(T^2),\\
E[\text{the (1,2) element}] &= E\left[\sum_{t=1}^{T}\varepsilon_t\sum_{t=1}^{T}v_t\sum_{s=1}^{t}s\right] = \sigma_\varepsilon^2\sum_{t=0}^{T-1}\frac{T^3-t^3}{6}\tilde{\psi}_t + O(T^2),\\
E[\text{the (2,2) element}] &= E\left[\sum_{t=1}^{T}t\varepsilon_t\sum_{t=1}^{T}v_t\sum_{s=1}^{t}s\right] = \sigma_\varepsilon^2\sum_{t=0}^{T-1}\left(\frac{T^4}{8}+\frac{t^4}{24}-\frac{tT^3}{6}\right)\tilde{\psi}_t + O(T^3),
\end{aligned}$$

where we used (25) with $f_t = 1$ and $g_t = t$ for the (1,1) element, $f_t = t$ and $g_t = t$ for the (2,1) element, $f_t = 1$ and $g_t = t(t+1)/2$ for the (1,2) element, and $f_t = t$ and $g_t = t(t+1)/2$ for the (2,2) element. Then, using (20) it is shown that

$$E[R_{22}] = \sigma_\varepsilon^2 T\sum_{t=0}^{T-1}\frac{1}{2}\left(1-\frac{4t^2}{T^2}+\frac{2t}{T}+\frac{t^4}{T^4}\right)\tilde{\psi}_t + O(1). \tag{28}$$

For R_{23}, we note that

$$\begin{aligned}
\sum_{t=1}^{T}\sum_{s=1}^{t}\varepsilon_s\sum_{s=1}^{t}d_s' &= \left[\sum_{t=1}^{T}t\sum_{s=1}^{t}\varepsilon_s, \sum_{t=1}^{T}\sum_{s=1}^{t}s\sum_{s=1}^{t}\varepsilon_s\right]\\
&= \left[\sum_{t=1}^{T}\left(\frac{T^2-t^2}{2}+O(T)\right)\varepsilon_t,\ \sum_{t=1}^{T}\left(\frac{T^3-t^3}{6}+O(T^2)\right)\varepsilon_t\right]
\end{aligned}$$

where the last expression is obtained by using (26). Hence, using (25), it is shown that

$$\begin{aligned}
&E\left[\left(\sum_{t=1}^{T}d_t\Delta v_t\right)\left(\sum_{t=1}^{T}\sum_{s=1}^{t}\varepsilon_s\sum_{s=1}^{t}d_s'\right)\right]\\
&= \sigma_\varepsilon^2 E\begin{bmatrix} \sum_{t=0}^{T-1}\frac{T^2-(T-t)^2}{2}\tilde{\psi}_t + O(T) & \sum_{t=0}^{T-1}\frac{T^3-(T-t)^3}{6}\tilde{\psi}_t + O(T^2) \\ \sum_{t=0}^{T-1}\left(-\frac{T^3}{3}+tT^2-\frac{t^3}{6}\right)\tilde{\psi}_t + O(T^2) & \sum_{t=0}^{T-1}\left(-\frac{T^4}{8}+\frac{1}{24}t^4+\frac{tT^3}{2}-\frac{t^2T^2}{4}\right)\tilde{\psi}_t + O(T^3)\end{bmatrix}
\end{aligned}$$

Using this result and (20), we have

$$E[R_{23}] = \sigma_\varepsilon^2 T\sum_{t=0}^{T-1}\frac{1}{2}\left(1-\frac{4t^2}{T^2}+\frac{2t}{T}+\frac{t^4}{T^4}\right)\tilde{\psi}_t + O(1). \tag{29}$$

Exactly in the same way, for R_{24}, we have,

$$E\left[\left(\sum_{t=1}^{T} d_t \Delta v_t\right)\left(\sum_{t=1}^{T} \varepsilon_t d_t'\right)\right] = \sigma_\varepsilon^2 \left[\begin{array}{cc} \sum_{t=0}^{T-1} \tilde{\psi}_t & \sum_{t=1}^{T-1} (T-t)\tilde{\psi}_t \\ \sum_{t=0}^{T-1} t\tilde{\psi}_t + O(1) & \sum_{t=0}^{T-1} \frac{T^2-t^2}{2}\tilde{\psi}_t + O(T) \end{array}\right].$$

Then, using (22), it is shown that

$$E[R_{24}] = \sigma_\varepsilon^2 T \sum_{t=0}^{T-1} \left(\frac{19}{30} - \frac{9t^2}{15T^2}\right)\tilde{\psi}_t + O(1). \tag{30}$$

From (24), (28), (29) and (30), we obtain

$$E[R_2] = \frac{2\sigma_\varepsilon^2 \psi(1)}{T} \sum_{t=0}^{T-1} \left(\frac{19}{30} - \frac{3t}{T} + \frac{17t^2}{5T^2} - \frac{t^4}{T^4}\right)\tilde{\psi}_t + o\left(\frac{1}{T}\right).$$

Furthermore, since $\sum_{j=0}^{\infty} |\tilde{\psi}_j| < \infty$, the above summation converges to $\sum_{j=0}^{\infty}(19/30)\tilde{\psi}_t$ and we get

$$E[R_2] = \frac{19\sigma_\varepsilon^2 \psi(1)}{15T} \sum_{j=0}^{\infty} \tilde{\psi}_j + o\left(\frac{1}{T}\right).$$

Noting that $\psi(1) = 1/\phi(1)$ and

$$\sum_{j=0}^{\infty} \tilde{\psi}_j = \psi'(1) = \left(\frac{1}{\phi(1)}\right)' = -\frac{\phi'(1)}{\phi(1)^2},$$

we finally obtain (13).

We obtain a similar expression for the constant case in exactly the same manner and we omit the proof.■

References

[1] Andrews, D., and J. C. Monahan (1992). An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator, *Econometrica* 60, 953-956.

[2] Aznar, A. and M. Ayuda (2008). A Point Optimal Test for the Null of Near Integration, presented at EEA-ESEM meeting 2008, University of Bocconi.

[3] Berk, K. N. (1974). Consistent Autoregressive Spectral Estimates, *Annals of Statistics* 2, 289-502.

[4] Brillinger, D. R. (1981). *Time Series Data Analysis and Theory.* Holden-Day, San Francisco.

[5] Caner, M., and L. Kilian (2001). Size Distortions of Tests of the Null Hypothesis of Stationarity: Evidence and Implications for the PPP Debate, *Journal of International Money and Finance* 20, 639-657.

[6] Carrion-i-Silvestre, J. L., and A. Sansó (2006). A Guide to the Computation of Stationarity Tests, *Empirical Economics* 31, 433-448.

[7] Cheung, Y. W. and M. D. Chinn (1997). Further Investigation of the Uncertain Unit Root in GNP, *Journal of Business and Economic Statistics* 15, 68-73.

[8] Hadri, K. (2000). Testing for Stationarity in Heterogeneous Panel Data. *Econometrics Journal* 3, 148-161.

[9] Harris, D., S. Leybourne and B. McCabe (2007). Modified KPSS Tests for Near Integration, *Econometric Theory* 23, 355-363.

[10] Kuo, B. S. and A. Mikkola (1999). Re-Examining Long-Run Purchasing Power Parity, *Journal of International Money and Finance* 18, 251-266.

[11] Kurozumi, E. (2009). Construction of Stationarity Tests with Less Size Distortions, *Hitotsubashi Journal of Economics* 50, 87-105.

[12] Kwiatkowski, D., P. C. B. Phillips, P. Schmidt and Y. Shin (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root, *Journal of Econometrics* 54, 159-178.

[13] Lanne, M. and P. Saikkonen (2003). Reducing Size Distortions of Parametric Stationarity Tests, *Journal of Time Series Analysis* 24, 423-439.

[14] Leybourne, S. J. and B. P. M. McCabe (1994). A Consistent Test for a Unit Root, *Journal of Business and Economic Statistics* 12, 157-166.

[15] Leybourne, S. J. and B. P. M. McCabe (1999). Modified Stationarity Tests with Data-Dependent Model-Selection Rules, *Journal of Business and Economic Statistics* 17, 264-270.

[16] Müller, U. K. (2005). Size and Power of Tests of Stationarity in Highly Autocorrelated Time Series, *Journal of Econometrics* 128, 264-270.

[17] Müller, U. K. (2008). The Impossibility of Consistent Discrimination between $I(0)$ and $I(1)$ Processes, *Econometric Theory* 24, 616-630.

[18] Perron, P. and S. Ng (1996). Useful Modifications to Some Unit Root Tests with Dependent Errors and Their Local Asymptotic Properties, *Review of Economic Studies* 63, 435-463.

[19] Rothman, P. (1997). More Uncertainty About the Unit Root in U.S. Real GNP, *Journal of Macroeconomics* 19, 771-780.

[20] Saikkonen, P. and R. Luukkonen (1993a). Testing for a Moving Average Unit Root in Autoregressive Integrated Moving Average Models, *Journal of the American Statistical Association* 88, 596-601.

[21] Saikkonen, P. and R. Luukkonen (1993b). Point Optimal Tests for Testing the Order of Differencing in ARIMA Models, *Econometric Theory* 9, 343-362.

[22] Sul, D., P. C. B. Phillips and C. Y. Choi (2005). Prewhitening Bias in HAC Estimation, *Oxford Bulletin of Economics and Statistics* 67, 517-546.