A Simple Cointegrating Rank Test Without Vector Autoregression

by

Mototsugu Shintani

Working Paper No. 00-W44

September 1998
Revised September 2000

DEPARTMENT OF ECONOMICS
VANDERBILT UNIVERSITY
NASHVILLE, TN 37235

www.vanderbilt.edu/econ

A Simple Cointegrating Rank Test Without Vector Autoregression

Mototsugu Shintani*

September 1998

Revised: September 2000

Abstract

This paper proposes a fully nonparametric test for cointegrating rank which does not require estimation of a vector autoregressive model. The test exploits the fact that the degeneracy in the moment matrix of the variables with mixed integration order corresponds to the notion of cointegration. With an appropriate standardization, the test statistics are shown to have a nuisance parameter free limiting distribution and to be consistent under reasonable conditions. Monte Carlo experiments also suggest that the performance of the test is satisfactory with a moderate sample size. The proposed tests are applied to the stochastic growth model using the U.S. aggregate data.

Key words: Cointegration; Long-run variance; Nonparametric spectrum estimation; Unit roots; Variance ratio

JEL classification: C32; C52

*Department of Economics, Vanderbilt University, Nashville, TN 37235 USA. E-mail: mototsugu.shintani@vanderbilt.edu

1 Introduction

Since Engle and Granger's (1987) seminal contribution, determining the number of cointegrating relations in a set of integrated series has received considerable attention from researchers. The most frequently used test in application is Johansen's (1991) likelihood ratio (LR) test based on the parametric estimation of a vector autoregressive (VAR) model. This test utilizes the technique of reduced rank regression since cointegration corresponds to the reduced rank in the first order VAR coefficient in the error correction format. Unlike the residual-based test using single equation cointegrating regression, this LR test has a preferable normalization-free property. However, since the procedure requires the assumption of a correctly specified finite order VAR process, the performance of the test is known to be sensitive to the misspecification.[1] To allow for more general dependency, Saikkonen and Luukkonen (1997) have employed the VAR approximation with lag length that increases to infinity with a rate slower than the sample size. With such a modification, the test can be justified to have a nuisance parameter free limit distribution.

This paper proposes an alternative means of determining the cointegrating rank in a partially nonstationary multivariate system with general dependence. The distinctive feature of the new test is that the test does not require estimation of a VAR model.[2] The test exploits the well-known fact that the number of cointegrating vectors is identical to the degree of degeneration in the space spanned by the sample moment matrix in the limit. This idea of using the degeneracy of the variance matrix was first employed by Phillips and Ouliaris (1990) in the context of testing the no cointegration hypothesis ($\widehat{P}_z$ test in their notation). As emphasized in their paper, such a variance matrix-based test also has the normalization-free property.[3]

[1]The sensitivity to the misspecification is reported in many simulation studies, including Boswijk and Franses (1992), Cheung and Lai (1993), Bewley and Yang (1995b), Ho and Sorensen (1996), and Haug (1996).

[2]VAR-based procedures other than Johansen's (1991) LR test include Stock and Watson's (1988) common trend test, a canonical correlation type test by Bewley and Yang (1995a), the FM-VAR rank test by Quintos (1998), and Snell's (1999) test with the replaced null and alternative hypotheses.

[3]This normalization-free property follows from the fact that it uses an idea related to the residual from the orthogonal regression, which does not distinguish dependent variables from independent variables. See Phillips and Ouliaris (1990), p

We consider the possibility of using the variance matrix for determining the cointegrating rank, which was not explicitly discussed in their paper.

The test statistic is simply based on a standardized version of the inverse of the sample variance matrix, and therefore it can easily be calculated. Unlike the VAR-based tests, it is a fully nonparametric cointegration test for the following two reasons.[4] First, it utilizes the degeneracy in the sample variance matrix rather than the degeneracy in the first order VAR coefficient estimates. Second, for the purpose of eliminating the effect of dependency in the limiting distribution, the sample moment is standardized using the nonparametric long-run variance estimation rather than lags in VAR models. Since the long-run variance is defined by the spectral density at frequency zero (multiplied by a constant), this second aspect is analogous to the choice between the nonparametric spectrum estimation and the autoregressive spectrum estimation without kernel smoothing.[5] Using the properties of nonparametric spectrum estimation for both differenced [or $I(0)$] and overdifferenced [or $I(-1)$] variables available in the literature (Hannan, 1970, and Phillips, 1995), we show that the test statistic based on the eigenvalues follows a familiar matrix unit root distribution in the limit under the null hypothesis of a specified cointegrating rank.[6]

Our test can also be interpreted as a multivariate version of the nonparametric variance ratio test for a unit root. This view is in contrast to the fact that the VAR-based test of Johansen (1991) and Saikkonen and Luukkonen (1997) can be interpreted as a multivariate augmented Dickey-Fuller (ADF) test.[7] The von Neumann (VN) ratio, the ratio of the sample variances of the first differences and the levels of a time

172.

[4] Recently, another nonparametric cointegrating rank test was proposed by Bierens (1997). The advantage of our procedure is the simplicity in both the test statistic and its limit distribution.

[5] See Robinson and Velasco (1997) and Parzen (1983) for the general discussions of the nonparametric spectrum estimation method and the autoregressive spectrum estimation method, respectively.

[6] The test statistics by Stock and Watson (1988), Harris (1997) and Snell (1999) require estimation of the $I(0)/I(1)$ subsystems. For this purpose, they utilized principal components analysis which requires the eigenvalues of sample variance. In contrast, our approach employs the eigenvalues of a standardized version of the inverse of the sample variance and uses them directly as a test statistic.

[7] In this context, the semiparametric version of Stock and Watson (1988) test is a multivariate Z test of Phillips (1987).

series, was first applied to test the random walk hypothesis by Sargan and Bhargava (1983). In a later work, Nabeya and Tanaka (1990) pointed out that the iid error assumption can be relaxed to allow for serial dependence without altering the limit distribution of the VN ratio if a nonparametric correction term is introduced. Alternatively, the modified VN ratio can be also constructed by simply replacing the numerator with a sample long-run variance of the differenced [or $I(0)$] variables, and the ratio obtained corresponds to our proposed test statistic in the scalar case. Furthermore, given the fact that the sample variance and the sample long-run variance of an integrated variable have the same limiting distribution with different rates of convergence (Phillips, 1991, and Corbae, Ouliaris and Phillips, 1999), we can also construct a variant of the VN ratio by replacing the denominator with the sample long-run variance of the level [or $I(1)$] variables. This idea is the basis of our second test statistic for cointegrating rank.

The remainder of the paper is organized as follows: Section 2 explains the model and assumptions. In Section 3, two types of cointegrating rank tests are introduced and their asymptotic properties are examined. Section 4 reports Monte Carlo results regarding the finite sample behavior of the test. The comparisons with parametric and semiparametric procedures are also provided. Section 5 reports the results of the test of the stochastic growth model as an empirical application of our procedure. Some concluding remarks are made in Section 6. All the proofs are presented in the Appendix.

Throughout this paper, we use the symbols "$\xrightarrow{d}$" and "$\xrightarrow{p}$" to signify convergence in distribution and convergence in probability, respectively. The inequality "> 0" denotes positive definite when applied to matrices. All the limits in the paper are taken as the sample size $T \to \infty$.

2 Model and assumptions

Let $\{z_t\}_{t=1}^{\infty}$ be an n-vector process generated by

$$z_t = z_{t-1} + \xi_t \tag{1}$$

where $\{\xi_t\}_{t=1}^{\infty}$ is a n-vector stationary innovation sequence and z_t is initialized at $t = 0$ with $z_0 = 0$. Cointegration can be understood as a singularity of the long-run variance matrix of ξ_t defined by $\Omega_{\xi\xi} \equiv \sum_{k=-\infty}^{\infty} E\left(\xi_k \xi_0'\right)$. In other words, z_t is cointegrated if there exist cointegrating vectors which lie in the null space of $\Omega_{\xi\xi}$.

For convenience, we assume ξ_t to be a linear process and express cointegration as a singularity restriction on the long-run moving average coefficient of the linear process.

Assumption EC (Error Condition):

(a) $\xi_t = C(L)\varepsilon_t = \sum_{j=0}^{\infty} C_j \varepsilon_{t-j}$, $C_0 = I_n$.

(b) ε_t *is iid with zero mean, variance matrix* $\Sigma_{\varepsilon\varepsilon} > 0$*, finite fourth order cumulants, and* $\varepsilon_s = 0$ *for* $s \leq 0$.

(c) $\sum_{j=0}^{\infty} j^2 \|C_j\| < B < \infty$ *where* $\|C_j\| \equiv tr(C_j' C_j)^{1/2}$.

(d) $H = [H_1 : H_2]$ *is an orthogonal matrix where* H_1 *is a* $n \times r$ *matrix and* H_2 *is a* $n \times (n-r)$ *matrix such that* $H_1' C(1) = 0$*, and rank*$(H_2' C(1)) = n - r$, $0 \leq r \leq n$. *(If* $r = n$*, we take* $H = H_1 = I_n$*, if* $r = 0$*, we take* $H = H_2 = I_n$*.)*

(e)$[C^*(1)' H_1 : C(1)' H_2]$ *is nonsingular, where* $C^*(L) = \sum_{j=0}^{\infty} C_j^* L^j$ *and* $C_j^* = -\sum_{i=j+1}^{\infty} C_i$*. (This also implies the nonsingularity of* $C^*(1)$ *for* $r = n$*, and the nonsingularity of* $C(1)$ *for* $r = 0$*.)*

EC (d) implies that z_t has r cointegrating vectors (the columns of H_1) and $n - r$ unit roots. Using H, we can rotate z_t into $I(0)/I(1)$ subsystems as

$$\underline{z}_t \equiv H' z_t = \begin{bmatrix} H_1' z_t \\ H_2' z_t \end{bmatrix} = \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} I(0) \\ I(1) \end{bmatrix} \begin{matrix} r \\ n-r \end{matrix} . \tag{2}$$

Since EC (c) ensures the validity of decomposing z_t into the stochastic trend and stationary components using $C(L) = C(1) + (1-L)C^*(L)$ as in Phillips and Solo (1992, p.985), the transformed system can be rewritten as

$$z_{1t} = H_1' C^*(L)\varepsilon_t = u_{1t} \tag{3}$$

$$\Delta z_{2t} = H_2' C(L)\varepsilon_t = u_{2t} \tag{4}$$

with variance and long-run variance matrices of $u_t = (u_{1t}', u_{2t}')$ written as

$$\Sigma_{uu} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} > 0, \quad \Omega_{uu} = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} > 0, \tag{5}$$

respectively. Note that the positive definiteness of the long-run variance matrix follows from EC (e).

Under Assumption EC, the multivariate version of the limit theory for linear processes developed by Phillips and Solo (1992) can be used in our analysis, including the functional central limit theory. For example, for $s \in [0,1]$, $T^{-1/2}\sum_{t=1}^{[Ts]} u_{2t} \xrightarrow{d} B_2(s) \equiv \Omega_{22}^{1/2} W_{n-r}(s)$ where $[Ts]$ signifies the integer part of Ts and $W_{n-r}(s)$ denote an $(n-r)$-vector standard Brownian motion. (When there is no ambiguity, dimensional subscripts will be omitted to simplify the notation. We also write integrals with respect to Lebesgue measure such as $\int_0^1 W(s)ds$ by $\int_0^1 W(s)$ or $\int_0^1 W$.)

As pointed out by Engle and Granger (1987, p. 260), when both $I(0)$ and $I(1)$ elements appear in the system, the sample variance matrix of z_t defined by $M_{zz} \equiv T^{-1}\sum_{t=1}^{T} z_t z_t'$ with the additional normalizer $1/T$ degenerates in the limit. This can be seen using the sample variance matrix of the transformed data $\underline{z}_t$ since

$$T^{-1}M_{\underline{z}\underline{z}} = T^{-2}\sum_{t=1}^{T}\begin{bmatrix} z_{1t}z_{1t}' & z_{1t}z_{2t}' \\ z_{2t}z_{1t}' & z_{2t}z_{2t}' \end{bmatrix} \xrightarrow{d} \begin{bmatrix} 0 & 0 \\ 0 & \int_0^1 B_2 B_2' \end{bmatrix}. \tag{6}$$

Similarly, the singularity in the long-run variance matrix of Δz_t (namely, $\Omega_{\Delta z\Delta z} = \Omega_{\xi\xi}$) and in its consistent estimator ($\widehat{\Omega}_{\Delta z\Delta z}$) can be understood using those of transformed data $\Delta\underline{z}_t$ ($\Omega_{\Delta\underline{z}\Delta\underline{z}}$ and $\widehat{\Omega}_{\Delta\underline{z}\Delta\underline{z}}$) since

$$\widehat{\Omega}_{\Delta\underline{z}\Delta\underline{z}} \xrightarrow{p} \Omega_{\Delta\underline{z}\Delta\underline{z}} = \begin{bmatrix} \Omega_{\Delta u_1\Delta u_1} & \Omega_{u_2\Delta u_1} \\ \Omega_{\Delta u_1 u_2} & \Omega_{u_2u_2} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & \Omega_{22} \end{bmatrix} \tag{7}$$

where three zero submatrices follows from Δu_{1t} being an $I(-1)$ process and $\Omega_{u_2u_2} = \Omega_{22}$ follows from the definition in (5).

In this paper, we employ the kernel estimator of the long-run variance matrix of Δz_t given by

$$\widehat{\Omega}_{\Delta z\Delta z} = \sum_{j=-T+1}^{T-1} w(j/K)\widehat{\Gamma}_{\Delta z\Delta z}(j), \quad \widehat{\Gamma}_{\Delta z\Delta z}(j) = T^{-1}\Sigma'\Delta z_{t+j}\Delta z_t' \tag{8}$$

where $w(\cdot)$ is a kernel function, K is a bandwidth parameter and Σ' signifies summation over $1 \leq t$, $t + j \leq T$. The class of kernels in our analysis is similar to the one employed by Phillips (1995), and is assumed to satisfy the following condition:

Assumption KL (Kernel Condition): *The kernel function $w(\cdot) : R \to [-1, 1]$ is a twice continuously differentiable even function with:*

(a) $w(0) = 1$, $w'(0) = 0$, $w''(0) < 0$*; and either*

(b) $w(x) = 0$, $|x| \geq 1$, *with* $\lim_{|x|\to 1} w(x)/(1 - |x|)^2 =$*constant, or* (b') $w(x) = O(x^{-2}),\ as\ |x| \to 1.$

In the subsequent section, we examine the asymptotic properties of two simple nonparametric tests for cointegrating rank based on $T^{-1}M_{zz}$ and $\widehat{\Omega}_{\Delta z \Delta z}$.

3 Test statistics

We consider testing the following hypotheses about the cointegrating rank of the n variable system,

$$\mathcal{H}_0 : r = s, \qquad \mathcal{H}_1 : r > s$$

where $0 \leq s < n$. Let $\lambda_i(A)$ be the i-th largest eigenvalue of a matrix A. Our first test statistic is

$$P(n, s) = T \sum_{i=s+1}^{n} \widehat{\lambda}_i \tag{9}$$

where $\widehat{\lambda}_i = \lambda_i\left(\widehat{\Omega}_{\Delta z \Delta z} M_{zz}^{-1}\right)$. Since $\widehat{\lambda}_i$'s are the eigenvalues in descending order, we expect that $P(n, s)$ consists of $n - r$ zero eigenvalues of $\widehat{\Omega}_{\Delta z \Delta z} M_{zz}^{-1}$ in the limit under the null hypothesis.

For the no cointegration hypothesis ($r = 0$), $P(n, s)$ can also be written as $T\mathrm{tr}\left(\widehat{\Omega}_{\Delta z \Delta z} M_{zz}^{-1}\right)$. The important difference between the present test statistic and the original Phillips-Ouliaris $\widehat{P}_z$ test statistic is in the selection of standardizing matrix. Instead of $\widehat{\Omega}_{\Delta z \Delta z}$ calculated from the first differences, the original $\widehat{P}_z$ test employs $\widehat{\Omega}_{\widehat{\xi}\widehat{\xi}}$ based on the residuals from the first order vector autoregression $z_t = \widehat{\Phi} z_{t-1} + \widehat{\xi}_t$. Such a residual-based long-run variance matrix estimator converges to a nonsingular matrix even in the case of cointegration and thus is a convenient approach in establishing the consistency of the test for no cointegration hypothesis. On the other hand, we allow the degeneracy in the limit of $\widehat{\Omega}_{\Delta z \Delta z}$ as well as the degeneracy in the limit of $T^{-1}M_{zz}$.[8]

Let us next consider a sample long-run variance matrix of the level series z_t , or $\widehat{\Omega}_{zz}$, obtained from (8) with Δz_t and K replaced by z_t and M. In the same manner as for the normalized sample moment $T^{-1}M_{zz}$ [see (6) above], we can show that the normalized sample long-run variance matrix $(MT)^{-1}\widehat{\Omega}_{zz}$ converges to a (random) singular matrix. This similarity leads to the possibility of constructing another test statistic using $\widehat{\Omega}_{zz}$ in place of M_{zz} in (9). Our second test statistics for cointegrating rank is defined as

$$P^*(n,s) = MT\sum_{i=s+1}^{n}\lambda_i^* \tag{10}$$

where $\lambda_i^* = \lambda_i\left(\overline{w}\widehat{\Omega}_{\Delta z\Delta z}\widehat{\Omega}_{zz}^{-1}\right)$ and $\overline{w} = \int_{-\infty}^{\infty} w(s)ds$.[9] If M is set to unity with a kernel satisfying $\overline{w} = 1$, then $P^*(n,s)$ reduces to $P(n,s)$. This second test may therefore be viewed as a generalization of the first test.

The main result of the paper is stated in the following theorem:

Theorem 3.1: *Suppose that* $\{z_t\}_1^\infty$ *is generated by* (1) *and assumptions EC and KL are satisfied. In addition, let* $P(n,s)$ *use bandwidth* $K = k_0T^k$ *with* $k_0 > 0$ *and* $k \in (0,1/2)$. *Also let* $P^*(n,s)$ *use bandwidths* $K = k_0T^k$ *and* $M = m_0T^m$ *with* $k_0, m_0 > 0$, $k \in (0,1)$ *and* m *satisfying*

$$\begin{cases} 0 \le m < 2k & \text{for } k \in (0,1/3] \\ 0 \le m < (k+1)/2 & \text{for } k \in (1/3,1/2] \\ 2k-1 < m < (k+1)/2 & \text{for } k \in (1/2,1). \end{cases}$$

[8]We do not use the residual-based estimator $\widehat{\Omega}_{\hat{\xi}\hat{\xi}}$ in our test statistics because such a test does not satisfy the assumption required in Lemma A.1 in the Appendix.

[9]For example, $\overline{w}$ for Parzen, Tukey-Hanning and QS kernels are 3/4, 1, and 5/4, respectively.

(a) *Under* $\mathcal{H}_0$,

$$P(n,s), P^*(n,s) \xrightarrow{d} \text{tr}\left\{ \left(\int_0^1 W_{n-r} W_{n-r}' \right)^{-1} \right\}.$$

(b) *Under* $\mathcal{H}_1$,

$$(K^2/T)P(n,s) \xrightarrow{p} -w''(0) \sum_{i=s+1}^{r} \lambda_i \left(\Omega_{11} \Sigma_{11}^{-1} \right) \quad > \quad 0, \qquad \textit{and}$$
$$(K^2/MT)P^*(n,s) \xrightarrow{p} -\overline{w}w''(0)(r-s) \quad > \quad 0.$$

Remarks: (a) Part (a) of Theorem 3.1 shows that the two proposed test statistics have the same limit distribution under the null hypothesis, while part (b) shows that both tests are consistent and test statistics diverge at the rate of $O_p(T^{1-2k})$ for $P(n,s)$ and $O_p(T^{1+m-2k})$ for $P^*(n,s)$ under the alternative. It should be noted that $P(n,s)$ requires the expansion rate of the bandwidth K to be slower than $T^{1/2}$ to establish the consistency of the test. The optimal growth rate of the bandwidth [for $I(0)$ variables] given in Andrews (1991) is $k = 1/(2q+1)$ where q is the characteristic exponent of the kernel. Since the characteristic exponent under assumption KL is 2, we can establish the consistency of $P(n,s)$ by using the optimal rate $k = 1/5$. On the other hand, for $P^*(n,s)$, the expansion rate of the bandwidth, k, can be any value between 0 and 1 if an appropriate choice of m is made. From the allowable combination of k and m given in Theorem 3.1, it can easily be seen that we can simply set $m = k$ without imposing any restrictions [with the rate of divergence under the alternative given by $O_p(T^{1-k})$]. This seems to be a convenient choice in practice. Alternatively, we can choose m which gives the maximum rate of divergence under the alternative. For example, a combination of $k = 1/5$ and $m = 2/5 - \epsilon$, with ϵ being an arbitrarily small number, gives $O_p(T^{1-\epsilon})$ as the rate of divergence under the alternative.[10] It should also be noted

[10] This example has been suggested by a referee.

that the asymptotic results of Theorem 3.1 are derived using the class of kernels satisfying Assumption KL which include the Parzen, Tukey-Hanning and quadratic spectral (QS) kernels. In fact, we can relax the assumption and employ other types of kernels, e.g., the Bartlett kernel, while maintaining the consistency of the test. However, in such a case, the asymptotics in Theorem 3.1 should be modified as well as the choice of the expansion rate of the bandwidth discussed above.

(b) Instead of investigating the rate of divergence under the fixed alternative of $\mathcal{H}_1$, we can also consider the local asymptotic power of the tests. For example, suppose we use the local alternative hypothesis of $n-r$ roots close to unity given in Phillips (1988, p.1029), and replace (4) by

$$z_{2t} = \left(I_{n-r} + T^{-1}C\right) z_{2t-1} + u_{2t} \tag{11}$$

where C =diag$(c_1, \ldots, c_{n-r})$. Then, $W(s)$ given in the Part (a) of Theorem 3.1 is replaced by the matrix diffusion process $J_C(s) \equiv \int_0^s \exp\{(s-t)C\}dW(t)$. This outcome implies that both $P(n,s)$ and $P^*(n,s)$ have the same local asymptotic power despite the fact that they behave differently under the fixed alternative. Furthermore, this result can be used in comparison with the asymptotic power function of the VAR-based test obtained by Johansen (1995, ch. 14) and Saikkonen and Lütkepohl (1999). For the scalar case, Elliott, Rothenberg and Stock (1996, figure 1) have shown that both the VN ratio and the Dickey-Fuller type tests for a unit root have power functions which are almost identical to the power envelope under the assumption of no deterministic component.[11] Since the power functions of our tests and the VAR-based tests are simply their multivariate generalizations, both types of tests are expected to have similar asymptotic local powers.

(c) As in other cointegrating rank tests, our test statistics can be extended to allow for a more general

[11] See also Stock (1994) for explicit distribution of these two unit root tests under the local alternative.

model with a deterministic trend. In place of (1), suppose that $\{z_t\}_1^\infty$ is generated by

$$z_t = \Gamma x_t + z_t^s, \qquad z_t^s = z_{t-1}^s + \xi_t \tag{12}$$

where $x_t = (1, t, \ldots, t^p)'$ is a p-th order polynomial deterministic trend. If we employ the traditional method of detrending and replace z_t by the residuals from regression z_t on x_t, then $W(s)$ in part (a) of Theorem 3.1 is replaced by the detrended Brownian motion process $W_X(s) \equiv W(s) - \int_0^1 WX' \left(\int_0^1 XX'\right)^{-1} X(s)$ where $X(s) = (1, s, \ldots, s^p)'$. The percentiles of the limiting distribution for the constant case where $x_t = 1$ and the linear trend case where $x_t = (1, t)'$ as well as the case of no deterministic trend are available in Phillips and Ouliaris (1990, tables 4a to 4c).[12] Since these three cases are used in most applications, their small sample properties will be considered in the next section. Hereafter, we denote the demeaned and the detrended version of the test statistics by subscripts μ and τ, respectively.

An important contribution of Bhargava (1986) is that he performed the detrending after taking the first difference in the VN ratio test for a unit root. Such an approach to detrending turned out to be efficient under the null hypothesis since it eliminates the redundant trend variable. Using a similar detrending method in our cointegrating rank test is not simple, but possible. For example, let us first transform the data to $I(0)/I(1)$ subsystems (3) and (4) based on a consistent estimate of H in an approach analogues to Stock and Watson (1988). Second, we conduct detrending separately using the usual approach for (3) while using the first difference detrending method for (4) and calculate the test statistic using both detrended series together.[13] If we employ this semiparametric variant of our test, then, $W(s)$ in part (a) is now

[12] To be precise, since Phillips and Ouliaris' tables are expressed in terms of the number of independent variables in the cointegrating regression, they cover dimensions from two to six. See the footnote of Table 1 in the present paper for the one-dimensional case.

[13] There are several different ways to conduct the first difference detrending in cointegrated systems. See Lütkepohl and Saikkonen (2000), for example. Alternatively, we can also employ the local GLS detrending which is efficient under the specific local alternative using an analogy to the DF-GLS test proposed by Elliot, Rothenberg and Stock (1996).

replaced by the generalized Brownian bridge process $V_X(s) \equiv W(s) - \int_0^1 dW \widetilde{X}' \left(\int_0^1 \widetilde{X}\widetilde{X}' \right)^{-1} \int_0^s \widetilde{X}(t)$ where $\widetilde{X}(s) = (1, s, \ldots, s^{p-1})'$.[14] Furthermore, we can also investigate the asymptotic properties of various types of detrended tests under the local alternatives. Let us denote the first difference detrended test for the linear trend case with $x_t = (1, t)'$ by subscript $\widetilde{\tau}$. When the null hypothesis is $r = n - 1$, the simulation results of Elliott, Rothenberg and Stock (1996, figures 2 and 3) imply that $P_\mu(n, s)$ and $P_{\widetilde{\tau}}(n, s)$ uniformly have higher local power than those of the VAR-based cointegrating rank tests. We can expect that the same is true for other cases.

4 Experimental evidence

Since the testing procedure developed in the previous section is based on asymptotic theory, it is of interest to examine its performance with sample sizes that are typical for economic time series. For this purpose, this section reports the result of a Monte Carlo experiment designed to assess the small sample properties of the proposed test. In addition, comparisons with VAR-based tests — Johansen's (1991) LR test and the semiparametric version of Stock and Watson's (1988) test (hereafter the SW test)— are provided.

Consider the following bivariate VAR(1) model with iid errors:

$$\begin{aligned} z_{1t} &= \phi_1 z_{1t-1} + \varepsilon_{1t}, \\ z_{2t} &= \phi_2 z_{2t-1} + \varepsilon_{2t} \end{aligned} \tag{13}$$

where $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})' \equiv iidN(0, I_2)$ and $z_{1,0} = z_{2,0} = 0$. The true cointegrating rank of this model is

[14] Such an approach is semiparametric in the sense that it requires the parametric estimation of H. Schmidt and Phillips (1992) have pointed out that the Lagrange multiplier (LM) principle leads to the VN ratio test for a unit root with first difference detrending. Recently, a parametric version of the LM type unit root test has been extended to the cointegrating rank test by Lütkepohl and Saikkonen (2000). In the special case where the null hypothesis is $r = n - 1$, the limit distribution of Lütkepohl and Saikkonen's test is identical to that of our test up to a multiplicative constant. In general, however, the two tests have different asymptotics.

controlled by the number of ones in the pair of the parameters (ϕ_1, ϕ_2).[15] We first investigate the small sample size of the test for the null hypothesis $r = 0$ with $(\phi_1, \phi_2) = (1.0, 1.0)$ and for the null hypothesis $r = 1$ with $(\phi_1, \phi_2) = (1.0, 0.0)$ using the sample variance of Δz_t as a standardized matrix (or $K = 1$). The empirical sizes of $P(n, s)$, $P_\mu(n, s)$ and $P_\tau(n, s)$ with various nominal levels (1, 5 and 10 percents) and various sample sizes ($T = 100$, 200 and 500) are presented in the upper half of Table 1.[16] The results indicate that there are slight size distortions especially with tests for $r = 0$. However, they become less severe by increasing the sample size from 100 to 500.

In practice, the determination of cointegrating rank using the LR test is often conducted by successively testing the hypotheses starting from $r = 0$. For such a sequential procedure with the significance level α, the probability of selecting true rank converges to $1 - \alpha$ while those of selecting the smaller rank converges to zero.[17] This point is discussed in Johansen (1995, ch.12) and the small sample performance of this procedure using the LR test is investigated in Toda (1995). In line with this approach, we next consider the performance of the proposed test in terms of a selected cointegrating rank with a sample size of 200, which is a typical sample size found in macroeconomic data.[18]

Table 2 shows the relative frequencies of selected ranks using $P_\mu(n, s)$ and $P^*_\mu(n, s)$ based on the 5 percent significance level. The Parzen kernel with various choices of bandwidth parameter ($K = 2$, 4 and 8) is employed. For the additional bandwidth parameter required in $P^*_\mu(n, s)$, we simply set $M = K$. Note that for the full unit roots case [$r = 0$ with $(\phi_1, \phi_2) = (1.0, 1.0)$], the theoretical selection frequency for $s = 0$ is 0.95. For the cointegrated case [$r = 1$ with $(\phi_1, \phi_2) = (0.8, 1.0)$, $(0.9, 1.0)$ and $(0.95, 1.0)$], the

[15] Since both our test and the VAR-based test are invariant to the nonsingular transformation of the data, we can restrict our attention to $I(0)/I(1)$ subsystems (2) in the simulation without loss of generality.

[16] All data used in this section were generated using the standard normal (pseudo) random variable generator included in the GAUSS programming language. Each experiment was replicated 10,000 times to obtain the sampling distribution.

[17] This implies that the rank selected by the sequential procedure is not consistent. See Lütkepohl and Poskitt (1998) and Chao and Phillips (1999) for examples of consistent rank selection procedures using various information criteria.

[18] The basic simulation program for the sequential rank selection procedure was kindly provided by Hiro Toda.

selection frequencies for $s = 0$, 1, and 2 are 0.00, 0.95 and 0.05, respectively. Similarly, those frequencies are 0.00, 0.00 and 1.00, respectively, for the stationary case [$r = 2$ with $(\phi_1, \phi_2) = (0.8, 0.8)$, $(0.9, 0.9)$ and $(0.95, 0.95)$]. For the purpose of comparison, the results from the LR test and the SW test based on the VAR(1) model with a constant regressor are also shown in Table 2.[19] Since the VAR model is correctly specified without serial dependence in the error, these parametric tests have a significant advantage over our tests in this simulation design.

There are two notable findings from this experiment. First, increasing bandwidth (in this case to $K = 8$) generally leads to poor performance of $P_\mu(n, s)$. Since the long-run variance is not necessary for the iid case, this result resembles the over-specified lag length case for the VAR-based method. In contrast, the result based on $P_\mu^*(n, s)$ shows that they are relatively robust to variation in the bandwidths. Second, with the exception of $P_\mu(n, s)$ when $K = 8$, our proposed test performs as well as the two other VAR-based tests. In some cases, the performance of our test is better than parametric tests despite the fact that the parametric model is correctly specified.

An important motivation of our test is that it allows for a general class of data generating processes with serial dependence. To see this aspect in small samples, we conduct another simulation with the following experimental design. Consider the same model as before but with MA(1) errors as:

$$\begin{aligned} z_{1t} &= \phi_1 z_{1t-1} + \varepsilon_{1t} + \theta \varepsilon_{1t-1}, \\ z_{2t} &= \phi_2 z_{2t-1} + \varepsilon_{2t} + \theta \varepsilon_{2t-1}. \end{aligned} \tag{14}$$

Similarly to the iid case, we first present the size properties of our test statistics. The lower half of Table 1 reports the empirical sizes of $P(n, s)$, $P_\mu(n, s)$ and $P_\tau(n, s)$ followed by those of $P^*(n, s)$, $P_\mu^*(n, s)$ and

[19] To be more specific, Johansen (1991) proposed two different trace tests depending on the assumption on the constant term; one with no restriction and the other constrained as part of the cointegrating vector. Only tests using the assumption of no restriction are reported here.

$P^*_\tau(n,s)$. The parameters are same as before except for the MA(1) parameter $\theta = 0.1$, and the bandwidth parameters $K(=M) = 2$, 3 and 4 for the sample sizes $T = 100$, 200 and 500, respectively, used for the long-run variance estimation with the Parzen kernel. The result shows that there is more size distortion and a slower convergence of the nominal size to the asymptotic size with increasing sample size than the iid case. Next, we investigate the small sample properties of the sequential procedure with comparisons to VAR-based tests when the sample size is 200. Table 3 shows the relative frequencies of selected ranks using $P_\mu(n,s)$, $P^*_\mu(n,s)$, the LR test and the SW test with three different pairs of AR(1) parameters (ϕ_1, ϕ_2), namely, $(1.0, 1.0)$ for the full unit roots case, $(0.8, 1.0)$ for the cointegrated case and $(0.8, 0.8)$ for the stationary case. The MA(1) parameters, namely θ, are 0.8, 0.5, -0.5 and -0.8 for the first case, and 0.8 and 0.5 for the latter two cases. In addition to the fixed bandwidth for $P_\mu(n,s)$ and $P^*_\mu(n,s)$ [only $K = 4$ case is reported here] and the simple VAR(1) model for the LR test and the SW test considered in the iid case, data-based methods are employed. To be more specific, the automatic bandwidth selection procedure suggested in Andrews (1991) is used for the nonparametric test and the semiparametric correction term for the SW test, while the BIC minimized VAR lag length is used for the LR test.

The result from the MA(1) experiment can be summarized by the following two points. First, for the fixed bandwidth case (with $K = 4$), the two nonparametric tests perform uniformly better than the two other tests based on the simple VAR(1) model for all parameter values. This result is not surprising since VAR-based methods are misspecified in this case and their asymptotic distributions are affected by nuisance parameters caused by serial dependency in the error. Second, for the data-based methods, the performance of VAR-based methods greatly improve for both the parametric case (LR test) and the semiparametric case (SW test). In contrast, introduction of automatic bandwidths results in the less favorable performance of $P_\mu(n,s)$ for the $r = 1$ case. This result is caused by the large value of bandwidths (K being larger than 10) chosen in this case. At the same time, the performance of $P^*_\mu(n,s)$ is relatively unaffected, which implies its robustness to the choice of bandwidth. However, one should be cautious about using automatic bandwidth selection methods in our nonparametric procedure. Since available methods are not designed for the degenerating case ($\widehat{\Omega}_{\Delta z \Delta z}$) nor the integrated case ($\widehat{\Omega}_{zz}$), they may not be a meaningful way of selecting bandwidth from the viewpoint of optimality.

5 Empirical application

Based on the simulation results in the previous section, our new test performs as well as other existing tests with a moderate sample size. In this section, we apply our method to test the implications on cointegrating rank derived from economic theory.

The stochastic growth model has recently become one of the most widely used approaches in macroeconomics literature [see Campbell (1994) and Cooley (1995), for example]. This model frequently employs a permanent productivity shock as the main source of economic growth and fluctuations, and such an assumption implies the cointegrating rank of the system. We re-examine two different types of stochastic growth models considered in King et al. (1991) with our new test statistics. The first system consists of three real variables, namely, output (y_t), consumption (c_t) and investment (i_t).[20] The theory predicts that the system $z_t = (y_t, c_t, i_t)'$ should have two cointegrating vectors. The second system includes three nominal variables in addition to the first three real variables. Additional variables are real money supply ($m_t - p_t$), interest rate (R_t) and inflation (Δp_t). The cointegrating rank of this six-variable model $z_t = (y_t, c_t, i_t, m_t - p_t, R_t, \Delta p_t)'$ should be three according to the model. We use seasonally adjusted, quarterly U.S. data with the sample period from 1947:2 to 1998:1 giving a maximum sample size of 204 observations.[21]

[20] All variables are expressed in logarithms.

[21] The source of the data and variable construction method are almost identical to those used by King et al.(1991, footnote 5).

Table 4 presents the results from $P_\mu(n,s)$, $P^*_\mu(n,s)$, the LR test and the semiparametric version of SW test. First, we directly test the null hypothesis of cointegrating rank predicted by the model. The null of $r=2$ is not rejected by $P_\mu(n,s)$, $P^*_\mu(n,s)$ and the SW test, but is rejected by the LR test in the three-variable system at the 5 percent level of significance. On the other hand, none of the four tests reject the null of $r=3$ in the six-variable system using the same level of significance. Therefore, with the exception of the LR test in the three-variable system, the stochastic growth model is supported by the single hypothesis testing approach.

Next, we consider the sequential procedure to determine the cointegrating rank of the system, which imposes a stronger requirement on the model to be tested. At the 5 percent significance level, the selected cointegrating ranks by $P_\mu(n,s)$, $P^*_\mu(n,s)$, the LR test and the SW test are 0, 1, 3, 1 for the three-variable system and 0, 1, 3, 1 for the six-variable system, respectively. In contrast to the single hypothesis testing approach with its weaker requirement, only one test out of eight yields a result that supports the stochastic growth model. Thus, the empirical result in this section provides a good example of how that our nonparametric cointegrating rank test can provide different conclusions from those obtained with VAR-based methods.

6 Concluding remarks

This paper has proposed a fully nonparametric approach to testing for cointegrating rank which does not require the estimation of the VAR model. The first test utilizes the degeneracy of the standardized sample moment matrix, while the second test uses the degeneracy of the standardized sample long-run variance matrix. Based on the asymptotic analysis, both types of tests were shown to have a nuisance parameter free distribution and to be consistent under reasonable conditions. Monte Carlo experiments have suggested that the performance of our test is also satisfactory in moderate sample sizes. Furthermore, both asymptotic and simulation results have shown some advantages in using the second type of the test.

There are several ways in which this work can be extended. They include: (i) inclusion of the trend break in the trend function and employment of efficient local GLS detrending; (ii) construction of an optimal bandwidth selection procedure which would be suitable for our nonparametric cointegration test; and (iii) conducting a Monte Carlo experiment with a more general data generating process or possibly providing the finite sample comparisons of our sequential procedure to other information criterion-based consistent procedures. These issues remain for future research. Nevertheless, we believe that the proposed nonparametric procedure in this paper can serve as a useful alternative to the other cointegration tests available in this literature.

Acknowledgements

The author would like to thank Peter Phillips and Chris Sims for their helpful suggestions. The author also thanks the editor and two anonymous referees, as well as Markus Goldstein, Oliver Linton, Kosuke Oya, John Rust, Hiro Toda, and seminar participants at Yale University and Osaka University for their helpful comments. The author gratefully acknowledges the Cowles Foundation for financial support under a Carl A. Anderson Fellowship.

Appendix

Before proving Theorem 3.1, we state a lemma which gives the condition on a degenerating matrix for the off diagonal blocks to have an asymptotically negligible effect in the calculation of the eigenvalues.

Lemma A.1: Let

$$X_T = \begin{bmatrix} A_T & T^{-\beta_1} B_T \\ T^{-\beta_2} C_T & T^{-\alpha} D_T \end{bmatrix}, \quad \alpha, \beta_1, \beta_2 > 0$$

be an $n \times n$ random matrix where the dimensions of A_T, B_T, C_T and D_T are $(r \times r), (r \times (n-r)), ((n-r) \times r)$ and $((n-r) \times (n-r))$, respectively, and A_T is a positive definite matrix (D_T can be a singular matrix). Suppose $A_T \xrightarrow{p} A > 0, B_T \xrightarrow{p} B, C_T \xrightarrow{p} C$ and $D_T \xrightarrow{d} D$ where A, B, C are constant matrices and D is a random matrix.

If $\alpha < \beta = \beta_1 + \beta_2$, then

$$\begin{cases} \lambda_i(X_T) \xrightarrow{p} \lambda_i(A) & \text{for } i = 1, \ldots, r, \\ T^{\alpha}\lambda_i(X_T) \xrightarrow{d} \lambda_{i-r}(D) & \text{for } i = r+1, \ldots, n, \end{cases}$$

and

$$\sum_{i=1}^{r} \lambda_i(X_T) \xrightarrow{p} \operatorname{tr}(A), \qquad T^{\alpha} \sum_{i=r+1}^{n} \lambda_i(X_T) \xrightarrow{d} \operatorname{tr}(D).$$

Alternatively, for the normalized version of the same matrix with normalizer T^{α},

$$T^{\alpha}X_T = \begin{bmatrix} T^{\alpha}A_T & T^{\gamma_1}B_T \\ T^{\gamma_2}C_T & D_T \end{bmatrix}$$

where $\gamma_1 = \alpha - \beta_1$, $\gamma_2 = \alpha - \beta_2$, the condition is replaced by $\alpha > \gamma = \gamma_1 + \gamma_2$.

Proof of Lemma A.1: First, suppose $\det(T^{-\alpha}D_T - \lambda I) \neq 0$, then the characteristic equation of X_T in the limit can be written as

$$\begin{aligned} \det(X_T - \lambda I) &= \det \begin{bmatrix} A_T - \lambda I & T^{-\beta_1}B_T \\ T^{-\beta_2}C_T & T^{-\alpha}D_T - \lambda I \end{bmatrix} \\ &= \det(T^{-\alpha}D_T - \lambda I) \times \det\left[(A_T - \lambda I) - T^{-\beta}B_T(T^{-\alpha}D_T - \lambda I)^{-1}C_T\right] \xrightarrow{p} 0. \end{aligned}$$

This convergence requires $\det(A_T - \lambda I) \xrightarrow{p} 0$. Since $A_T \xrightarrow{p} A > 0$, we have $\lambda(X_T) \xrightarrow{p} \lambda(A) > 0$ where $\lambda(\cdot)$ is an eigenvalue of a matrix.

Next, suppose $\det(A_T - \lambda I) \neq 0$, then

$$\det(X_T - \lambda I) = \det(A_T - \lambda I) \times \det\left[(T^{-\alpha}D_T - \lambda I) - T^{-\beta}C_T(A_T - \lambda I)^{-1}B_T\right] \xrightarrow{p} 0.$$

Note that $\alpha < \beta$ implies

$$(T^{-\alpha}D_T - \lambda I) - T^{-\beta}C_T(A_T - \lambda I)^{-1}B_T = (T^{-\alpha}D_T - \lambda I) + o_p(T^{-\alpha}).$$

Therefore the characteristic equation requires $\det(T^{-\alpha}D_T - \lambda I) \xrightarrow{p} 0$ implying $\lambda(X_T) \xrightarrow{p} 0$ and $T^{\alpha}\lambda(X_T) \xrightarrow{d} \lambda(D)$.

Finally, required results follow from the fact that the eigenvalues in the first set are larger than those in the second set in the limit. □

Proof of Theorem 3.1: The basic premise of the proof is to show that the standardized inverse of moment matrix satisfies the condition given in Lemma A.1. In this proof, upper case letters are used to denote data matrices constructed from the observations. For example, $Z_1' = [z_{11}, \ldots, z_{1T}]$.

$P(n, s)$ part: First we rotate the data with H as

$$\begin{aligned} \lambda_i\left(T\widehat{\Omega}_{\Delta z \Delta z}M_{zz}^{-1}\right) &= \lambda_i\left(TH'\widehat{\Omega}_{\Delta z \Delta z}H \times H'M_{zz}^{-1}H\right) = \lambda_i\left(T\widehat{\Omega}_{\Delta \underline{z} \Delta \underline{z}}M_{\underline{zz}}^{-1}\right) \\ &= \lambda_i\left(TD_T^{-1}\widehat{\Omega}_{\Delta \underline{z} \Delta \underline{z}}D_T^{-1} \times D_T M_{\underline{zz}}^{-1} D_T\right) \end{aligned}$$

where

$$D_T = \text{diag}\left[I_r, T^{1/2}I_{n-r}\right].$$

Using the results from Hannan (1970) and lemma 8.1 (a) and (b) in Phillips (1995), under assumption KL, we have the following expression for the first term,

$$\begin{aligned}
TD_T^{-1}\widehat{\Omega}_{\Delta\underline{z}\Delta\underline{z}}D_T^{-1} &= \begin{bmatrix} T\widehat{\Omega}_{\Delta u_1\Delta u_1} & T^{1/2}\widehat{\Omega}_{u_2\Delta u_1} \\ T^{1/2}\widehat{\Omega}_{\Delta u_1 u_2} & \widehat{\Omega}_{u_2 u_2} \end{bmatrix} \\
&= \begin{cases} \begin{bmatrix} -(T/K^2)w''(0)\Omega_{11} + o_p(T/K^2) & O_p(T^{1/2}/K^2) \\ O_p(T^{1/2}/K^2) & \Omega_{22} + o_p(1) \end{bmatrix}, \text{ for } k \in (0, 1/3] \\ \begin{bmatrix} -(T/K^2)w''(0)\Omega_{11} + o_p(T/K^2) & O_p(1/\sqrt{K}) \\ O_p(1/\sqrt{K}) & \Omega_{22} + o_p(1) \end{bmatrix}, \text{ for } k \in (1/3, 1/2). \end{cases}
\end{aligned}$$

For the second term,

$$\begin{aligned}
D_T M_{\underline{zz}}^{-1} D_T &= TD_T \begin{bmatrix} Z_1'Z_1 & Z_1'Z_2 \\ Z_2'Z_1 & Z_2'Z_2 \end{bmatrix}^{-1} D_T \\
&= \begin{bmatrix} T(Z_1'Q_2Z_1)^{-1} & -T^{3/2}(Z_1'Z_1)^{-1}Z_1'Z_2(Z_2'Q_1Z_2)^{-1} \\ -T^{3/2}(Z_2'Q_1Z_2)^{-1}Z_2'Z_1(Z_1'Z_1)^{-1} & T^2(Z_2'Q_1Z_2)^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \Sigma_{11}^{-1} + o_p(1) & O_p(T^{-1/2}) \\ O_p(T^{-1/2}) & T^2(Z_2'Z_2)^{-1} + o_p(1) \end{bmatrix}
\end{aligned}$$

where $Q_i = I - Z_i(Z_i'Z_i)^{-1}Z_i'$ for $i = 1, 2$. Combining the results for the first term and the second term yields

$$\lambda_i\left(T\widehat{\Omega}_{\Delta z\Delta z}M_{zz}^{-1}\right) = \lambda_i\left(\begin{bmatrix} -(T/K^2)w''(0)\Omega_{11}\Sigma_{11}^{-1} + o_p(T/K^2) & O_p(T^{1/2}/K^2) + O_p(1/\sqrt{K}) \\ O_p(T^{1/2}/K^2) + O_p(1/\sqrt{K}) & \Omega_{22}T^2(Z_2'Z_2)^{-1} + o_p(1) \end{bmatrix}\right)$$

where order of off diagonal blocks depend on k. Without loss of generality, suppose $k_0 = 1$. The rate of expansion for the (1,1) block can then be written as $T/K^2 = T^{1-2k} = T^\alpha$ with $\alpha = 1 - 2k > 0$. Similarly, let T^{γ_1} and T^{γ_2} be the rates of expansion for the (1,2) and (2,1) blocks, respectively. Since

$$\gamma_1 = \gamma_2 = \begin{cases} 1/2 - 2k, \text{ for } k \in (0, 1/3] \\ -k/2\ , \text{ for } k \in (1/3, 1/2), \end{cases}$$

we have

$$\begin{aligned}
\alpha - \gamma &= \alpha - (\gamma_1 + \gamma_2) \\
&= \begin{cases} 2k > 0, \text{ for } k \in (0, 1/3] \\ 1 - k > 0, \text{ for } k \in (1/3, 1/2). \end{cases}
\end{aligned}$$

Therefore, in either case, the condition required in Lemma A.1 is satisfied. From Lemma A.1 and

$T^2(Z_2'Z_2)^{-1} \xrightarrow{d} \left(\int_0^1 B_2 B_2'\right)^{-1}$, we have

$$(K^2/T)\lambda_i\left(T\widehat{\Omega}_{\Delta z \Delta z}M_{zz}^{-1}\right) \xrightarrow{p} -w''(0)\lambda_i\left(\Omega_{11}\Sigma_{11}^{-1}\right)$$

for $i = 1, \ldots, r$, and

$$\lambda_i\left(T\widehat{\Omega}_{\Delta z \Delta z}M_{zz}^{-1}\right) \xrightarrow{d} \lambda_{i-r}\left(\Omega_{22}\left(\int_0^1 B_2 B_2'\right)^{-1}\right)$$

for $i = r+1, \ldots, n$. Under $\mathcal{H}_0$,

$$\begin{aligned} P(n,s) &= \sum_{i=s+1}^{n}\lambda_i\left(T\widehat{\Omega}_{\Delta z \Delta z}M_{zz}^{-1}\right) = \sum_{i=r+1}^{n}\lambda_i\left(T\widehat{\Omega}_{\Delta z \Delta z}M_{zz}^{-1}\right) \\ &\xrightarrow{d} \sum_{i=1}^{n-r}\lambda_i\left(\Omega_{22}\left(\int_0^1 B_2 B_2'\right)^{-1}\right) = \mathrm{tr}\left\{\Omega_{22}\left(\int_0^1 B_2 B_2'\right)^{-1}\right\} = \mathrm{tr}\left\{\left(\int_0^1 W_{n-r}W_{n-r}'\right)^{-1}\right\} \end{aligned}$$

as required in part (a). Under $\mathcal{H}_1$,

$$\begin{aligned} (K^2/T)P(n,s) &= (K^2/T)\sum_{i=s+1}^{n}\lambda_i\left(T\widehat{\Omega}_{\Delta z \Delta z}M_{zz}^{-1}\right) \\ &= (K^2/T)\sum_{i=s+1}^{r}\lambda_i\left(T\widehat{\Omega}_{\Delta z \Delta z}M_{zz}^{-1}\right) + (K^2/T)\sum_{i=r+1}^{n}\lambda_i\left(T\widehat{\Omega}_{\Delta z \Delta z}M_{zz}^{-1}\right) \\ &= -w''(0)\sum_{i=s+1}^{r}\lambda_i\left(\Omega_{11}\Sigma_{11}^{-1}\right) + o_p(1) \end{aligned}$$

as required in part (b).

$P^*(n,s)$ part: We first rotate the data as

$$\begin{aligned} \lambda_i\left(MT\widehat{\Omega}_{\Delta z \Delta z}\widehat{\Omega}_{zz}^{-1}\right) &= \lambda_i\left(MTH'\widehat{\Omega}_{\Delta z \Delta z}H \times H'\widehat{\Omega}_{zz}^{-1}H\right) = \lambda_i\left(MT\widehat{\Omega}_{\Delta \underline{z} \Delta \underline{z}}^{-1}\widehat{\Omega}_{\underline{zz}}^{-1}\right) \\ &= \lambda_i\left(MTD_{MT}^{-1}\widehat{\Omega}_{\Delta \underline{z} \Delta \underline{z}}D_{MT}^{-1} \times D_{MT}\widehat{\Omega}_{\underline{zz}}^{-1}D_{MT}\right) \end{aligned}$$

where

$$D_{MT} = \mathrm{diag}\left[I_r, (MT)^{1/2}I_{n-r}\right].$$

By the same line of argument used in the proof of $P(n,s)$ part, under assumption KL, we have the following expression for the first term,

$$MTD_{MT}^{-1}\widehat{\Omega}_{\Delta \underline{z} \Delta \underline{z}}D_{MT}^{-1} = \begin{bmatrix} MT\widehat{\Omega}_{\Delta u_1 \Delta u_1} & (MT)^{1/2}\widehat{\Omega}_{u_2 \Delta u_1} \\ (MT)^{1/2}\widehat{\Omega}_{\Delta u_1 u_2} & \widehat{\Omega}_{u_2 u_2} \end{bmatrix}$$

$$= \begin{cases} \begin{bmatrix} -(MT/K^2)w''(0)\Omega_{11} + o_p(MT/K^2) & O_p((MT)^{1/2}/K^2) \\ O_p((MT)^{1/2}/K^2) & \Omega_{22} + o_p(1) \end{bmatrix}, \text{ for } k \in (0, 1/3] \\ \begin{bmatrix} -(MT/K^2)w''(0)\Omega_{11} + o_p(MT/K^2) & O_p(\sqrt{M/K}) \\ O_p(\sqrt{M/K}) & \Omega_{22} + o_p(1) \end{bmatrix}, \text{ for } k \in (1/3, 1). \end{cases}$$

Next, from the argument on the asymptotic properties of long-run variance estimator for $I(1)$ series used in Phillips (1991) and Corbae, Ouliaris and Phillips (1999), we can show that

$$\frac{1}{MT}\widehat{\Omega}_{z_2z_2} \xrightarrow{d} \left(\int_{-\infty}^{\infty} w(s)ds\right)\int_0^1 B_2B_2' \qquad \text{and} \qquad \frac{1}{M}\widehat{\Omega}_{z_2z_1} = \frac{1}{M}\widehat{\Omega}_{z_2u_1} = O_p(1).$$

Using these results, we have the following asymptotics for the second term,

$$D_{MT}\widehat{\Omega}_{\underline{zz}}^{-1}D_{MT} = D_{MT}\begin{bmatrix} \widehat{\Omega}_{z_1z_1} & \widehat{\Omega}_{z_1z_2} \\ \widehat{\Omega}_{z_2z_1} & \widehat{\Omega}_{z_2z_2} \end{bmatrix}^{-1} D_{MT}$$

$$= \begin{bmatrix} (\widehat{\Omega}_{z_1z_1} - \widehat{\Omega}_{z_1z_2}\widehat{\Omega}_{z_2z_2}^{-1}\widehat{\Omega}_{z_2z_1})^{-1} & -(MT)^{1/2}\widehat{\Omega}_{z_1z_1}^{-1}\widehat{\Omega}_{z_1z_2}(\widehat{\Omega}_{z_2z_2} - \widehat{\Omega}_{z_2z_1}\widehat{\Omega}_{z_1z_1}^{-1}\widehat{\Omega}_{z_1z_2})^{-1} \\ -(MT)^{1/2}(\widehat{\Omega}_{z_2z_2} - \widehat{\Omega}_{z_2z_1}\widehat{\Omega}_{z_1z_1}^{-1}\widehat{\Omega}_{z_1z_2})^{-1}\widehat{\Omega}_{z_2z_1}\widehat{\Omega}_{z_1z_1}^{-1} & MT(\widehat{\Omega}_{z_2z_2} - \widehat{\Omega}_{z_2z_1}\widehat{\Omega}_{z_1z_1}^{-1}\widehat{\Omega}_{z_1z_2})^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} \Omega_{11}^{-1} + o_p(1) & O_p(\sqrt{M/T}) \\ O_p(\sqrt{M/T}) & MT\widehat{\Omega}_{z_2z_2}^{-1} + o_p(1) \end{bmatrix}.$$

Combining the results for the first term and the second term yields

$$\lambda_i\left(MT\widehat{\Omega}_{\Delta z\Delta z}\widehat{\Omega}_{zz}^{-1}\right) = \lambda_i\left(\begin{bmatrix} -(MT/K^2)w''(0)I_r + o_p(MT/K^2) & E \\ F & \Omega_{22}MT\widehat{\Omega}_{z_2z_2}^{-1} + o_p(1) \end{bmatrix}\right)$$

where

$$E = \begin{cases} O_p(M^{3/2}T^{1/2}/K^2), \text{ for } k \in (0, 1/3] \text{ or for } k \in (1/3, 1) \text{ with } m \geq (3k-1)/2 \\ O_p(\sqrt{M/K}), \text{ for } k \in (1/3, 1) \text{ with } m < (3k-1)/2 \end{cases}$$

and

$$F = \begin{cases} O_p((MT)^{1/2}/K^2), \text{ for } k \in (0, 1/3] \\ O_p(\sqrt{M/K}), \text{ for } k \in (1/3, 1). \end{cases}$$

Suppose, without loss of generality, $k_0 = m_0 = 1$. By the same argument used in the proof for $P(n, s)$, with respect to expansion rate for the (1,1) block, we have

$$\alpha = 1 + m - 2k > 0$$

if $m > 2k - 1$. For the off diagonal blocks, we have

$$\begin{aligned} \gamma_1 &= \frac{1+3m}{2} - 2k, \quad \gamma_2 = \frac{1+m}{2} - 2k \text{ for } k \in (0, 1/3], \\ \gamma_1 &= \frac{1+3m}{2} - 2k, \quad \gamma_2 = \frac{m-k}{2} \text{ for } k \in (1/3, 1) \text{ with } m \geq (3k-1)/2, \text{ and} \end{aligned}$$

$$\gamma_1 = \gamma_2 = \frac{m-k}{2} \text{ for } k \in (1/3, 1) \text{ with } m < (3k-1)/2.$$

When $k \in (0, 1/3]$,

$$\alpha - \gamma = (1 + m - 2k) - \left\{ \left(\frac{1+3m}{2} - 2k \right) + \left(\frac{1+m}{2} - 2k \right) \right\} = 2k - m > 0$$

if $m < 2k$. Similarly, when $k \in (1/3, 1)$ with $m \geq (3k-1)/2$,

$$\alpha - \gamma = (1 + m - 2k) - \left\{ \left(\frac{1+3m}{2} - 2k \right) + \frac{m-k}{2} \right\} = \frac{1}{2} - m + \frac{k}{2} > 0$$

if $m < (k+1)/2$. Finally, when $k \in (1/3, 1)$ with $m < (3k-1)/2$,

$$\alpha - \gamma = (1 + m - 2k) - 2 \times (\frac{m-k}{2}) = 1 - k > 0.$$

Therefore, for all cases, the condition required in Lemma A.1 is satisfied under the assumption on the growth rate of bandwidth given in Theorem 3.1. From Lemma A.1, we have

$$(K^2/MT)\lambda_i \left(MT\widehat{\Omega}_{\Delta z \Delta z} \widehat{\Omega}_{zz}^{-1} \right) \xrightarrow{p} -w''(0)\lambda_i (I_r) = -w''(0)$$

for $i = 1, \ldots, r$, and

$$\lambda_i \left(MT\widehat{\Omega}_{\Delta z \Delta z} \widehat{\Omega}_{zz}^{-1} \right) \xrightarrow{d} \lambda_{i-r} \left(\Omega_{22} \overline{w}^{-1} \left(\int_0^1 B_2 B_2' \right)^{-1} \right)$$

for $i = r+1, \ldots, n$ where $\overline{w} = \int_{-\infty}^{\infty} w(s)ds$. Under $\mathcal{H}_0$,

$$\begin{aligned} P^*(n,s) &= \sum_{i=s+1}^{n} \lambda_i \left(MT\overline{w}\widehat{\Omega}_{\Delta z \Delta z} \widehat{\Omega}_{zz}^{-1} \right) = \sum_{i=r+1}^{n} \lambda_i \left(MT\overline{w}\widehat{\Omega}_{\Delta z \Delta z} \widehat{\Omega}_{zz}^{-1} \right) \\ &\xrightarrow{d} \sum_{i=1}^{n-r} \lambda_i \left(\Omega_{22} \left(\int_0^1 B_2 B_2' \right)^{-1} \right) = \text{tr} \left\{ \left(\int_0^1 W_{n-r} W_{n-r}' \right)^{-1} \right\} \end{aligned}$$

as required in part (a). Under $\mathcal{H}_1$,

$$\begin{aligned} (K^2/MT)P^*(n,s) &= (K^2/MT) \sum_{i=s+1}^{n} \lambda_i \left(MT\overline{w}\widehat{\Omega}_{\Delta z \Delta z} \widehat{\Omega}_{zz}^{-1} \right) \\ &= (K^2/MT) \sum_{i=s+1}^{r} \lambda_i \left(MT\overline{w}\widehat{\Omega}_{\Delta z \Delta z} \widehat{\Omega}_{zz}^{-1} \right) + (K^2/MT) \sum_{i=r+1}^{n} \lambda_i \left(MT\overline{w}\widehat{\Omega}_{\Delta z \Delta z} \widehat{\Omega}_{zz}^{-1} \right) \\ &= -\overline{w}w''(0)(r-s) + o_p(1) \end{aligned}$$

as required in part (b). □

References

Andrews, D. W. K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica 59(3), 817-858.

Bewley, R. and M. Yang, 1995a, Tests for cointegration based on canonical correlation analysis, Journal of the American Statistical Association 90(431), 990-996.

Bewley, R. and M. Yang, 1995b, Testing for cointegration: the effects of mis-specifying the lag length, Mathematics and Computers in Simulation 39, 251-255.

Bhargava, A., 1986, On the theory of testing for unit roots in observed time series, Review of Economic Studies 53, 369-384.

Bierens, H. J., 1997, Nonparametric cointegration analysis, Journal of Econometrics 77, 379-404.

Boswijk, P. and P. H. Franses, 1992, Dynamic specification and cointegration, Oxford Bulletin of Economics and Statistics 54(3), 369-381.

Campbell, J. Y., 1994, Inspecting the mechanism: An analytical approach to the stochastic growth model, Journal of Monetary Economics 33, 463-506.

Chao, J. C. and P. C. B. Phillips, 1999, Model selection in partially nonstationary vector autoregressive processes with reduced rank structure, Journal of Econometrics 91(2), 227-271.

Cheung, Y.-W. and K. S. Lai, 1993, Finite-sample sizes of Johansen's likelihood ratio tests for cointegration, Oxford Bulletin of Economics and Statistics 55(3), 313-328.

Cooley, T. F., 1995, Frontiers of Business Cycle Research (Princeton University Press, Princeton).

Corbae, D., S. Ouliaris and P. C. B. Phillips, 1999, Band spectral regression with trending data, mimeograph, University of Pittsburgh, National University of Singapore and Yale University.

Elliott, G., T. J. Rothenberg and J. H. Stock, 1996, Efficient tests for an autoregressive unit root, Econometrica 64(4), 813-836.

Engle, R. F. and C. W. J. Granger, 1987, Cointegration and error correction: representation, estimation, and testing, Econometrica 55(2), 251-276.

Hannan, E. J., 1970, Multiple Time Series (Wiley, New York).

Harris, D., 1997, Principal components analysis of cointegrated time series, Econometric Theory 13, 529-557.

Haug, A. A., 1996, Tests for cointegration: A Monte Carlo comparison, Journal of Econometrics 71, 89-115.

Ho, M. S. and B. E. Sorensen, 1996, Finding cointegrating rank in high dimensional systems using the Johansen test: An illustration using data based Monte Carlo simulations, Review of Economics and Statistics, 726-732.

Johansen, S., 1991, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, Econometrica 59(6), 1551-1580.

Johansen, S., 1995, Likelihood-based inference in cointegrated vector autoregressive models (Oxford University Press, Oxford).

King, R. G., C. I. Plosser, J. H. Stock and M. W. Watson, 1991, Stochastic trends and economic fluctuations, American Economic Review 81(4), 819-840.

Lütkepohl, H. and D. S. Poskitt, 1998, Consist estimation of the number of cointegration relations in a vector autoregressive model, in: R. Galata and H. Kuchenhoff, eds., Econometrics in Theory and Practice (Springer Verlag, Berlin) 87-100.

Lütkepohl, H. and P. Saikkonen, 2000, Testing for the cointegrating rank of a VAR process with a time trend, Journal of Econometrics 95(1), 177-198.

Nabeya, S. and K. Tanaka, 1990, Limiting power of unit-root tests in time-series regression, Journal of Econometrics 46, 247-271.

Osterwald-Lenum, M., 1992, A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics, Oxford Bulletin of Economics and Statistics 54(3), 461-472.

Parzen, E., 1983, Autoregressive spectral estimation, in: D. R. Brillinger and P. R. Krishnaiah, eds., Handbook of Statistics, Vol. 3 (North-Holland, Amsterdam) 221-247.

Phillips, P. C. B., 1987, Time series regression with a unit root, Econometrica 55(2), 277-301.

Phillips, P. C. B., 1988, Regression theory for near-integrated time series, Econometrica 56(5), 1021-1043.

Phillips, P. C. B., 1991, Spectral regression for cointegrated time series. in: W. A. Barnett, J. Powell and G. E. Tauchen, eds., Nonparametric and semiparametric methods in econometrics and statistics: Proceedings of the fifth international symposium in economic theory and econometrics (Cambridge University Press, Cambridge) 413-435.

Phillips, P. C. B., 1995, Fully modified least squares and vector autoregression, Econometrica 63(5), 1023-1078.

Phillips, P. C. B. and S. Ouliaris, 1990, Asymptotic properties of residual based tests for cointegration, Econometrica 58(1), 165-193.

Phillips, P. C. B. and V. Solo, 1992, Asymptotics for linear processes, Annals of Statistics 20(2), 971-1001.

Quintos, C. E., 1998, Fully modified vector autoregressive inference in partially nonstationary models, Journal of the American Statistical Association 93(442), 783-795.

Robinson, P. M. and C. Velasco, 1997, Autocorrelation-robust inference, in: G. S. Maddala and C. R. Rao, eds., Handbook of Statistics, Vol. 15 (North-Holland, Amsterdam) 267-298.

Saikkonen, P. and H. Lütkepohl, 1999, Local power of likelihood ratio tests for the cointegrating rank of a VAR process, Econometric Theory 15, 50-78.

Saikkonen, P. and R. Luukkonen, 1997, Testing cointegration in infinite order vector autoregressive processes, Journal of Econometrics 81, 93-126.

Sargan, J. D. and A. Bhargava, 1983, Testing residuals from least squares regression for being generated by the Gaussian random walk, Econometrica 51(1), 153-174.

Schmidt, P. and P. C. B. Phillips, 1992, LM tests for a unit root in the presence of deterministic trends, Oxford Bulletin of Economics and Statistics 54, 257-287.

Snell, A., 1999, Testing for r versus r-1 cointegrating vectors, Journal of Econometrics 88, 151-191.

Stock, J. H., 1994, Unit roots, structural breaks and trends, in: R. F. Engle and D. McFadden eds., Handbook of Econometrics, Vol. 4 (North-Holland, Amsterdam) 2740-2841.

Stock, J. H. and M. W. Watson, 1988, Testing for common trends, Journal of American Statistical Association 83(404), 1097-1107.

Toda, H. Y., 1995, Finite sample performance of likelihood ratio tests for cointegrating ranks in vector autoregressions, Econometric Theory 11, 1015-1032.

Table 1
Empirical Size

(1) iid error

Nominal Size	$T = 100$			$T = 200$			$T = 500$		
	P	P_μ	P_τ	P	P_μ	P_τ	P	P_μ	P_τ
(a) $s = r = 0, \quad (\phi_1, \phi_2) = (1.0, 1.0)$									
10.0	8.2	7.5	7.6	8.9	8.7	8.6	9.8	9.8	9.4
5.0	3.6	3.4	3.3	4.1	4.2	4.0	4.5	4.6	4.6
1.0	0.7	0.6	0.4	0.8	0.8	0.6	1.0	0.9	0.7
(b) $s = r = 1, \quad (\phi_1, \phi_2) = (0.0, 1.0)$									
10.0	9.2	9.5	8.8	9.7	9.8	9.1	9.6	10.5	10.4
5.0	4.3	4.5	4.0	4.8	4.5	4.3	4.9	5.0	5.1
1.0	0.7	0.8	0.8	0.9	0.9	0.9	0.9	1.0	1.0

(2) MA(1) error with $\theta = 0.1$

Nominal Size	$T = 100$			$T = 200$			$T = 500$		
	P	P_μ	P_τ	P	P_μ	P_τ	P	P_μ	P_τ
(a) $s = r = 0, \quad (\phi_1, \phi_2) = (1.0, 1.0)$									
10.0	4.9	3.7	2.7	6.1	5.0	4.0	8.1	7.1	6.1
5.0	1.9	1.5	0.8	2.7	2.2	1.4	3.6	3.1	2.7
1.0	0.3	0.1	0.1	0.4	0.2	0.1	0.6	0.5	0.3
(b) $s = r = 1, \quad (\phi_1, \phi_2) = (0.0, 1.0)$									
10.0	6.5	6.1	4.9	8.1	7.7	6.3	9.2	8.5	7.9
5.0	2.8	2.2	1.9	3.6	3.1	2.4	4.3	4.0	3.3
1.0	0.5	0.3	0.2	0.5	0.4	0.4	0.7	0.6	0.6
Nominal Size	$T = 100$			$T = 200$			$T = 500$		
	P^*	P^*_μ	P^*_τ	P^*	P^*_μ	P^*_τ	P^*	P^*_μ	P^*_τ
(a) $s = r = 0, \quad (\phi_1, \phi_2) = (1.0, 1.0)$									
10.0	7.0	5.3	4.9	9.6	6.9	6.3	12.5	8.4	7.9
5.0	3.2	2.3	1.9	4.5	3.2	2.7	6.0	3.8	3.7
1.0	0.5	0.3	0.2	0.8	0.5	0.4	1.2	0.6	0.5
(b) $s = r = 1, \quad (\phi_1, \phi_2) = (0.0, 1.0)$									
10.0	8.4	7.0	6.7	11.7	8.4	7.8	14.4	9.0	9.0
5.0	3.6	2.8	2.9	5.4	3.6	3.4	7.1	4.4	4.0
1.0	0.7	0.4	0.4	0.9	0.6	0.7	1.5	0.7	0.8

Notes: $n = 2$. P, P_μ, P_τ, P^*, P^*_μ and P^*_τ represent $P(n, s)$, $P_\mu(n, s)$, $P_\tau(n, s)$, $P^*(n, s)$, $P^*_\mu(n, s)$ and $P^*_\tau(n, s)$, respectively. Critical values for part (a) are taken from Tables 4a to 4c in Phillips and Ouliaris (1990). Those for part (b) are obtained by following the procedure described in Appendix B of Phillips and Ouliaris. 10% (5%, 1%) level critical values for $P(n, s)$, $P_\mu(n, s)$ and $P_\tau(n, s)$ are 13.0 (17.6, 29.3), 21.5 (27.5, 40.2) and 35.5 (42.6, 56.9), respectively. Parzen kernel is used for the MA(1) error case with the bandwidth parameters $K (= M) = 2$, 3 and 4 for the sample sizes $T = 100$, 200 and 500, respectively. All numbers are based on 10,000 replications.

Table 2
Determination of Cointegrating Rank with IID Error

$\mathcal{H}_0$	$P_\mu(n,s)$			$P^*_\mu(n,s)$			LR	SW
	$K=2$	$K=4$	$K=8$	$K=2$	$K=4$	$K=8$	$VAR(1)$	$VAR(1)$
				(a) $r=0$				
				$(\phi_1,\phi_2)=(1.0,1.0)$				
$s=0$	**96.3**	**97.2**	**98.8**	**95.3**	**94.9**	**94.0**	**94.7**	**95.4**
$s=1$	3.6	2.7	1.2	4.6	5.0	5.8	4.9	4.3
$s=2$	0.1	0.1	0.0	0.1	0.1	0.1	0.4	0.3
				(b) $r=1$				
				$(\phi_1,\phi_2)=(0.8,1.0)$				
$s=0$	0.3	1.9	22.8	0.1	0.4	1.6	2.9	0.3
$s=1$	**95.5**	**94.5**	**74.8**	**95.4**	**94.8**	**93.6**	**92.0**	**95.3**
$s=2$	4.2	3.6	2.4	4.5	4.7	4.8	5.1	4.5
				$(\phi_1,\phi_2)=(0.9,1.0)$				
$s=0$	36.8	47.4	69.3	31.9	32.9	36.3	55.7	40.8
$s=1$	**59.9**	**50.0**	**29.2**	**64.6**	**63.4**	**60.1**	**40.3**	**56.7**
$s=2$	3.3	2.6	1.5	3.5	3.8	3.6	4.0	2.5
				$(\phi_1,\phi_2)=(0.95,1.0)$				
$s=0$	78.0	82.7	90.9	74.4	73.8	73.4	84.8	82.0
$s=1$	**20.8**	**16.3**	**8.7**	**24.1**	**24.7**	**25.0**	**13.6**	**16.9**
$s=2$	1.2	0.9	0.5	1.5	1.5	1.6	1.6	1.1
				(c) $r=2$				
				$(\phi_1,\phi_2)=(0.8,0.8)$				
$s=0$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$s=1$	0.0	0.0	0.7	0.0	0.0	0.1	0.0	0.0
$s=2$	**100.0**	**100.0**	**99.3**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
				$(\phi_1,\phi_2)=(0.9,0.9)$				
$s=0$	0.3	0.8	5.0	0.1	0.3	0.6	4.9	21.9
$s=1$	14.2	20.9	37.6	12.4	14.6	20.9	26.9	10.5
$s=2$	**85.5**	**78.3**	**57.4**	**87.5**	**85.0**	**78.5**	**68.2**	**67.6**
				$(\phi_1,\phi_2)=(0.95,0.95)$				
$s=0$	35.6	42.0	57.3	32.0	31.8	32.9	62.1	71.8
$s=1$	48.6	45.1	35.2	50.2	50.6	50.3	29.1	20.0
$s=2$	**15.8**	**12.9**	**7.5**	**17.9**	**17.6**	**16.9**	**8.8**	**8.2**

Notes: $n=2, T=200$. Each row shows the relative frequency of selecting the cointegrating rank to be s by sequential procedure with the 5% significance level. Bold numbers indicate the frequencies of selecting the true rank. Critical values for the LR test and the SW test are from Tables 1.1* in Osterwald-Lenum (1992) and from Table 2 in Stock and Watson (1988), respectively. Parzen kernel is used for nonparametric tests [$P_\mu(n,s)$ and $P^*_\mu(n,s)$] with bandwidth $K(=M)$. All numbers are based on 10,000 replications.

Table 3
Determination of Cointegrating Rank with MA(1) Error

		$P_\mu(n,s)$		$P^*_\mu(n,s)$		LR		SW	
θ	$\mathcal{H}_0$	$K=4$	*data*	$K=4$	*data*	$VAR(1)$	*data*	$VAR(1)$	*data*
				(a) $r=0$, $(\phi_1,\phi_2)=(1.0,1.0)$					
0.8	$s=0$	**98.7**	**99.5**	**97.8**	**93.8**	**91.7**	**90.4**	**100.0**	**99.1**
	$s=1$	1.2	0.5	2.1	6.0	7.5	8.7	0.0	0.9
	$s=2$	0.1	0.1	0.1	0.1	0.8	0.9	0.0	0.1
0.5	$s=0$	**98.6**	**99.2**	**97.6**	**94.5**	**93.1**	**89.6**	**99.9**	**98.4**
	$s=1$	1.3	0.8	2.3	5.4	6.2	9.7	0.1	1.5
	$s=2$	0.1	0.1	0.1	0.2	0.7	0.8	0.0	0.1
-0.5	$s=0$	**65.6**	**86.9**	**50.1**	**67.1**	**12.2**	**70.3**	**10.4**	**33.0**
	$s=1$	30.9	12.1	44.1	30.4	58.2	25.8	48.6	51.3
	$s=2$	3.5	1.0	5.8	2.6	29.5	3.9	41.0	15.7
-0.8	$s=0$	**0.2**	**10.4**	**0.0**	**0.8**	**0.0**	**9.1**	**0.0**	**0.0**
	$s=1$	27.1	53.4	19.0	41.4	0.5	39.2	0.2	6.3
	$s=2$	72.8	36.2	81.0	57.8	99.5	51.7	99.8	93.7
				(b) $r=1$, $(\phi_1,\phi_2)=(0.8,1.0)$					
0.8	$s=0$	6.2	62.8	2.7	5.8	70.7	25.4	56.3	23.0
	$s=1$	**91.9**	**35.8**	**94.8**	**89.7**	**27.5**	**68.7**	**43.7**	**75.4**
	$s=2$	1.9	1.4	2.5	4.5	1.8	5.9	0.1	1.6
0.5	$s=0$	5.5	38.3	2.1	3.7	61.0	10.2	38.7	10.7
	$s=1$	**92.5**	**59.8**	**95.2**	**92.1**	**37.0**	**82.4**	**61.1**	**86.9**
	$s=2$	2.0	1.9	2.6	4.2	2.0	7.4	0.2	2.3
				(c) $r=2$, $(\phi_1,\phi_2)=(0.8,0.8)$					
0.8	$s=0$	0.0	0.0	0.0	0.0	4.5	0.7	43.0	6.4
	$s=1$	0.1	3.1	0.0	0.3	18.8	7.1	3.7	0.3
	$s=2$	**99.9**	**96.9**	**100.0**	**99.7**	**76.6**	**92.2**	**53.3**	**93.3**
0.5	$s=0$	0.0	0.0	0.0	0.0	1.2	0.1	22.5	1.4
	$s=1$	0.0	0.7	0.0	0.1	10.5	1.9	1.9	0.1
	$s=2$	**100.0**	**99.3**	**100.0**	**99.9**	**88.3**	**98.1**	**75.6**	**98.5**

Notes: $n=2, T=200$. *data* implies that bandwidths for nonparametric tests [$P_\mu(n,s)$ and $P^*_\mu(n,s)$], the semiparametric SW test, and VAR lag length for the LR test are selected by using a data-based method. All numbers are based on 10,000 replications. See also notes for Table 2.

Table 4
Empirical Results for the Stochastic Growth Models

(a) *Three-Variable Model (n=3, r=2)*

$\mathcal{H}_0$	$P_\mu(n,s)$	$P_\mu^*(n,s)$	*LR*	*SW*
$s=0$	75.88	112.9*	40.27*	-26.34*
$s=1$	30.15	44.21	18.62*	-22.99
$s=2$	0.19	0.20	4.93*	-0.730

(b) *Six-Variable Model (n=6, r=3)*

$\mathcal{H}_0$	$P_\mu(n,s)$	$P_\mu^*(n,s)$	*LR*	*SW*
$s=0$	200.38	290.04*	131.97*	-111.51*
$s=1$	117.27	145.92	82.89*	-20.75
$s=2$	66.29	76.77	47.66*	-18.48
$s=3$	30.96	35.84	23.57	-13.15
$s=4$	6.39	6.74	7.02	-11.15
$s=5$	0.18	0.18	2.38*	-7.89

Notes: Numbers with asterisks imply that the null hypothesis is rejected at the 5% significance level. Critical values for nonparametric tests [$P_\mu(n,s)$ and $P_\mu^*(n,s)$] are taken from Table 4b in Phillips and Ouliaris (1990) and footnotes from Table 1. Those for the LR test and the SW test are taken from Table 1 in Osterwald-Lenum (1992) and Table 2 in Stock and Watson (1988), respectively. Parzen kernel with optimal bandwidth is used for both nonparametric tests and the semiparametric SW test. VAR lag length for the LR test is selected by minimizing BIC.