Discussion Paper #2011-05

Linear Nonstationary Models
- A Review of the Work of Professor P.C.B. Phillips -

Katsuto Tanaka
Department of Economics, Hitotsubashi University
tanaka@stat.hit-u.ac.jp

April, 2011

Abstract

The work of Professor P.C.B. Phillips, even if it is focused on the area of linear nonstationary models, is enormous. So it is hard for me to explore the whole of his work in this paper. Therefore I have decided to take up only a few results of his work. The topics chosen here are applications of the martingale approximation and the problem of choosing between stochastic and deterministic trends, which I discuss and, hopefully extend.

An earlier version of this paper was presented at the special meeting of the New Zealand Econometrics Study Group held at the University of Auckland in March 2008 in honor of Professor P.C.B. Phillips. I am grateful to Professors P.C.B. Phillips, B.E. Hansen and two anonymous referees for their useful comments and suggestions.

1. Introduction

There is no doubt that Professor P.C.B. Phillips is one of the greatest and most influential econometricians in the history of our profession. His research extends over a wide range of the theoretical and empirical problems in econometrics (see, for details, his home page "http://korora.econ.yale.edu/phillips/index.htm"). In particular, his work is enormous on the asymptotic as well as finite sample distribution theory in simultaneous econometric models, the asymptotic inference in integrated and long memory time series models, and a general asymptotic theory for nonstationary panel data.

It is a great honor to overview his work on this occasion, but it is quite difficult for me to give a general and fair view. Among many contributions of Professor Phillips, I have decided to take up the following two topics, the choice of which is, to some extent, in line with my interest, but does reflect the relative importance that his work conveyed to us. In each topic I try to extend and reinterpret his results.

In Section 2 we discuss the first topic, which is the martingale approximation or the so-called B-N decomposition in the econometrics literature. The former idea was first suggested in Gordin (1969) in a wide context, and was extensively discussed in Hall and Heyde (1980), whereas the latter idea was first given in Beveridge and Nelson (1981) in the case of a scalar linear stationary process. Phillips (1988) dealt with a vector case and discussed weak convergence of the sum of matrix products to the Ito integral by using the martingale approximation. The proof of the vector case is quite different from that of the scalar case and is much involved. A simplified proof will be given. This approximation or decomposition is based on expanding a polynomial in a Taylor series about unity. Phillips and Solo (1992) and Phillips (1999), on the other hand, considered the expansion about an arbitrary complex value on the unit circle. It turns out that this latter expansion is suitable for frequency domain applications and is useful for establishing asymptotic results related to Fourier transforms, periodograms and spectra. The expansion will be applied to the analysis of the complex unit root model, extending the original model discussed by Ahtola and Tiao (1987).

In Section 3 we deal with somewhat a philosophical problem, that is, the problem of differentiating between deterministic and stochastic trends. Phillips (1998, 2002) considered a situation where the I(1) process, which is a stochastic trend, is regressed on deterministic trends, and developed an asymptotic theory when K, the number of regressors, becomes large in additon to T, the sample size. This situation is referred to as K-asymptotics. It turns out that the stochastic trend can be well approximated by deterministic trends under K-asymptotics. This poses a question about the validity of unit root tests that aim to differentiate between deterministic and stochastic trends. The discussion will be extended to the near I(1) case, where it is shown that the unit root test becomes invalid under K-asymptotics in the sense that the test gives rise to no power under the local alternative.

In the following we use the symbol "$\Rightarrow$" to signify weak convergence. The limit is taken with respect to the sample size T or the number of deterministic regressors K when T or K goes to ∞.

2. Martingale approximation

2.1. Time domain

Let us consider a q-dimensional I(1) process $\{\boldsymbol{y}_j\}$ defined by

$$\boldsymbol{y}_j = \boldsymbol{y}_{j-1} + \boldsymbol{u}_j, \quad \boldsymbol{y}_0 = \boldsymbol{0}, \quad (j = 1, \ \ldots, \ T), \tag{1}$$

where $\{\boldsymbol{u}_j\}$ is a stationary linear process given by

$$\boldsymbol{u}_j = \sum_{l=0}^{\infty} A_l \boldsymbol{\varepsilon}_{j-l}, \quad \sum_{l=1}^{\infty} l \, ||A_l|| < \infty, \quad A = \sum_{l=0}^{\infty} A_l \neq 0. \tag{2}$$

Here $||B||$ is the matrix norm of B defined by $\max_{a,b}|B_{ab}|$ with B_{ab} being the (a, b)th element of B, whereas $\{\boldsymbol{\varepsilon}_j\}$ follows i.i.d.$(\boldsymbol{0}, I_q)$ with I_q being the identity matrix of dimension q. It holds that the process $\{\boldsymbol{u}_j\}$ is strictly stationary and ergodic with the continuous spectral matrix given by

$$f_{\boldsymbol{u}}(\omega) = \frac{1}{2\pi} \sum_{j=0}^{\infty} A_j e^{ij\omega} \sum_{j=0}^{\infty} A'_j e^{-ij\omega}.$$

Note in passing that A_0 is not assumed to be the identity matrix so that $\mathrm{V}(\boldsymbol{\varepsilon}_j) = I_q$ is justified.

Under the above situation Phillips (1988) proved the following important fact:

$$\frac{1}{T} \sum_{j=1}^{T} \boldsymbol{y}_{j-1} \boldsymbol{u}'_j \ \Rightarrow \ A \int_0^1 \boldsymbol{W}(t) \, d\boldsymbol{W}'(t) \, A' + \Lambda, \tag{3}$$

where $\{\boldsymbol{W}(t)\}$ is the q-dimensional standard Brownian motion and

$$\Lambda = \sum_{h=1}^{\infty} \mathrm{E}(\boldsymbol{u}_0 \, \boldsymbol{u}'_h) = \sum_{h=1}^{\infty} \Gamma(h) = \lim_{T \to \infty} \frac{1}{T} \sum_{j=1}^{T} \mathrm{E}(\boldsymbol{y}_{j-1} \boldsymbol{u}'_j).$$

In the proof, the following martingale approximation (B-N decomposition) plays an important role:

$$\boldsymbol{u}_j = A \, \boldsymbol{\varepsilon}_j + \tilde{\boldsymbol{\varepsilon}}_{j-1} - \tilde{\boldsymbol{\varepsilon}}_j, \tag{4}$$

where

$$\tilde{\boldsymbol{\varepsilon}}_j = \sum_{l=0}^{\infty} \tilde{A}_l \, \boldsymbol{\varepsilon}_{j-l}, \quad \tilde{A}_l = \sum_{k=l+1}^{\infty} A_k, \quad \sum_{l=0}^{\infty} ||\tilde{A}_l|| < \infty. \tag{5}$$

Here the process $\{\tilde{\boldsymbol{\varepsilon}}_j\}$ is evidently a stationary linear process because of the last condition in (5), which follows from (2). Then we have

$$\boldsymbol{y}_j = \sum_{i=1}^{j} \boldsymbol{u}_i = A \sum_{i=1}^{j} \boldsymbol{\varepsilon}_i + \tilde{\boldsymbol{\varepsilon}}_0 - \tilde{\boldsymbol{\varepsilon}}_j = A \boldsymbol{z}_j + \tilde{\boldsymbol{\varepsilon}}_0 - \tilde{\boldsymbol{\varepsilon}}_j,$$

where

$$\boldsymbol{z}_j = \sum_{i=1}^{j} \boldsymbol{\varepsilon}_i = \boldsymbol{z}_{j-1} + \boldsymbol{\varepsilon}_j, \quad \boldsymbol{z}_0 = 0.$$

The following may be a simplified proof of Phillips (1988). First of all, we have

$$\begin{aligned}\sum_{j=1}^{T} \boldsymbol{y}_{j-1}\boldsymbol{u}_j' &= \sum_{j=1}^{T} [A\boldsymbol{z}_{j-1} + \tilde{\boldsymbol{\varepsilon}}_0 - \tilde{\boldsymbol{\varepsilon}}_{j-1}]\,[A\boldsymbol{\varepsilon}_j + \tilde{\boldsymbol{\varepsilon}}_{j-1} - \tilde{\boldsymbol{\varepsilon}}_j]' \\ &= A\sum_{j=1}^{T} \boldsymbol{z}_{j-1}\boldsymbol{\varepsilon}_j' A' + H_T + R_T, \end{aligned} \tag{6}$$

where

$$H_T = \sum_{j=1}^{T} \left[A\boldsymbol{\varepsilon}_j\tilde{\boldsymbol{\varepsilon}}_j' - \tilde{\boldsymbol{\varepsilon}}_{j-1}\boldsymbol{u}_j'\right], \quad R_T = \sum_{j=1}^{T} \left[A\boldsymbol{z}_{j-1}(\tilde{\boldsymbol{\varepsilon}}_{j-1} - \tilde{\boldsymbol{\varepsilon}}_j)' - A\boldsymbol{\varepsilon}_j\tilde{\boldsymbol{\varepsilon}}_j' + \tilde{\boldsymbol{\varepsilon}}_0\boldsymbol{u}_j'\right].$$

Here it holds

$$\begin{aligned}\frac{1}{T}H_T &\to \mathrm{E}\left(A\boldsymbol{\varepsilon}_j\tilde{\boldsymbol{\varepsilon}}_j' - \tilde{\boldsymbol{\varepsilon}}_{j-1}\boldsymbol{u}_j'\right) = A(A - A_0) - \sum_{l=0}^{\infty}\left(\sum_{k=l+1}^{\infty} A_k\right) A_{l+1}' \\ &= \sum_{l=0}^{\infty}\sum_{k=1}^{\infty} A_l A_{k+l}' = \sum_{h=1}^{\infty} \mathrm{E}(\boldsymbol{u}_0\boldsymbol{u}_h') = \sum_{h=1}^{\infty}\Gamma(h), \\ \frac{1}{T}R_T &= \frac{1}{T}\sum_{j=1}^{T}\left[A\boldsymbol{z}_{j-1}\tilde{\boldsymbol{\varepsilon}}_{j-1}' - A(\boldsymbol{z}_j - \boldsymbol{\varepsilon}_j)\tilde{\boldsymbol{\varepsilon}}_j' - A\boldsymbol{\varepsilon}_j\tilde{\boldsymbol{\varepsilon}}_j' + \tilde{\boldsymbol{\varepsilon}}_0\boldsymbol{u}_j'\right] \\ &= \frac{1}{T}\left[A(\boldsymbol{z}_0\tilde{\boldsymbol{\varepsilon}}_0' - \boldsymbol{z}_T\tilde{\boldsymbol{\varepsilon}}_T') + \tilde{\boldsymbol{\varepsilon}}_0\sum_{j=1}^{T}\boldsymbol{u}_j'\right] \to 0,\end{aligned}$$

where the symbol "$\to$" signifies convergence in probability. Moreover it follows from Chan and Wei (1988) that

$$\frac{1}{T}\sum_{j=1}^{T}\boldsymbol{z}_{j-1}\boldsymbol{\varepsilon}_j' = \frac{1}{T}\sum_{j=1}^{T}(\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2 + \ldots + \boldsymbol{\varepsilon}_{j-1})\boldsymbol{\varepsilon}_j' \Rightarrow \int_0^1 \boldsymbol{W}(t)\,d\boldsymbol{W}'(t). \tag{7}$$

Therefore we can establish (3) by the continuous mapping theorem (CMT).

In particular, if $\{y_j\}$ and $\{u_j\}$ are scalar processes, we have

$$\frac{1}{T}\sum_{j=1}^{T} y_{j-1}u_j \Rightarrow 2\pi f_u(0)\int_0^1 W(t)\,dW(t) + \lambda,$$

where $f_u(\omega)$ is the spectrum of $\{u_j\}$, whereas $\{W(t)\}$ is the one-dimensional standard Brownian motion, and

$$\lambda = \sum_{h=1}^{\infty}\gamma(h) = \frac{1}{2}\left(2\pi f_u(0) - \gamma(0)\right), \quad \gamma(h) = \mathrm{E}(u_0u_h).$$

It also holds that, for $y_j = \rho y_{j-1} + u_j$ with $\rho = 1$, the LSE $\hat{\rho}$ of ρ follows

$$T(\hat{\rho} - 1) = \frac{\sum_{j=2}^{T} y_{j-1}u_j/T}{\sum_{j=2}^{T} y_{j-1}^2/T^2} \Rightarrow \frac{\int_0^1 W(t)\,dW(t) + \frac{1}{2}(1 - r)}{\int_0^1 W^2(t)\,dt}, \tag{8}$$

where r is the ratio of the short-run variance to the long-run variance defined by

$$r = \frac{\gamma(0)}{2\pi f(0)} = \frac{\gamma(0)}{\sum_{h=-\infty}^{\infty}\gamma(h)} = \frac{1}{\sum_{h=-\infty}^{\infty}\rho(h)}.$$

Here $\rho(h) = \gamma(h)/\gamma(0)$ is the lag h auto-correlation of $\{u_j\}$.

As another application, let us consider the unit root seasonal model

$$y_j = \rho_m \, y_{j-m} + v_j, \quad \rho_m = 1, \quad (j = 1, \ldots, T), \tag{9}$$

where m is the period, whereas $y_j = 0$ for $j \le 0$, and

$$v_j = \sum_{l=0}^{\infty} \alpha_{lm} \, \varepsilon_{j-lm}, \quad \sum_{l=1}^{\infty} l \, |\alpha_{lm}| < \infty.$$

Then we have, for $N = [T/m]$, the LSE $\hat{\rho}_m$ of ρ_m follows

$$N(\hat{\rho}_m - 1) \ \Rightarrow \ \frac{\int_0^1 \boldsymbol{W}'(t) \, d\boldsymbol{W}(t) + \frac{m}{2} \{1 - \gamma(0)/(2\pi f_v(0))\}}{\int_0^1 \boldsymbol{W}'(t)\boldsymbol{W}(t) \, dt}, \tag{10}$$

where $\{\boldsymbol{W}(t)\}$ is the $m-$dimensional standard Brownian motion, whereas $f_v(\omega)$ is the spectrum of $\{v_j\}$.

2.2. Fourier transforms

The martingale approximation in (4) was derived by expanding a polynomial in a Taylor series about unity. Here we consider expanding a polynomial about an arbitrary complex value on the unit circle, which will prove useful when we consider the asymptotic distributions of statistics associated with Fourier transforms. Following Phillips and Solo (1992) and Phillips (1999), let us consider

$$u_j = \sum_{l=0}^{\infty} \alpha_l \, \varepsilon_{j-l} = \alpha(L)\varepsilon_j, \quad \sum_{j=1}^{\infty} l|\alpha_l| < \infty, \quad \alpha_0 = 1, \quad \sum_{j=0}^{\infty} \alpha \ne 0, \tag{11}$$

where $\{\varepsilon_j\} \sim$ i.i.d.$(0, \sigma^2)$, whereas $\alpha(L) = 1 + \alpha_1 L + \alpha_2 L^2 + \ldots$ with L being the lag operator. Then, expanding the polynomial $\alpha(z)$ about $z = e^{i\theta}$ with $0 < \theta < \pi$, we have

$$\alpha(z) = \alpha(e^{i\theta}) + \alpha(z) - \alpha(e^{i\theta}) = \alpha(e^{i\theta}) + (e^{-i\theta}z - 1)\tilde{\alpha}(z),$$

where

$$\tilde{\alpha}(z) = \sum_{l=0}^{\infty} \tilde{\alpha}_l z^l, \qquad \tilde{\alpha}_l = \sum_{k=l+1}^{\infty} \alpha_k e^{i(k-l)\theta}.$$

Thus we have

$$u_j = \alpha(L)\varepsilon_j = \left[\alpha(e^{i\theta}) + (e^{-i\theta}L - 1)\tilde{\alpha}(L)\right]\varepsilon_j = \alpha(e^{i\theta})\varepsilon_j + e^{-i\theta}\tilde{\varepsilon}_{j-1} - \tilde{\varepsilon}_j, \tag{12}$$

where

$$\tilde{\varepsilon}_j = \tilde{\alpha}(L)\varepsilon_j = \sum_{l=0}^{\infty} \tilde{\alpha}_l \varepsilon_{j-l}.$$

Note that the process $\{\tilde{\varepsilon}_j\}$ is complex-valued because coefficients $\tilde{\alpha}_l$'s are, and is stationary because of the assumption given in (11). It is also noticed that the expansion (12) reduces to that in the time domain when $\theta = 0$.

It then follows that

$$
\begin{aligned}
\sum_{j=1}^{T} e^{ij\theta} u_j &= \alpha(e^{i\theta}) \sum_{j=1}^{T} e^{ij\theta} \varepsilon_j + \sum_{j=1}^{T} e^{i(j-1)\theta} \tilde{\varepsilon}_{j-1} - \sum_{j=1}^{T} e^{ij\theta} \tilde{\varepsilon}_j \\
&= \alpha(e^{i\theta}) \sum_{j=1}^{T} e^{ij\theta} \varepsilon_j + \tilde{\varepsilon}_0 - e^{iT\theta} \tilde{\varepsilon}_T. \qquad (13)
\end{aligned}
$$

Taking the real and imaginary parts of this last expression separately, we obtain

$$
\sum_{j=1}^{T} \begin{pmatrix} \cos j\,\theta \\ \sin j\,\theta \end{pmatrix} u_j = \begin{pmatrix} a(\theta) & -b(\theta) \\ b(\theta) & a(\theta) \end{pmatrix} \sum_{j=1}^{T} \begin{pmatrix} \cos j\,\theta \\ \sin j\,\theta \end{pmatrix} \varepsilon_j + \begin{pmatrix} R_{1T} \\ R_{2T} \end{pmatrix}, \qquad (14)
$$

where

$$
a(\theta) = \mathrm{Re}[\alpha(e^{i\theta})],\ b(\theta) = \mathrm{Im}[\alpha(e^{i\theta})],\ R_{1T} = \mathrm{Re}[\tilde{\varepsilon}_0 - e^{iT\theta}\tilde{\varepsilon}_T],\ R_{2T} = \mathrm{Im}[\tilde{\varepsilon}_0 - e^{iT\theta}\tilde{\varepsilon}_T].
$$

It is seen that the quantities on the left of (14) are trigonometric coefficients computed from the stationary process $\{u_j\}$, which can be approximated by a linear transformation of trigonometric coefficients computed from the i.i.d. process. Note that R_{1T} and R_{2T} are negligible as compared with the dominant term.

From the above expansion, it is straightforward to derive the distribution of trigonometric coefficients

$$
X_T(\theta) = \sum_{j=1}^{T} u_j \cos j\theta, \qquad Y_T(\theta) = \sum_{j=1}^{T} u_j \sin j\theta, \qquad (0 < \theta < \pi). \qquad (15)
$$

Asymptotic normality of $X_T(\theta)$ and $Y_T(\theta)$ was earlier given in Anderson (1971) by a somewhat complicated approach. Here we use the expansion (14). First of all we employ results by Helland (1982) and Chan and Wei (1988), which implies that

$$
\frac{\sqrt{2}}{\sqrt{T}\sigma} \sum_{j=1}^{[Tt]} \begin{pmatrix} \cos j\,\theta \\ \sin j\,\theta \end{pmatrix} \varepsilon_j \ \Rightarrow\ \boldsymbol{W}(t), \qquad \frac{\sqrt{2}}{\sqrt{T}\sigma} \sum_{j=1}^{T} \begin{pmatrix} \cos j\,\theta \\ \sin j\,\theta \end{pmatrix} \varepsilon_j \ \Rightarrow\ \mathrm{N}(\mathbf{0}, I_2),
$$

where $\{\boldsymbol{W}(t)\}$ is the two-dimensional standard Brownian motion. Because of the above results, we can immediately establish, by using (14) and the CMT,

Theorem 2.1. For the linear process $\{u_j\}$ defined by (11) it holds that

$$
\frac{\sqrt{2}}{\sqrt{T}\sigma} \sum_{j=1}^{[Tt]} \begin{pmatrix} \cos j\,\theta \\ \sin j\,\theta \end{pmatrix} u_j \quad \Rightarrow \quad K(\theta)\, \boldsymbol{W}(t), \quad (0 \le t \le 1), \qquad (16)
$$

where

$$
K(\theta) = \begin{pmatrix} a(\theta) & -b(\theta) \\ b(\theta) & a(\theta) \end{pmatrix}, \quad a(\theta) = \mathrm{Re}[\alpha(e^{i\theta})], \quad b(\theta) = \mathrm{Im}[\alpha(e^{i\theta})]. \qquad (17)
$$

It follows from Theorem 2.1 that

$$
\frac{\sqrt{2}}{\sqrt{T}\sigma} \sum_{j=1}^{T} \begin{pmatrix} \cos j\,\theta \\ \sin j\,\theta \end{pmatrix} u_j \ \Rightarrow\ \mathrm{N}(\mathbf{0}, K(\theta)\, K'(\theta)) = \mathrm{N}\left(\mathbf{0}, \frac{2\pi}{\sigma^2} f(\theta)\, I_2\right),
$$

where $f(\theta) = \sigma^2(a^2(\theta)+b^2(\theta))/(2\pi)$ is the spectrum of $\{u_j\}$ evaluated at θ. More generally, let us consider trigonometric coefficients defined by

$$A(\theta_k) = \frac{1}{\sqrt{T}}\sum_{j=1}^{T} u_j \cos j\,\theta_k, \quad B(\theta_k) = \frac{1}{\sqrt{T}}\sum_{j=1}^{T} u_j \sin j\,\theta_k, \ (k = 1, \ \ldots, \ n),$$

where $0 < \theta_1 < \ \ldots, < \theta_n < \pi$. Then we have

$$A(\theta_1), B(\theta_1), \ \ldots, \ A(\theta_n), B(\theta_n) \ \Rightarrow \ \mathrm{N}(\mathbf{0}, D),$$

where

$$D = \mathrm{diag}\,(\pi f(\theta_1), \pi f(\theta_1), \ \ldots, \ \pi f(\theta_n), \pi f(\theta_n))$$

2.3. Complex unit roots

As another application of the expansion (14), let us consider the following model

$$(1 - e^{i\theta}L)(1 - e^{-i\theta}L)\, y_j = u_j \quad \Leftrightarrow \quad y_j = \phi_1 y_{j-1} + \phi_2 y_{j-2} + u_j, \tag{18}$$

where $\phi_1 = 2\cos\theta$, $\phi_2 = -1$, and $\{u_j\}$ is the stationary linear process given in (11). The present model describes a nonstationary cyclical behavior with period $2\pi/\theta$ and is also discussed in Bierens (2001). Ahtola and Tiao (1987) originally considered the model (18) with the error term being i.i.d., where the asymptotic distribution of the LSEs of ϕ_1 and ϕ_2 was discussed. Here we consider the same problem for the extended model (18), and compute, by numerical integration, the limiting distribution.

Assuming that $y_{-1} = y_0 = 0$, we have

$$\begin{aligned} y_j &= \frac{u_j}{(1 - e^{i\theta}L)(1 - e^{-i\theta}L)} = \frac{1}{\sin\theta}\,[\, z_{1j}\sin(j+1)\theta - z_{2j}\cos(j+1)\theta\,] \\ &= \frac{1}{\sin\theta}\boldsymbol{a}_j'\boldsymbol{z}_j, \end{aligned} \tag{19}$$

where

$$\boldsymbol{z}_j = \begin{pmatrix} z_{1j} \\ z_{2j} \end{pmatrix} = \sum_{l=1}^{j} \begin{pmatrix} \cos l\theta \\ \sin l\theta \end{pmatrix} u_l, \quad \boldsymbol{a}_j = \begin{pmatrix} \sin(j+1)\theta \\ -\cos(j+1)\theta \end{pmatrix}.$$

Applying the expansion (14) to u_l in the expression for $\boldsymbol{z}_j$, we obtain

$$\boldsymbol{z}_j = \sum_{l=1}^{j} \begin{pmatrix} \cos l\theta \\ \sin l\theta \end{pmatrix} u_l = K(\theta)\boldsymbol{x}_j + \boldsymbol{w}_j, \tag{20}$$

where $K(\theta)$ is defined in (17), whereas

$$\boldsymbol{x}_j = \sum_{l=1}^{j} \begin{pmatrix} \cos l\theta \\ \sin l\theta \end{pmatrix} \varepsilon_l, \quad \boldsymbol{w}_j = \begin{pmatrix} \mathrm{Re}[\tilde{\varepsilon}_0 - e^{ij\theta}\tilde{\varepsilon}_j] \\ \mathrm{Im}[\tilde{\varepsilon}_0 - e^{ij\theta}\tilde{\varepsilon}_j] \end{pmatrix}.$$

Then we obtain the asymptotic distribution of various second moments as follows (see, for more details, Tanaka (2008)).

$$\begin{aligned}
R_T(h) &= \frac{1}{T^2}\sum_{j=h+1}^{T} y_{j-h}\, y_j \;\Rightarrow\; \frac{\pi f(\theta)\cos h\theta}{2\sin^2\theta}\int_0^1 \boldsymbol{W}'(t)\boldsymbol{W}(t)\,dt,\\
S_T(h) &= \frac{1}{T}\sum_{j=h+1}^{T} y_{j-h}\, u_j\\
&\Rightarrow \frac{1}{\sin\theta}\left[\pi f(\theta)\int_0^1 \boldsymbol{W}'(t)\, J_h(\theta)\, d\boldsymbol{W}(t) + \sum_{j=h}^{\infty}\gamma(j)\sin(j-h+1)\theta\right],
\end{aligned}$$

where $f(\theta)$ is the spectrum of $\{u_j\}$ evaluated at θ, and $\gamma(h)$ is the lag h auto-covariance of $\{u_j\}$, whereas

$$J_h(\theta) = \begin{pmatrix} -\sin(h-1)\theta & \cos(h-1)\theta \\ -\cos(h-1)\theta & -\sin(h-1)\theta \end{pmatrix}. \tag{21}$$

Denoting by $\hat{\boldsymbol{\phi}}$ the LSE of $\boldsymbol{\phi} = (\phi_1, \phi_2)'$ in the present model, we can now establish

Theorem 2.2. For the AR(2) model with complex unit roots defined by (18) it holds that

$$\begin{aligned}
T(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) &= \left[\frac{1}{T^2\sigma^2}\sum_{j=3}^{T}\begin{pmatrix} y_{j-1}^2 & y_{j-1}y_{j-2} \\ y_{j-1}y_{j-2} & y_{j-2}^2 \end{pmatrix}\right]^{-1}\left[\frac{1}{T\sigma^2}\sum_{j=3}^{T}\begin{pmatrix} y_{j-1}u_j \\ y_{j-2}u_j \end{pmatrix}\right]\\
&\Rightarrow \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix},
\end{aligned}$$

$$\begin{aligned}
Z_1 &= 2\left[\pi f(\theta)\int_0^1 \boldsymbol{W}'(t)\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} d\boldsymbol{W}(t)\right.\\
&\qquad \left. + \sin\theta\sum_{j=1}^{\infty}\gamma(j)\cos(j-1)\theta\right] \Big/ \left\{\pi f(\theta)\int_0^1 \boldsymbol{W}'(t)\boldsymbol{W}(t)\,dt\right\}, \qquad (22)\\
Z_2 &= \frac{-2\left[\int_0^1 \boldsymbol{W}'(t)\,d\boldsymbol{W}(t) + 1 - \gamma(0)/\left(2\pi f(\theta)\right)\right]}{\int_0^1 \boldsymbol{W}'(t)\boldsymbol{W}(t)\,dt}. \qquad (23)
\end{aligned}$$

It is seen that the asymptotic distributions of $\hat{\phi}_1$ and $\hat{\phi}_2$ both depend on the parameter θ in the present model. Note that, when the error term $\{u_j\}$ becomes independent, the distribution of the former still depends on θ, but the latter does not. More specifically, it holds that, when $u_j = \varepsilon_j$ so that $\gamma(0) = 2\pi f(\theta) = \sigma^2$,

$$\begin{aligned}
T(\hat{\phi}_1 + 1) &\Rightarrow \frac{2\int_0^1 \boldsymbol{W}'(t)\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} d\boldsymbol{W}(t)}{\int_0^1 \boldsymbol{W}'(t)\boldsymbol{W}(t)\,dt},\\
T(\hat{\phi}_2 + 1) &\Rightarrow \frac{-2\int_0^1 \boldsymbol{W}'(t)\,d\boldsymbol{W}(t)}{\int_0^1 \boldsymbol{W}'(t)\boldsymbol{W}(t)\,dt}.
\end{aligned}$$

It is of some interest to compare the distribution of Z_2 in (23) with the unit root seasonal distribution in (10) which arises from the seasonal model in (9). In particular, it is seen that $-Z_2/2$ has the same distribution as the seasonal unit root distribution when $m = 2$ and $u_j = \varepsilon_j$.

The distribution of Z_2 may be called the complex unit root distribution, and can be computed as follows. Let us put $r = \gamma(0)/(2\pi f(\theta))$. Then we have the following theorem (see, for details, Tanaka (1996))

Theorem 2.3.

$$P\left(Z_2 \le x\right) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{1}{u}\mathrm{Im}[\psi(u)]\,du, \tag{24}$$

where $\psi(u)$ is the characteristic function (c.f.) of

$$X = -\frac{x}{2}\int_0^1 \boldsymbol{W}'(t)\boldsymbol{W}(t)\,dt - \int_0^1 \boldsymbol{W}'(t)\,d\boldsymbol{W}(t) - (1-r),$$

and is given by

$$\psi(u) = \mathrm{E}(e^{iuX}) = e^{iru} \Bigg/ \left(\cos\sqrt{2iuy} + iu\frac{\sin\sqrt{2iuy}}{\sqrt{2iuy}}\right), \quad y = -x/2.$$

Figure 1 draws the probability density of Z_2 for $r = 0.2, 0.5, 1, 1.5, 2$, which can be computed from (24) by numerical integration. It is seen that the distribution is shifted to the right as r becomes large. Table 1 reports percent points and moments of Z_2 for $r = 1$.

Table 1. Percent points and some moments of Z_2 for $r = 1$

0.01	0.05	0.1	0.5	0.9	0.95	0.99	Mean	SD
−2.956	−2.028	−1.539	0.775	6.010	8.389	14.115	1.664	3.460

Figure 1

3. K-asymptotics in integrated and near-integrated processes

3.1. Case of integrated processes

Let us consider the scalar I(1) process

$$y_j = y_{j-1} + u_j, \qquad y_0 = 0, \qquad (j = 1,\ \ldots,\ T), \tag{25}$$

where $\{u_j\}$ is a stationary linear process defined by

$$u_j = \sum_{l=0}^{\infty}\alpha_l\varepsilon_{j-l}, \quad \sum_{l=1}^{\infty} l\,|\alpha_l| < \infty, \quad \alpha_0 = 1, \quad \alpha \equiv \sum_{l=0}^{\infty}\alpha_l \neq 0, \tag{26}$$

with $\{\varepsilon_j\} \sim$ i.i.d.$(0,\ \sigma^2)$. We denote the short-run and long-run variances of $\{u_j\}$ by $\sigma_S^2 = \mathrm{V}(u_j)$ and $\sigma_L^2 = \alpha^2\sigma^2$, respectively. Then, for the partial sum process given by

$$X_T(t) = \frac{1}{\sqrt{T}\sigma_L}\, y_{[Tt]} = \frac{1}{\sqrt{T}\sigma_L}\sum_{j=1}^{[Tt]} u_j, \quad (0 \le t \le 1),$$

the following FCLT holds:

$$X_T(\cdot) \Rightarrow W(\cdot),$$

where $\{W(t)\}$ is the standard Brownian motion on $[0,1]$. This is a typical invariance principle in the weak version, while the strong version [Csörgő and Horváth (1993)] says that $\{X_T(t)\}$ can be well approximated by $\{W(t)\}$ in the sense that, if $\mathrm{E}(|\varepsilon_j|^p) < \infty$ for some $p > 2$, we can construct a standard Brownian motion such that

$$\sup_{0\le t\le 1} T^\delta \left|X_T(t) - W(t)\right| = \sup_{0\le t\le 1} T^\delta \left|\frac{1}{\sqrt{T}\sigma_L} y_{[Tt]} - W(t)\right| \longrightarrow 0$$

with probability 1, where $0 < \delta = 1/2 - 1/p < 1/2$.

On the other hand, it is known [Loève (1978), Chan and Wei (1988)] that $W(t)$ admits infinitely many series representations in the L_2-sense. For example, we have

$$W(t) = \sum_{k=1}^{m} g_k(t)\nu_k + \sum_{n=1}^{\infty} \frac{f_n(t)}{\sqrt{\lambda_n}}\xi_n\,, \tag{27}$$

where $\{\nu_k\} \sim$ NID(0,1), $\{\xi_n\} \sim$ NID(0,1), and the two sequences are independent of each other, whereas $g_k(t)$ is a continuous function. Moreover, λ_n is the nth smallest eigenvalue, and $f_n(t)$ is the corresponding orthonormal eigenfunction for the integral equation

$$f(t) = \lambda \int_0^1 K(s,t)\, f(s)\, ds, \tag{28}$$

where $K(s,t)$ is a continuous, symmmetric and positive definite kernel defined by

$$K(s,t) = \mathrm{Cov}\left(W(s) - \sum_{k=1}^{m} g_k(s)\nu_k, W(t) - \sum_{l=1}^{m} g_l(t)\nu_l\right) = \sum_{n=1}^{\infty} \frac{f_n(s)f_n(t)}{\lambda_n}\,. \tag{29}$$

The expansion in (27) is called the Karhunen-Loève expansion, where the λ_n is repeated as many times as its multiplicity which is defined by the number of linearly independent solutions of eigenfunctions corresponding to λ_n. Note also that the expansion in (29) is ensured to converge absolutely and uniformly to $K(s,t)$ by Mercer's theorem (Hochstadt (1973)).

Thus, allowing for various values of m and various functions $g_k(t)$, we obtain infinitely many series representations for $W(t)$ in the L_2-sense.Among such representations the most convenient for the present purpose is

$$W(t) = \sum_{n=1}^{\infty} \frac{\phi_n(t)}{(n-1/2)\pi}\xi_n\,, \quad \phi_n(t) = \sqrt{2}\sin[(n-1/2)\pi t], \tag{30}$$

where $\lambda_n = (n-1/2)^2\pi^2$ is the nth smallest eigenvalue of the kernel $K(s,t) = \min(s,t)$, while $\phi_n(t)$ is the corresponding orthonormal eigenfunction. It can be checked that the multiplicity is unity for each eigenvalue. Because of the martingale convergence theorem, the representation in (30) holds, not only in the L_2-sense, but also with probability 1, uniformly in $t \in [0,1]$. It follows that $\{W(t)\}$ can be represented by an infinite number of linear combinations of trigonometric functions with random coefficients.

On the basis of the above facts, Phillips (1998) considers approximating the I(1) process, that is, the process that contains purely stochastic trends, by trigonometric functions with stochastic coefficients. More specifically, the following regression was considered:

$$y_j = \sum_{k=1}^{K} \hat{b}_k \phi_k \left(\frac{j}{T}\right) + \hat{u}_j, \quad (j = 1,\ ...,\ T), \tag{31}$$

where $\hat{b}_1,\ ...,\ \hat{b}_K$ are LSEs and $\hat{u}_j$ is the OLS residual. We also denote the t-ratio statistic for $\hat{b}_k$ by $t_{\hat{b}_k}$, the coefficient of determination by R^2, and the Durbin-Watson statistic by DW.

Under the above setting, Phillips (1998) obtained T-asymptotics for the above statistics. The following are just a summary of those results.

(a) $\hat{b}_1/\sqrt{T},\ ...,\ \hat{b}_K/\sqrt{T}$ tend to be independently distributed as normal.

(b) $\sum_{j=1}^{T} \hat{u}_j^2 = O_p(T^2)$,

(c) $t_{\hat{b}_k} = O_p(\sqrt{T})$,

(d) R^2 has a non-degenerate limiting distribution.

(e) $DW = O_p(1/T)$.

The result (a) comes from (30) and the orthogonality of $\{\phi_n(t)\}$. The result (b) implies that the regression residuals still contain nonstaionary components, whereas it follows from (c) that the regression coefficients of deterministic trends are significant. This also applies when the robust t-ratio which accommodates serial dependence in the residuals is used. Moreover, the result (d) also signals that the fitted regression is valid. The result (e), however, serves as conventional wisdom that detects poor performance of the fitted model.

We move on to K-asymptotics by letting $T \to \infty$ and then $K \to \infty$. It holds that

(a) $\hat{b}_1/\sqrt{T},\ ...,\ \hat{b}_K/\sqrt{T}$ tend to be independently distributed as normal.

(b) $\sum_{j=1}^{T} \hat{u}_j^2/T^2 = O_p(1/K)$,

(c) $t_{\hat{b}_k}/\sqrt{T} = O_p(\sqrt{K})$,

(d) $R^2 \to 1$ in probability,

(e) $T \times DW = O_p(K)$.

All of the above statistics signal that the regression relation is valid in K-asymptotics. We have that the coefficients of deterministic trends are still significant because of (c), and the regression (31) fully captures the variation of $\{y_j\}$ because of (d). Moreover, as described in (e), the DW statistic does produce a nonnegligible value. More specifically, it holds that, as $T \to \infty$ and then $K \to \infty$, $T \times DW/K$ converges to π^2 in probability. In conclusion, stochastic trends cannot be distinguished from deterministic trends in K-asymptotics of the integrated process.

It is of great interest to study K-asymptotics in nested models for unit root tests. To this end we consider the regression relation

$$y_j = \hat{\rho} y_{j-1} + \sum_{k=1}^{K} \hat{b}_k \phi_k \left(\frac{j}{T} \right) + \hat{u}_j, \quad (j = 1,\ ...,\ T). \tag{32}$$

We first deal with T-asymptotics. Let ADF_ρ and Z_ρ be the unit root coefficient statistics suggested in Said and Dickey (1984) and Phillips (1987), respectively. Let also ADF_t and Z_t be the corresponding unit root t-ratio statistics. Then Phillips (2002) proved that it holds that, as $T \to \infty$,

$$ADF_\rho,\ Z_\rho \quad \Rightarrow \quad \frac{\int_0^1 W_{\boldsymbol{\phi}_K}(t)\, dW(t)}{\int_0^1 W^2_{\boldsymbol{\phi}_K}(t)\, dt}, \tag{33}$$

$$ADF_t,\ Z_t \quad \Rightarrow \quad \frac{\int_0^1 W_{\boldsymbol{\phi}_K}(t)\, dW(t)}{\left(\int_0^1 W^2_{\boldsymbol{\phi}_K}(t)\, dt \right)^{1/2}}, \tag{34}$$

where

$$W_{\boldsymbol{\phi}_K}(t) = W(t) - \int_0^1 \boldsymbol{\phi}(K,s)' W(s)\, ds\, \boldsymbol{\phi}(K,t), \quad \boldsymbol{\phi}(K,t) = (\phi_1(t),\ ...,\ \phi_K(t))'.$$

Note that the process $\{W_{\boldsymbol{\phi}_K}(t)\}$ is a detrended Brownian motion, which is the residual process of the Hilbert space projection of $W(t)$ on the space spanned by $\boldsymbol{\phi}(K,t)$.

It also follows that, unlike in the case of the purely deterministic regression in (31), the estimators $\hat{b}_1,\ ...,\ \hat{b}_K$ do not tend to normality, and the stochastic order of these estimators decrease to $1/\sqrt{(T)}$ so that the estimators are consistent and converge to the true value of 0. Nonetheless it holds that the corresponding t-ratio is $O_p(1)$ but not $o_p(1)$, which implies possible significance of the coefficients of deterministic trends.

This last statement becomes much clearer in K-asymptotics. In fact we have that $\sqrt{T}\hat{b}_k = O_p(K)$ and the t-ratio for $\hat{b}_k$ is of order $\sqrt{K}$. Moreover Phillips (2002) proved that, as K becomes large,

$$ADF_\rho,\ Z_\rho \quad \Rightarrow \quad \mathrm{N}\left(-\frac{\pi^2 K}{2},\ \frac{\pi^4 K}{6} \right) = \mathrm{N}(-4.93K,\ 16.23K), \tag{35}$$

$$ADF_t,\ Z_t \quad \Rightarrow \quad \mathrm{N}\left(-\frac{\pi\sqrt{K}}{2},\ \frac{\pi^2}{24} \right) = \mathrm{N}(-1.57\sqrt{K},\ 0.41). \tag{36}$$

The asymptotic distributions described in (35) and (36) do vary depending on deterministic trends chosen as regressors. In fact, when we consider a usual model for unit root tests that uses polynomials given by

$$y_j = \hat{\rho} y_{j-1} + \sum_{k=0}^{K} \hat{b}_k \left(\frac{j}{T} \right)^k + \hat{u}_t, \quad (j = 1,\ ...,\ T), \tag{37}$$

Nabeya (1999) obtained the following results as K becomes large.

$$ADF_\rho,\, Z_\rho \quad \Rightarrow \quad \mathrm{N}(-4K,\, 16K), \tag{38}$$

$$ADF_t,\, Z_t \quad \Rightarrow \quad \mathrm{N}\left(-\sqrt{2K},\, 1/2\right). \tag{39}$$

Note that (38) and (39) are polynomial versions of (35) and (36), respectively. It is seen that the asymptotic distributions based on trigonometric and polynomial functions are different, although they are close to each other.

3.2. Case of near-integrated processes

We next extend the above arguments to deal with near-integrated processes. Thus we consider

$$y_j = \rho y_{j-1} + u_j, \qquad y_0 = 0, \qquad \rho = 1 - (c/T), \qquad (j = 1,\, \ldots,\, T), \tag{40}$$

where c is a fixed positive constant, whereas $\{u_j\}$ is a stationary process described in (26) with the long-run variance σ_L^2. Then it is known that, for the partial sum process defined by

$$Y_T(t) = \frac{1}{\sqrt{T}\sigma_L} y_{[Tt]} = \frac{1}{\sqrt{T}\sigma_L} \sum_{j=1}^{[Tt]} \rho^{[Tt]-j} u_j\,, \tag{41}$$

the following FCLT holds:

$$Y_T(\cdot) \Rightarrow J^c(\cdot)\,,$$

where $\{J^c(t)\}$ is the Ornstein-Uhlenbeck (O-U) process given by

$$J^c(t) = e^{-ct}\int_0^t e^{cs}\, dW(s) \quad \Leftrightarrow \quad dJ^c(t) = -cJ^c(t)\, dt + dW(t), \quad J^c(0) = 0\,. \tag{42}$$

The O-U process $\{J^c(t)\}$ admits the following series representation:

$$J^c(t) = \sum_{n=1}^{\infty} \frac{f_n(t)}{\sqrt{\lambda_n}}\, \xi_n\,, \tag{43}$$

where $\{\xi_n\} \sim$ NID(0,1), whereas λ_n is the n-th smallest eigenvalue of the positive definite kernel

$$\mathrm{Cov}(J^c(s), J^c(t)) = \frac{e^{-c|s-t|} - e^{-c(s+t)}}{2c}\,,$$

and $f_n(t)$ is the corresponding orthonormal eigenfunction. Unlike in the expansion (30) of the standand Brownian motion $W(t)$, it is impossible to obtain λ_n and $f_n(t)$ explicitly, although numerically possible once c is given. We can show (see, for details, Tanaka (2001)) that λ_n is the n-th smallest positive solution to

$$\tan\sqrt{\lambda - c^2} = -\frac{\sqrt{\lambda - c^2}}{c}\,.$$

Then it can be checked easily that

$$(n - 1/2)^2\pi^2 + c^2 < \lambda_n < n^2\pi^2 + c^2\,.$$

We also obtain

$$f_n(t)=\frac{\sin\mu_n t}{M_n},\quad \mu_n=\sqrt{\lambda_n-c^2},\quad M_n=\sqrt{\frac{1}{2}-\frac{\sin 2\mu_n}{4\mu_n}}. \tag{44}$$

Under the above setting we first consider the regression relation

$$y_j=\sum_{k=1}^{K}\tilde{b}_k f_k\left(\frac{j}{T}\right)+\tilde{u}_j=\tilde{\boldsymbol{b}}(K)'\boldsymbol{f}(K,t/T)+\tilde{u}_j\,,\qquad (j=1,\ \ldots,\ T), \tag{45}$$

where $\tilde{b}_1,\ \ldots,\ \tilde{b}_K$ are LSEs and $\tilde{u}_j$ is the OLS residual, whereas

$$\tilde{\boldsymbol{b}}(K)=(\tilde{b}_1,\ \ldots,\ \tilde{b}_K)',\quad \boldsymbol{f}(K,t)=(f_1(t),\ \ldots,\ f_K(t))'.$$

Then it holds that, as $T\to\infty$ under $\rho=1-(c/T)$,

$$\frac{1}{\sqrt{T}\sigma_L}\,\tilde{u}_{[Tt]}\quad\Rightarrow\quad J^c_{fK}(t)=J^c(t)-\int_0^1\boldsymbol{f}(K,s)'J^c(s)\,ds\,\boldsymbol{f}(K,t)\,. \tag{46}$$

We now have the following T-asymptotics for near-integrated processes (Tanaka (2001)).

Theorem 3.1. When $\{y_j\}$ follows a near-integrated process defined by (40), it holds for the regression relation (45) that

(a) $\boldsymbol{c}(K)'\tilde{\boldsymbol{b}}(K)\Big/\sqrt{T}\quad\Rightarrow\quad\sigma_L\boldsymbol{c}(K)'\int_0^1\boldsymbol{f}(K,t)J^c(t)\,dt$,

(b) $\sum_{j=1}^{T}\tilde{u}_j^2\Big/T^2\quad\Rightarrow\quad\sigma_L^2\int_0^1\{J^c_{fK}(t)\}^2\,dt$,

(c) $t_{\boldsymbol{c}(K)'\tilde{\boldsymbol{b}}(K)}\Big/\sqrt{T}\quad\Rightarrow\quad\boldsymbol{c}(K)'\int_0^1\boldsymbol{f}(K,t)J^c(t)\,dt\Big/\left(\int_0^1\{J^c_{fK}(t)\}^2\,dt\right)^{1/2}$,

(d) $R^2=1-\sum_{j=1}^{T}\tilde{u}_j^2\Big/\sum_{j=1}^{T}y_j^2\quad\Rightarrow\quad 1-\int_0^1\{J^c_{fK}(t)\}^2\,dt\Big/\int_0^1\{J^c(t)\}^2\,dt$,

(e) $T\times DW\quad\Rightarrow\quad\sigma_S^2\Big/\left(\sigma_L^2\int_0^1\{J^c_{fK}(t)\}^2\,dt\right)$,

where $\boldsymbol{c}(K)=(c_1,\ \ldots,\ c_K)'$ is any $K\times 1$ vector such that $\boldsymbol{c}(K)'\boldsymbol{c}(K)=1$.

Theorem 3.1 implies that T-asymptotics in the near-integrated process give essentillay the same results as in the integrated process. The only difference is that the limiting process $J^c(t)$ and the eigenfunction $\boldsymbol{f}(K,t)$ have been substituted for $W(t)$ and $\boldsymbol{\phi}(K,t)$, respectively. Then it follows from (43) and the orthonormality of $\{f_n(t)\}$ that

$$\int_0^1\boldsymbol{f}(K,t)J^c(t)\,dt=\left(\xi_1/\sqrt{\lambda_1},\ \ldots,\ \xi_K/\sqrt{\lambda_K}\right)',$$

which implies that the components of $\tilde{\boldsymbol{b}}(K)$ are asymptotically normal and independent of each other.

It may be noted that this eigenvalue decomposition does not hold if $\boldsymbol{f}(K,t)$ is replaced by the much simpler function $\boldsymbol{\phi}(K,t)$. The reason will be explained shortly. In any case use of $\boldsymbol{\phi}(K,t)$ rather than $\boldsymbol{f}(K,t)$ in the above theorem makes arguments complicated.

We now discuss K-asymptotics by letting $K\to\infty$ in Theorem 3.1, which yields

Theorem 3.2. For the regression relation (45), it holds that, as $T\to\infty$ and then as K becomes large,

(a) $\boldsymbol{c}(K)'\tilde{\boldsymbol{b}}(K)\Big/\sqrt{T} \quad \Rightarrow \quad \mathrm{N}(0,\sigma_c^2), \quad \sigma_c^2 = \sigma_L^2 \sum_{n=1}^{\infty} c_n^2/\lambda_n,$

(b) $\sum_{j=1}^{T} \tilde{u}_j^2\Big/T^2 = O_p(1/K),$

(c) $t_{\boldsymbol{c}(K)'\tilde{\boldsymbol{b}}(K)}\Big/\sqrt{T} = O_p(\sqrt{K}),$

(d) $R^2 \to 1$ in probability,

(e) $T \times DW = O_p(K).$

It is seen that K-asymptotics in the near-integrated process are, like T-asymptotics, essentially similar to those in the integrated process. We note, however, that, if the deterministic trend $\boldsymbol{f}(K,t)$ is replaced by $\boldsymbol{\phi}(K,t)$ in the regression equation (45), the above result (a) has to be changed. This is because, in that case, we have, as $T \to \infty$,

$$\begin{aligned}
\tilde{b}_k\Big/\left(\sqrt{T}\sigma_L\right) \quad &\Rightarrow \quad \int_0^1 \phi_k(t) J^c(t)\, dt \\
&= \sum_{n=1}^{\infty} \frac{\xi_n}{\sqrt{\lambda_n}} \int_0^1 f_n(t)\phi_k(t)\, dt \\
&= \sum_{n=1}^{\infty} \frac{\sqrt{2}\,\xi_n}{M_n\sqrt{\lambda_n}} \int_0^1 \sin \mu_n t\, \sin(k-1/2)\pi t\, dt \\
&= \sum_{n=1}^{\infty} \frac{\xi_n}{M_n\sqrt{2\lambda_n}} \left[\frac{\sin\left(\mu_n - (k-1/2)\pi\right)}{\mu_n - (k-1/2)\pi} - \frac{\sin\left(\mu_n + (k-1/2)\pi\right)}{\mu_n + (k-1/2)\pi}\right],
\end{aligned}$$

so that this last sum need not be $\xi_k/\sqrt{\lambda_k}$, which is to be attained when $\phi_k(t)$ is replaced by $f_k(t)$. This means that the components of $\tilde{\boldsymbol{b}}(K)$ are asymptotically not independent, but a closer examination reveals that, in the above infinite sum, the term corresponding to $n = k$ dominates and yields a value which is close to $\xi_k/\sqrt{\lambda_k}$. This property will be effectively used when we formulate a model for unit root tests.

We move on to deal with the regression relation

$$y_j = \hat{\rho} y_{j-1} + \sum_{k=1}^{K} \hat{b}_k f_k\left(\frac{j}{T}\right) + \hat{u}_j = \hat{\rho} y_{j-1} + \hat{\boldsymbol{b}}(K)'\boldsymbol{f}(K, j/T) + \hat{u}_j, \tag{47}$$

and obtain the following results.

Theorem 3.3. For the regression relation (47) it holds that, as $T \to \infty$,

$$ADF_\rho,\, Z_\rho \quad \Rightarrow \quad \frac{\int_0^1 J_{fK}^c(t)\, dJ^c(t)}{\int_0^1 \{J_{fK}^c(t)\}^2\, dt}, \tag{48}$$

$$ADF_t,\, Z_t \quad \Rightarrow \quad \frac{\int_0^1 J_{fK}^c(t)\, dJ^c(t)}{\left(\int_0^1 \{J_{fK}^c(t)\}^2\, dt\right)^{1/2}}, \tag{49}$$

$$\sqrt{T}\boldsymbol{c}(K)'\hat{\boldsymbol{b}}_K\Big/\sigma_L \quad \Rightarrow \quad \boldsymbol{c}(K)'\boldsymbol{Y}(K,f)\,, \tag{50}$$

$$t_{\boldsymbol{c}(K)'\hat{\boldsymbol{b}}(K)} \quad \Rightarrow \quad \boldsymbol{c}(K)'\boldsymbol{Y}(K,f) \Big/ \sqrt{\boldsymbol{c}(K)'\Sigma(K,f)^{-1}\boldsymbol{c}(K)} \ , \tag{51}$$

where

$$\begin{aligned}
\boldsymbol{Y}(K,f) &= \int_0^1 \boldsymbol{f}(K,t)\, dW(t) - \left(\frac{\int_0^1 J^c_{fK}(t)\, dJ^c(t) + (1-\sigma_S^2/\sigma_L^2)/2}{\int_0^1 \{J^c_{fk}(t)\}^2\, dt} + c \right) \\
&\quad \times \int_0^1 \boldsymbol{f}(K,t) J^c(t)\, dt \,, \\
\Sigma(K,f) &= I_K - \int_0^1 \boldsymbol{f}(K,t) J^c(t)\, dt \int_0^1 \boldsymbol{f}(K,t)' J^c(t)\, dt \Big/ \int_0^1 \{J^c(t)\}^2\, dt.
\end{aligned}$$

Theorem 3.4. For the regression relation (47) it holds that, as $T \to \infty$ and then as K becomes large,

$$ADF_\rho,\ Z_\rho \quad \Rightarrow \quad \mathrm{N}\left(-\frac{\pi^2 K}{2}, \frac{\pi^4 K}{6} \right), \tag{52}$$

$$ADF_t,\ Z_t \quad \Rightarrow \quad \mathrm{N}\left(-\frac{\pi\sqrt{K}}{2}, \frac{\pi^2}{24} \right), \tag{53}$$

$$\sqrt{T}\boldsymbol{c}(K)'\hat{\boldsymbol{b}}(K) \Big/ \sigma_S \quad \Rightarrow \quad \mathrm{N}\left(0,\ \frac{\pi^4 K^2}{4} \sum_{k=1}^{K} \frac{c_k^2}{\lambda_k} \right), \tag{54}$$

$$t_{\boldsymbol{c}(K)'\hat{\boldsymbol{b}}(K)} \quad = \quad O_p(\sqrt{K}). \tag{55}$$

It is seen that the T- and K-asymptotics in the near-integrated process are essentially the same as in the integrated process. In particular, it is quite interesting to notice that, in K-asymptotics, the statistics ADF_ρ and ADF_t are normally distributed independently of the near-integration parameter c. Note, however, that these statitsics do depend on c in T-asymptotics.

The regression relation (47) cannot be used as a model for tesing a unit root hypothesis $H_0 : \rho = 1$ because the deterministic regressor $\boldsymbol{f}(K,t)$ depends on the unknown parameter c. We should use the model (32) discussed in Section 3.1 as a suitable model for this purpose, for which Phillips (2002) derived the limiting distributions of test statistics ADF_ρ and ADF_t under H_0. To derive the limiting power under $H_1 : \rho = 1 - (c/T)$, we need consider the regression (32) under H_1, that is, under the situation where the true model is the near-integrated process (40). In that case, results on T-asymptotics can be obtained in the same way as in Theorem 3.3 by replacing $f_k(t)$ by $\phi_k(t)$. For instance, we have, as $T \to \infty$ under $\rho = 1 - (c/T)$,

$$ADF_\rho \quad \Rightarrow \quad \frac{\int_0^1 J^c_{\phi K}\, dJ^c(t)}{\int_0^1 \{J^c_{\phi K}(t)\}^2\, dt}.$$

It, however, turns out that results on K-asymptotics in the present case are not clear-cut because of the reason described below Theorem 3.2. We also mentioned there that replacing $f_k(t)$ by $\phi_k(t)$ affects K-asymptotics little. Thus it is expected that results similar to those in Theorem 3.4 hold true in this case. It follows that the unit root test loses its local power in K-asymptotics. This last point will be examined by simulations in the next subsection.

3.3. Some simulations

We examine, by simulations, the finite sample performance of T- and K-asymptotics developed in previous sections. For simplicity we assume the process $\{y_j\}$ to be generated by

$$y_j = \rho y_{j-1} + \varepsilon_j, \qquad y_0 = 0, \qquad \rho = 1 - (c/T), \qquad (j = 1, \ldots, T), \tag{56}$$

where $\{\varepsilon_j\} \sim \mathrm{NID}(0, 1)$ and c is a nonnegative constant.

The regression relations considered are

$$y_j = \sum_{k=1}^{K} \tilde{b}_k g_k\left(\frac{j}{T}\right) + \tilde{\varepsilon}_j\,, \qquad (j = 2, \ldots, T)\,, \tag{57}$$

$$y_j = \hat{\rho} y_{j-1} + \sum_{k=1}^{K} \hat{b}_k g_k\left(\frac{j}{T}\right) + \hat{\varepsilon}_j\,, \qquad (j = 2, \ldots, T)\,, \tag{58}$$

where $g_k(t)$ is a deterministic function equal to $\phi_k(t)$ in (30) when $c = 0$ ($\rho = 1$) and equal to $f_k(t)$ in (44) when $c > 0$ ($\rho < 1$). Note that $f_k(t)$ cannot be given explicitly so that it has to be computed numerically.

Table 2 is concerned with R^2 and DW for the model (57) with ρ=1. The entries are the means and standard deviations (SDs) of these statistics computed from 1,000 replications. We fix the number of replications at 1,000 throughout simulations. The sample sizes used here are T=400 and 800, for which six values of the number of terms K are examined. It is seen from Table 2 that the distribution of R^2 with K fixed depends little on T, as was described in Theorem 3.1, although R^2 tends to 1 as K becomes large. On the other hand the distribution of DW does depend on T even if T is large and K is fixed. Both the mean and SD decrease to half as the sample size doubles. This is because $DW = O_p(1/T)$ with K fixed. When K becomes large with large T fixed, DW increases in proportion to K because $T \times DW = O_p(K)$.

Table 2

Table 3 is concerned with the means and SDs of $T(\hat{\rho}-1)$ obtained from (58) for ρ=1, 0.975, and 0.95, respectively with the sample size $T = 400$. The entries in parentheses are the corresponding theoretical values derived from K-asymptotics described in (35) and (52). It is observed from Table 3 that the distribution depends little on ρ when K is moderately large ($K > 10$ in the present case). This last fact is a consequence of Theorem 3.4.

Table 3

Figure 2 draws the histogram of $T(\hat{\rho}-1)$ with $T=400$, $\rho=1$ and $K=1$, together with the density of N$(-4.93,\ 16.23)$ derived from K-asymptotics.The approximation is evidently poor because of a very small value of K. Figure 3 draws the same graph as in Figure 2, but, with $\rho=1$ and $K=20$, where the density is N$(-98.7,\ 324.7)$. It is seen that the approximation is fairly good.

Figure 2 Figure 3

Figure 4 draws the histogram with ρ=0.95 and K=20 using $\boldsymbol{f}(K,t)$ as regressors, whereas Figure 5 draws the same histogram using $\boldsymbol{\phi}(K,t)$ as regressors. The densities are N$(-98.7,\ 324.7)$ in both figures. It is seen that these two figures are much alike, as was mentioned earlier. Moreover, these figures are quite similar to Figure 3, which implies that the distribution of $T(\hat{\rho}-1)$ depends little on ρ close to 1 when K is reasonably large. This means that the unit root test loses its local power in K-asymptotics.

Figure 4 Figure 5

4. Concluding Remarks

We have discussed two topics to which Professor Phillips greatly contributed. The first topic was concerned with applications of the martingale approximation or the BN decomposition. It proved to be very useful when we deal with weak convergence to the matrix-valued Ito integral and derive the asymptotic distribution of Fourier transforms of stationary processes. The complex unit root distribution was also discussed as a by-product.

The other topic was somewhat philosophical. We considered K-asymptotics in near-integrated as well as integrated processes, where K, the number of regressors of deterministic trends of trigonometric functions, increases after the sample size T goes to infinity. The results obtained may be summarized into three respects as follows:

i) The vector of deterministic trends $\boldsymbol{g}(K,t) = (g_1(t),\ \ldots,\ g_K(t))'$ tends to explain fully the true process $\{y_j\}$ that contains purely stochastic trends in the sense that the regression of y_j on $\boldsymbol{g}(K,t)$ yields significant t-ratios for the fitted coefficients, R^2 close to 1, and DW exhibiting little indication of serial correlation. The situation remains unchanged between the integrated and near-integrated processes.

ii) The deterministic trends are still significant if y_j is regressed on $\boldsymbol{g}(K,t)$ in addition to y_{j-1}, although the true process for y_j is purely integrated.

iii) The unit root test based on the regression of y_j on y_{j-1} and $\boldsymbol{g}(K,t)$ loses its local power against near-integration since the unit root distribution in the integrated process is the same as that in the near-integrated process.

Needless to say, the model with stochastic trends is preferred to the one with deterministic trends on the ground of parsimony. The truth, however, may be that the actual process is generated by an infinite number of deterministic trends with random coefficients. There is no way of choosing between the two in K-asymptotics of the integrated and near-integrated processes. This raises a question of what the trend is, which arises just because the trend, if any, is unobservable. More recent papers by Phillips (2005, 2006) also discuss the trending problem in other contexts.

References

Ahtola, J. & G. C. Tiao (1987) Distributions of least squares estimators of autoregressive parameters for a process with complex roots on the unit circle. *Journal of Time Series Analysis* 8, 1-14.

Anderson, T. W. (1971) *The Statistical Analysis of Time Series.* New York: Wiley.

Beveridge, S. & C. R. Nelson (1981) A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cysle'. *Journal of Monetary Economics* 7, 151-174.

Bierens, H. J. (2001) Complex unit roots and business cycles: Are they real? *Econometric Theory* 17, 962-983.

Chan, N. H. & C. Z. Wei (1988) Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics* 16, 367-401.

Csörgő, K & L. Horváth (1993) *Weighted Approximations in Probability and Statistics.* New York: Wiley.

Gordin, M. I. (1969) The central limit theorem for stationary processes. *Soviet Mathematics Doklady* 10, 1174-1176.

Hall, P. & C. C. Heyde (1980) *Martingale Limit Theory and Its Application.* New York: Academic Press.

Helland, I. S. (1982) Central limit theorems for martingales with discrete or continuous time. *Scandinavian Journal of Statistics* 9, 79-94.

Hochstadt, H. (1973) *Integral Equations*, New York: Wiley.

Loève, M. (1978) *Probability Theory II*, Fourth Edition, New York: Springer-Verlag.

Nabeya, S. (2001) Unit root seasonal autoregressive models with a polynomial trend of higher degree. *Econometric Theory* 17, 357-385.

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

Phillips, P.C.B. (1988) Weak convergence of sample covariance matrices to stochastic integrals via martingale approximations. *Econometric Theory* 4, 528-533.

Phillips, P.C.B. (1998) New tools for understanding spurious regressions. *Econometrica* 66, 1299-1325.

Phillips, P.C.B. (1999) Discrete Fourier transforms of fractional processes. *Cowles Foundation Discussion Paper 1243.*

Phillips, P.C.B. (2002) New unit root asymptotics in the presence of deterministic trends. *Journal of Econometrics* 111, 323-353.

Phillips, P.C.B. (2005) HAC estimation by automated regression, *Econometric Theory* 21, 116-142.

Phillips, P.C.B. (2006) Optimal estimation of cointegrated systems with irrelevant instruments, *Cowles Foundation Discussion Paper 1547.*

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

Said, E. S. & D. A. Dickey (1984) Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71, 599-607.

Tanaka, K. (1996) *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*, New York: Wiley.

Tanaka, K. (2001) K-asymptotics associated with deterministic trends in the integrated and near-integrated processes. *The Japanese Economic Review* 52, 35-63.

Tanaka, K. (2008) Analysis of models with complex roots on the unit circle. *Journal of the Japan Statistical Society* 38, 145-155.

Table 2. R^2 and DW Statistics in (57) with $\rho = 1$

	$K=1$	2	5	10	20	50
			R^2			
$T = 400$						
Mean	0.594	0.752	0.887	0.943	0.971	0.989
SD	0.325	0.236	0.121	0.060	0.032	0.013
$T = 800$						
Mean	0.591	0.752	0.888	0.942	0.971	0.988
SD	0.314	0.235	0.117	0.063	0.032	0.012
			DW			
$T = 400$						
Mean	0.039	0.064	0.136	0.256	0.484	1.100
SD	0.024	0.031	0.047	0.063	0.081	0.104
$T = 800$						
Mean	0.020	0.033	0.069	0.129	0.249	0.586
SD	0.012	0.016	0.023	0.032	0.043	0.062

Table 3. Distributions of $T(\hat{\rho} - 1)$ in (58) with $g_k(t) = \phi_k(t)$ or $f_k(t)$

	$K = 1$	2	5	10	20	50
			$\rho = 1$			
Mean	−7.03	−11.91	−26.32	−50.60	−96.18	−219.83
	(−4.93)	(−9.87)	(−24.67)	(−49.35)	(−98.70)	(−246.74)
SD	5.10	6.56	9.71	12.77	16.61	21.04
	(4.03)	(5.70)	(9.01)	(12.74)	(18.02)	(28.49)
			$\rho = 0.975$			
Mean	−14.87	−18.08	−29.67	−52.36	−96.97	−220.10
SD	6.26	7.37	9.67	12.78	16.59	21.06
			$\rho = 0.95$			
Mean	−24.50	−27.08	−36.56	−56.69	−99.12	−220.78
SD	7.60	8.38	10.09	12.88	16.58	21.07

Figure 1. Complex unit root distribution: $T(\hat{\phi}_2 + 1)$

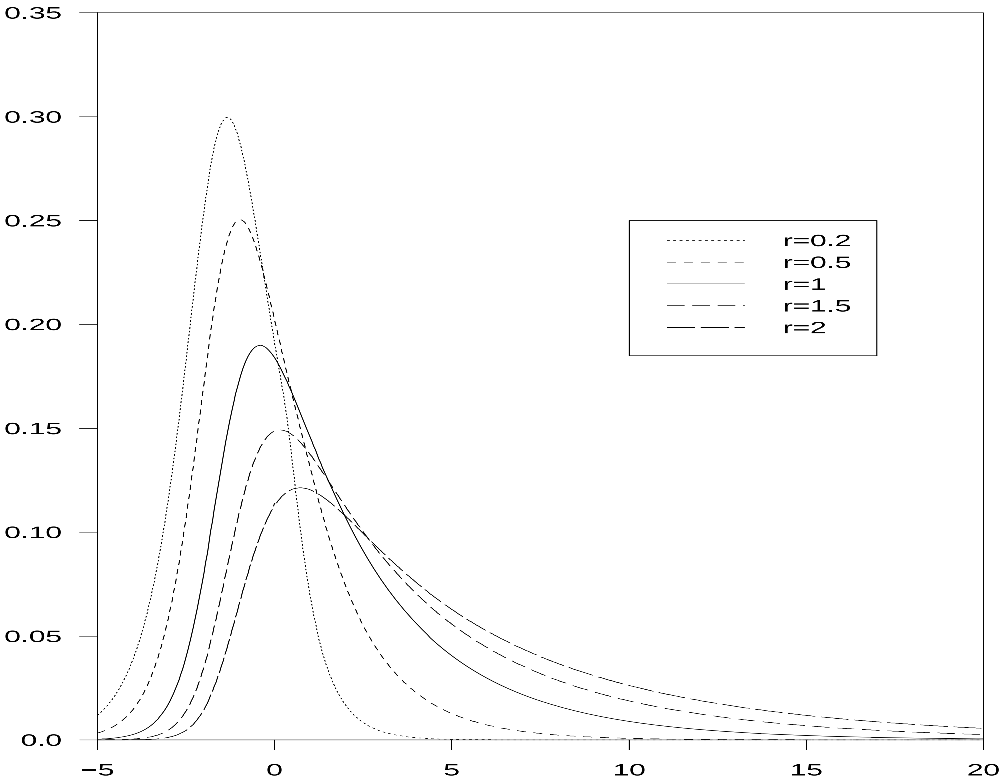

Figure 2. Distributions of $T(\hat{\rho}-1)$ with $T = 400$, $K = 1$, $\rho = 1$

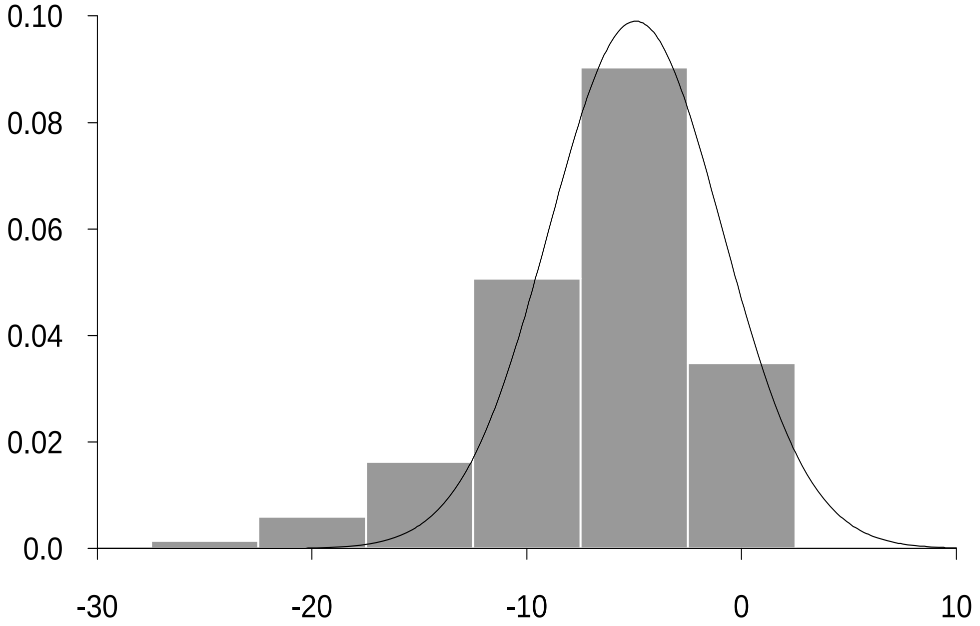

Figure 3. Distributions of $T(\hat{\rho}-1)$ with $T = 400$, $K = 20$, $\rho = 1$

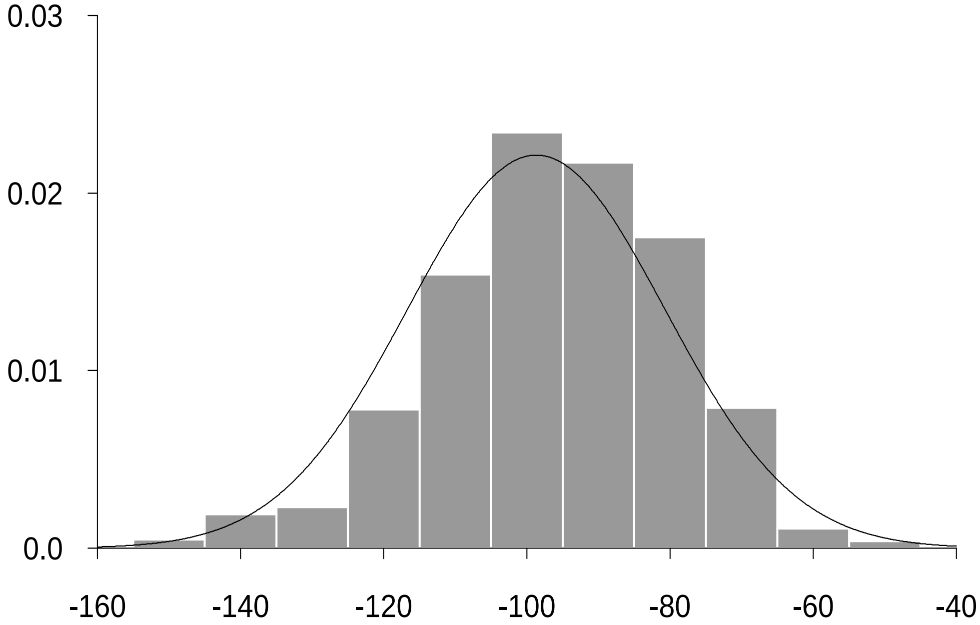

Figure 4. Distributions of $T(\hat{\rho}-1)$ with $T=400$, $K=20$, $\rho=0.95$
(Regression with $\boldsymbol{f}(K,t)$ as deterministic trends)

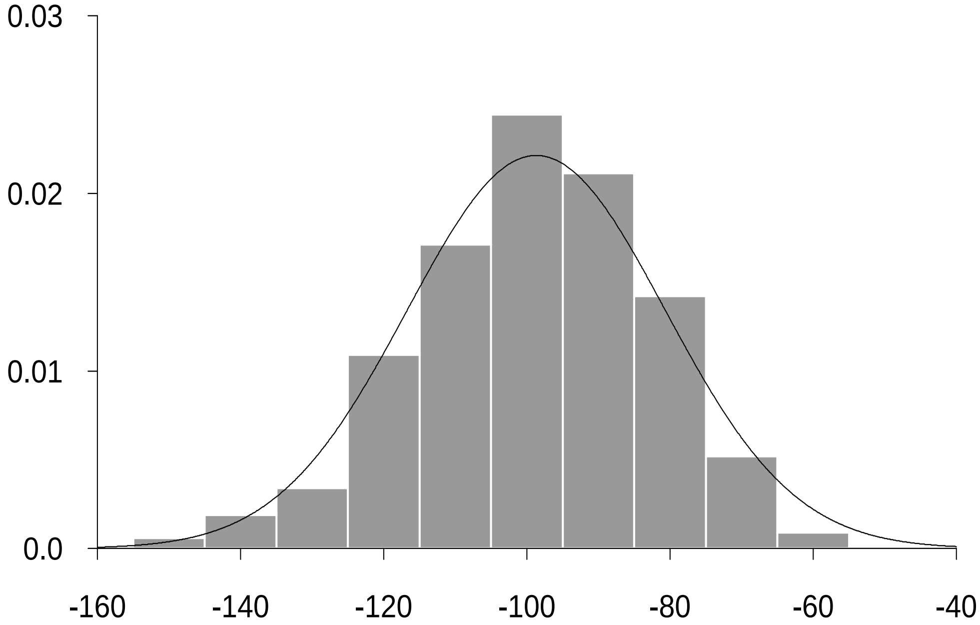

Figure 5. Distributions of $T(\hat{\rho}-1)$ with $T=400$, $K=20$, $\rho=0.95$
(Regression with $\boldsymbol{\phi}(K,t)$ as deterministic trends)