K-Asymptotics Associated with Deterministic Trends in the Integrated and Near-Integrated Processes*

Katsuto Tanaka
Hitotsubashi University

Abstract

Dealing with the integrated and near-integrated processes, this paper investigates the validity of regression on deterministic trends of K terms as K becomes large. It is found that the regression tends to be valid in spite of the true process being free from the deterministic trends, which implies that the distinction between stochastic and deterministic trends disappears in K-asymptotics of the integrated and near-integrated processes. It is also shown that, in K-asymptotics, the usual unit root test based on the model with deterministic trends of K terms becomes useless against near integration since the unit root distribution remains unchanged.

---

* This paper was prepared for the Invited Lecture at the Annual Meeting of the Japanese Economic Association held at the Osaka Prefecture University on 16-17 September, 2000. I am very grateful to Professor Seiji Nabeya for valuable suggestions on the computational aspects of simulations reported in this paper. I am also grateful to an anonymous referee for useful comments.
JEL Classification Numbers: C12, C15, C22.

1. Introduction

In the analysis of economic time series it has been taken for granted that it is possible to differentiate stochastic trends from deterministic trends, where stochastic trends refer to I(1) processes, that is, nonstationary processes whose first differences become stationary. The so-called unit root tests have been devised to discriminate between the two with the belief in this dichotomy.

There are, however, some phenomena that shake this belief. Spurious detrending in regression is one typical example, where the I(1) process is regressed on deterministic time trends to observe that the regression coefficients exhibit statistical significance as the sample size T goes to infinity, although their true values are zero, which is discussed in Durlauf and Phillips (1988). [1] In addition, R^2, the coefficient of determination, has a non-degenerate limiting distribution. The only evident indication of spurious regression seems to be a very low value of the Durbin-Watson (DW) statistic, which implies that the detrended residuals are still of the nonstationary I(1) nature.

The above findings are based on the situation where the sample size $T \to \infty$ with K, the number of terms of deterministic trends used for regression, fixed. We shall call this usual type of asymptotics "T- asymptotics". On the other hand, Phillips (1998, 1999) recently extended the arguments to "K-asymptotics", where the number of deterministic terms $K \to \infty$ maintaining T-asymptotics. In K-asymptotics the regression coefficients of deterministic trends are still significant. Moreover, R^2 converges to 1 in probability, implying that the stochastic trends can be fully explained by deterministic trends. The DW statistic is found to be of no help in detecting spurious regression. In other words, spurious regression becomes valid in K-asymptotics.

The purpose of the present paper is two-fold. One is to explore various statistical properties in K-asymptotics, extending the arguments of Phillips (1998, 1999) to near-integrated processes. The other is to investigate how K-asymptotics work in finite samples. In doing so attention is directed to the effect on the unit root tests after detrending. Section 2 describes briefly Phillips' arguments, while Section 3 extends his arguments. It seems necessary to deal with these two cases separately because of some mathematical reasons which will become clear later. Section 4 examines, by simulations, the finite-sample performance of K-asymptotics. Section 5 concludes.

[1] Spurious regression is usually referred to as the regression of one I(1) process on another independent I(1) process. This case can be detected in terms of cointegration, with which we are not concerned here.

Since the Fredholm theory on integral equations plays an important role in this paper, its brief summary is provided as Appendix A, whereas proofs of theorems are all given in Appendix B.

2. K-Asymptotics in the Integrated Process

In this section we deal with the I(1) process whose data generating process (DGP) is given by

$$y_t = y_{t-1} + u_t, \qquad y_0 = 0, \qquad (t = 1, \ldots, T), \tag{1}$$

where $\{u_t\}$ is a short memory stationary process defined by

$$u_t = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} j|\alpha_j| < \infty, \quad \alpha \equiv \sum_{j=0}^{\infty} \alpha_j \neq 0, \tag{2}$$

with $\{\varepsilon_t\} \sim$ i.i.d.$(0, \sigma^2)$. Then the process $\{u_t\}$ has the long-run variance

$$\sigma_L^2 = \lim_{T\to\infty} \frac{1}{T} V\left(\sum_{t=1}^{T} u_t\right) = \sigma^2 \left(\sum_{j=0}^{\infty} \alpha_j\right)^2, \tag{3}$$

which is finite and positive. The short run variance $V(u_t)$ is denoted as σ_S^2. We further assume that $E\left(|\varepsilon_t|^p\right) < \infty$ for some $p > 2$.

We also define the partial sum process

$$X_T(r) = \frac{1}{\sqrt{T}\sigma_L} y_{[Tr]} = \frac{1}{\sqrt{T}\sigma_L} \sum_{t=1}^{[Tr]} u_t, \tag{4}$$

where $r \in [0,1]$ and $[Tr]$ is the integer part of Tr. Then, as $T \to \infty$, the following functional central limit theorem (FCLT) holds:

$$X_T(\cdot) \Rightarrow W(\cdot), \tag{5}$$

where $\Rightarrow$ signifies weak convergence throughout this paper, whereas $\{W(r)\}$ is the standard Brownian motion defined on [0,1]. This is a typical invariance principle in the weak version, while the strong version [Csörgő and Horváth (1993)] says that, as $T \to \infty$, we can construct a standard Brownian motion such that

$$\sup_{0\le r\le 1} T^{\delta} \left| \frac{1}{\sqrt{T}\sigma_L} y_{[Tr]} - W(r) \right| \longrightarrow 0 \tag{6}$$

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

Moreover it is known [Loève (1978), Chan and Wei (1988)] that $W(r)$ admits infinitely many ways of series representations. For example, we have

$$W(r) = \sum_{i=1}^{m} g_i(r)\nu_i + \sum_{n=1}^{\infty} \frac{f_n(r)}{\sqrt{\lambda_n}}\xi_n ,$$

where $\{\nu_i\} \sim$ NID(0,1), $\{\xi_n\} \sim$ NID(0,1), and the two sequences are independent of each other, whereas $g_i(r)$ is a continuous function. Moreover, λ_n is the eigenvalue of the positive definite kernel

$$K(r,s) = \mathrm{Cov}\left(W(r) - \sum_{i=1}^{m} g_i(r)\nu_i, W(s) - \sum_{i=1}^{m} g_i(s)\nu_i\right) ,$$

and $f_n(r)$ is the corresponding orthonormal eigenfunction (see, for details, Appendix A). Thus, allowing for various values of m and various functions $g_i(r)$, we obtain infinitely many ways of series representations of $W(r)$ that converge with probability 1 and in mean square, uniformly in $r \in [0,1]$. Among such representations the most convenient for the present purpose is

$$W(r) = \sum_{n=1}^{\infty} \frac{\phi_n(r)}{(n-1/2)\pi}\xi_n , \qquad \phi_n(r) = \sqrt{2}\sin[(n-1/2)\pi r], \tag{7}$$

where $(n-1/2)\pi$ is the square root of the n-th smallest eigenvalue of the positive definite kernel $\mathrm{Cov}(W(r), W(s)) = \min(r,s)$, while $\phi_n(r)$ is the corresponding orthonormal eigenfunction.

It is anticipated from (6) and (7) that the I(1) process $y_t = y_{t-1} + u_t$ that contains purely stochastic trends can be validly represented by deterministic trends. To see this we consider, following Phillips (1998), the regression relation

$$y_t = \sum_{k=1}^{K} \tilde{b}_k \phi_k\left(\frac{t}{T}\right) + \tilde{u}_t = \tilde{b}(K)'\phi(K, t/T) + \tilde{u}_t \tag{8}$$

where $\tilde{b}(K)' = (\tilde{b}_1, \cdots, \tilde{b}_K)$, $\phi(K, t/T) = (\phi_1(t/T), \cdots, \phi_K(t/T))'$ and $\tilde{u}_t$ is the OLS residual. Note that $\phi(K, t/T)$ is the vector of deterministic trends. One may argue that use of polynomials is more appropriate as a trend. It is certainly true, but we do prefer to use $\phi(K, t/T)$ because this is also used in the series representation of the Brownian motion in (7), which yields mathematical convenience. For our purpose, however, this choice will lose no generality, as we shall demonstrate later.

To study the asymptotic behavior of various statistics arising from the regression relation (8), the partial sum process based on residuals plays an important role, which we describe as

$$\tilde{U}_T(r) = \frac{1}{\sqrt{T}\sigma_L}\,\tilde{u}_{[Tr]} \quad \Rightarrow \quad W_{\phi K}(r) = W(r) - \int_0^1 \phi(K,s)'W(s)\,ds\,\phi(K,r)\,. \qquad (9)$$

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

Let $c(K) = (c_1,\ \ldots,\ c_K)'$ be any $K \times 1$ vector with $c(K)'c(K) = 1$ and $t_{c(K)'\tilde{b}(K)}$ the usual regression t-ratio constructed from $c(K)'\tilde{b}(K)$. We also denote the coefficient of determination and the Durbin-Watson statistic by R^2 and DW, respectively. Then Phillips (1998) obtained the following results on T-asymptotics.

(a) $c(K)'\tilde{b}(K)/\sqrt{T} \quad \Rightarrow \quad \sigma_L c(K)' \int_0^1 \phi(K,r)W(r)\,dr\,,$

(b) $\sum_{t=1}^T \tilde{u}_t{}^2/T^2 \quad \Rightarrow \quad \sigma_L^2 \int_0^1 W_{\phi K}^2(r)\,dr\,,$

(c) $t_{c(K)'\tilde{b}(K)}/\sqrt{T} \quad \Rightarrow \quad c(K)' \int_0^1 \phi(K,r)W(r)\,dr / \left(\int_0^1 W_{\phi K}^2(r)\,dr\right)^{1/2}\,,$

(d) $R^2 = 1 - \sum_{t=1}^T \tilde{u}_t^2 / \sum_{t=1}^T y_t^2 \quad \Rightarrow \quad 1 - \int_0^1 W_{\phi K}^2(r)\,dr / \int_0^1 W^2(r)\,dr\,,$

(e) $DW \to 0$ in probability.

The result (a) together with (7) and the orthogonality of $\phi_n(r)$ implies that the K components of $\tilde{b}(K)/\sqrt{T}$ are asymptotically independently normally distributed with the variance of its k-th component being $\sigma_L^2/((k-1/2)^2\pi^2)$. In fact, it holds that

$$\int_0^1 \phi(K,r)W(r)\,dr = (\xi_1/a_1,\ \ldots,\ \xi_K/a_K)'\,, \quad a_k = (k-1/2)\pi\,.$$

The result (b) implies that the regression residuals still contain nonstaionary components while 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. This last statement can be made more rigorous by the fact that

$$T \times DW \quad \Rightarrow \quad \frac{1}{\int_0^1 W_{\phi K}^2(r)\,dr}\,.$$

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

(a) $c(K)'\tilde{b}(K)/\sqrt{T} \quad \Rightarrow \quad \mathrm{N}(0,\sigma_0^2), \quad \sigma_0^2 = \sigma_L^2 \sum_{n=1}^{\infty} c_n^2/\left((n-1/2)^2\pi^2\right),$

(b) $\sum_{t=1}^{T} \tilde{u}_t{}^2/T^2 = O_p(1/K),$

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

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

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

All of the above statistics signal that the regression relation (8) is valid in K-asymptotics. We have that the coefficients of deterministic trends are still significant because of (c), and the regression (8) fully captures the variation of $\{y_t\}$ 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 models for unit root tests. To this end we consider the regression relation

$$y_t = \hat{\rho} y_{t-1} + \sum_{k=1}^{K} \hat{b}_k \phi_k\left(\frac{t}{T}\right) + \hat{u}_t = \hat{\rho} y_{t-1} + \hat{b}(K)'\phi(K,t/T) + \hat{u}_t, \quad (t = 2,\ \ldots,\ T). \tag{10}$$

We first deal with T-asymptotics, for which Phillips (1999) proved that it holds that, as $T \to \infty$,

$$ADF_\rho,\ Z_\rho \quad \Rightarrow \quad \frac{\int_0^1 W_{\phi K}(r)\, dW(r)}{\int_0^1 W_{\phi K}^2(r)\, dr}, \tag{11}$$

$$ADF_t,\ Z_t \quad \Rightarrow \quad \frac{\int_0^1 W_{\phi K}(r)\, dW(r)}{\left(\int_0^1 W_{\phi K}^2(r)\, dr\right)^{1/2}}, \tag{12}$$

where ADF_ρ and Z_ρ are the unit root coefficient statistics suggested in Said and Dickey (1984) and Phillips (1987), respectively, whereas ADF_t and Z_t are the corresponding unit root t-ratio statistics. Note that, in the simplest case where the error term $\{u_t\}$ in (1) follows i.i.d.$(0,\sigma^2)$, $ADF_\rho = T(\hat{\rho}-1)$ and

$$ADF_t = \frac{\hat{\rho}-1}{\hat{\sigma} \Big/ \left(\sum y_{t-1}^2 - \sum y_{t-1}\phi(K,t/T)' A_T^{-1} \sum y_{t-1}\phi(K,t/T)\right)^{1/2}},$$

where $\hat{\sigma}^2 = \sum \hat{u}_t^2/(T-K-1)$ and $A_T = \sum \phi(K,t/T)\phi(K,t/T)'$.

The following results concerning the significance of the coefficients of deterministic trends can also be obtained by routine calculation.

$$\sqrt{T}c(K)'\hat{b}(K)/\sigma_L \quad \Rightarrow \quad c(K)'X(K,\phi)\,, \tag{13}$$

$$t_{c(K)'\hat{b}(K)} \quad \Rightarrow \quad \frac{\sigma_L}{\sigma_S}\,c(K)'X(K,\phi)\Big/\sqrt{c(K)'\Omega(K,\phi)^{-1}c(K)}\,, \tag{14}$$

where

$$X(K,\phi) = \int_0^1 \phi(K,r)dW(r) - \frac{\left(\int_0^1 W_{\phi K}(r)\,dW(r) + (1-\sigma_S^2/\sigma_L^2)/2\right)\int_0^1 \phi(K,r)W(r)\,dr}{\int_0^1 W_{\phi K}^2(r)\,dr}\,,$$

$$\Omega(K,\phi) = I_K - \int_0^1 \phi(K,r)W(r)\,dr \int_0^1 \phi(K,r)'W(r)\,dr\Big/\int_0^1 W^2(r)\,dr.$$

It is seen from (13) that the asymptotic distributions of the coefficients are non-normal, unlike in the case of the purely deterministic regression (8), and that the stochastic order of $\hat{b}(K)$ has decreased to $1/\sqrt{T}$ so that $\hat{b}(K)$ is consistent and converges to the true coefficient of zero. Nonetheless it follows from (14) that the corresponding t-ratio is $O_p(1)$, which implies possible significance of the coefficients of deterministic trends.

This last statement becomes much clearer in K-asymptotics. In fact we have the following results on K-asymptotics, which can be derived by letting $K \to \infty$ in the expressions on the right sides of (11) through (14). [2]

$$ADF_\rho,\, Z_\rho \quad \Rightarrow \quad \mathrm{N}\left(-\frac{\pi^2 K}{2},\, \frac{\pi^4 K}{6}\right)\,, \tag{15}$$

$$ADF_t,\, Z_t \quad \Rightarrow \quad \mathrm{N}\left(-\frac{\pi\sqrt{K}}{2},\, \frac{\pi^2}{24}\right)\,, \tag{16}$$

$$\sqrt{T}c(K)'\hat{b}(K)/\sigma_S \quad \Rightarrow \quad \mathrm{N}\left(0,\, \pi^2 K^2 \sum_{k=1}^{K} \frac{c_k^2}{(2k-1)^2}\right)\,, \tag{17}$$

$$t_{c(K)'\hat{b}(K)} \quad = \quad O_p(\sqrt{K})\,. \tag{18}$$

It is quite interesting to observe in (15) that the unit root coefficient statistic, which is $T(\hat{\rho}-1)$ in the simplest case, is $O_p(K)$, diverging to $-\infty$, and tends to normality after suitable recentering and rescaling. It is also interesting to notice in (16) that the t-ratio is $O_p(\sqrt{K})$, diverging to $-\infty$, and tends to normality after recentering without rescaling. It follows from (17) that $\sqrt{T}c(K)'\hat{b}(K)/K = O_p(1)$ so that the coefficients

[2] The results (17) and (18) are not given in Phillips (1999), whose proof will be provided in Section 3 when we deal with the near-integrated process.

of deterministic trends can take nonnegligible values in K-asymptotics. The result in (18) further signals the statistical significance of deterministic trends.

The limiting distributions described in (15) and (16) 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

$$y_t = \hat{\rho} y_{t-1} + \sum_{k=0}^{K} \hat{b}_k \left(\frac{t}{T}\right)^k + \hat{u}_t = \hat{\rho} y_{t-1} + \hat{b}(K)' p(K, t/T) + \hat{u}_t, \tag{19}$$

where $p(K, r) = (1,\ r,\ ..,\ r^K)'$, Nabeya (1999) obtained the following K-asymptotics.[3]

$$ADF_\rho,\ Z_\rho \quad \Rightarrow \quad \mathrm{N}(-4K,\ 16K)\,, \tag{20}$$

$$ADF_t,\ Z_t \quad \Rightarrow \quad \mathrm{N}\left(-\sqrt{2K},\ 1/2\right)\,. \tag{21}$$

Note that (20) and (21) are polynomial versions of (15) and (16), respectively. Since

$$\mathrm{N}\left(-\frac{\pi^2 K}{2},\ \frac{\pi^4 K}{6}\right) = \mathrm{N}(-4.93K,\ 16.23K)\,, \quad \mathrm{N}\left(-\frac{\pi\sqrt{K}}{2},\ \frac{\pi^2}{24}\right) = \mathrm{N}(-1.57\sqrt{K},\ 0.41)\,,$$

it is seen that the limiting distributions based on trigonometric and polynomial functions are close to each other. This fact will be corroborated by simulations in Section 4. In any case, use of trigonometric functions as deterministic trends has been justified.

3. K-Asymptotics in the Near-Integrated Process

In this section we extend the arguments discussed in the last section to deal with near-integrated processes. Thus we consider, as the DGP,

$$y_t = \rho y_{t-1} + u_t, \qquad y_0 = 0, \qquad \rho = 1 - (c/T), \qquad (t = 1,\ \ldots,\ T), \tag{22}$$

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

$$Y_T(r) = \frac{1}{\sqrt{T}\sigma_L} y_{[Tr]} = \frac{1}{\sqrt{T}\sigma_L} \sum_{t=1}^{[Tr]} \rho^{[Tr]-t} u_t\,, \tag{23}$$

the following FCLT holds:

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

[3] An extension of the analysis of the polynomial case to the near-integrated process seems difficult and remains to be solved. This is a main reason we concentrate on deterministic trends of trigonometric functions.

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

$$J^c(r) = e^{-cr} \int_0^r e^{cs}\, dW(s) \quad \Leftrightarrow \quad dJ^c(r) = -cJ^c(r)\, dr + dW(r),\ Y(0) = 0\,. \tag{24}$$

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

$$J^c(r) = \sum_{n=1}^{\infty} \frac{f_n(r)}{\sqrt{\lambda_n}}\, \xi_n\,, \tag{25}$$

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

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

and $f_n(r)$ is the corresponding orthonormal eigenfunction. Unlike in the expansion (7) of the standand Brownian motion $W(r)$, it is impossible to obtain λ_n and $f_n(r)$ analytically, although numerically possible once c is given, which will be demonstrated in Section 4. We can show (see Appendix A) 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

$$(n - 1/2)\pi < \sqrt{\lambda_n - c^2} < n\pi \quad \text{and} \quad (n-1/2)^2\pi^2 + c^2 < \lambda_n < n^2\pi^2 + c^2\,.$$

We also obtain

$$f_n(r) = \frac{\sin \mu_n r}{M_n}, \quad \mu_n = \sqrt{\lambda_n - c^2}, \quad M_n = \sqrt{\frac{1}{2} - \frac{\sin 2\mu_n}{4\mu_n}}. \tag{26}$$

Under the above setting we first consider

$$y_t = \sum_{k=1}^{K} \tilde{b}_k f_k\left(\frac{t}{T}\right) + \tilde{u}_t = \tilde{b}(K)' f(K, t/T) + \tilde{u}_t\,, \qquad (t = 1,\ \ldots,\ T), \tag{27}$$

where $f(K, r) = (f_1(r),\ \ldots,\ f_K(r))'$ and the partial sum process based on the OLS residulas $\tilde{u}_1,\ \ldots,\ \tilde{u}_T$ satisfies

$$\frac{1}{\sqrt{T}\sigma_L}\, \tilde{u}_{[Tr]} \quad \Rightarrow \quad J^c_{fK}(r) = J^c(r) - \int_0^1 f(K, s)' J^c(s)\, ds\, f(K, r)\,. \tag{28}$$

Then, proceeding in the same way as before, we have the following theorem on T-asymptotics.

Theorem 1. *For the regression relation* (27) *it holds that, as* $T \to \infty$,

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

(b) $\sum_{t=1}^T \tilde{u}_t{}^2/T^2 \quad \Rightarrow \quad \sigma_L^2 \int_0^1 \{J_{fK}^c(r)\}^2\, dr$,

(c) $t_{c(K)'\tilde{b}(K)}/\sqrt{T} \quad \Rightarrow \quad c(K)' \int_0^1 f(K,r) J^c(r)\, dr / \left(\int_0^1 \{J_{fK}^c(r)\}^2\, dr\right)^{1/2}$,

(d) $R^2 = 1 - \sum_{t=1}^T \tilde{u}_t^2 / \sum_{t=1}^T y_t^2 \quad \Rightarrow \quad 1 - \int_0^1 \{J_{fK}^c(r)\}^2\, dr / \int_0^1 \{J^c(r)\}^2\, dr$,

(e) $DW \to 0$ in probability,

where $c(K)$ is any $K \times 1$ vector such that $c(K)'c(K) = 1$.

Theorem 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 $J^c(r)$ and the eigenfunction $f(K,r)$ have been substituted for $W(r)$ and $\phi(K,r)$, respectively. Then it follows from (25) and the orthonormality of $\{f_n(r)\}$ that

$$\int_0^1 f(K,r) J^c(r)\, dr = \left(\xi_1/\sqrt{\lambda_1},\ \ldots,\ \xi_K/\sqrt{\lambda_K}\right)',$$

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

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

We now discuss K-asymptotics by letting $K \to \infty$ in Theorem 1, which yields

Theorem 2. *For the regression relation* (27)*, it holds that, as $T \to \infty$ and then $K \to \infty$,*

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

(b) $\sum_{t=1}^T \tilde{u}_t{}^2/T^2 = O_p(1/K)$,

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

(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 $f(K,r)$ is replaced by $\phi(K,r)$ in the regression equation (27), the above result (a) has to be changed. This is because, in that case, we have, as $T \to \infty$,

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

so that this last sum need not be $\xi_k/\sqrt{\lambda_k}$, which is to be attained when $\phi_k(r)$ is replaced by $f_k(r)$. This means that the components of $\tilde{b}(K)$ 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

$$
y_t = \hat{\rho} y_{t-1} + \sum_{k=1}^{K} \hat{b}_k f_k\left(\frac{t}{T}\right) + \hat{u}_t = \hat{\rho} y_{t-1} + \hat{b}(K)' f(K, t/T) + \hat{u}_t, \tag{29}
$$

and obtain the following results on T-asymptotics (Theorem 3) and K-asymptotics (Theorem 4).

Theorem 3. *For the regression relation* (29) *it holds that, as* $T \to \infty$,

$$
ADF_\rho,\ Z_\rho \quad \Rightarrow \quad \frac{\int_0^1 J_{fK}^c(r)\, dJ^c(r)}{\int_0^1 \{J_{fK}^c(r)\}^2\, dr}, \tag{30}
$$

$$
ADF_t,\ Z_t \quad \Rightarrow \quad \frac{\int_0^1 J_{fK}^c(r)\, dJ^c(r)}{\left(\int_0^1 \{J_{fK}^c(r)\}^2\, dr\right)^{1/2}}, \tag{31}
$$

$$
\sqrt{T} c(K)' \hat{b}_K / \sigma_L \quad \Rightarrow \quad c(K)' Y(K, f), \tag{32}
$$

$$
t_{c(K)'\hat{b}(K)} \quad \Rightarrow \quad c(K)' Y(K, f) \Big/ \sqrt{c(K)' \Sigma(K, f)^{-1} c(K)}, \tag{33}
$$

where

$$
Y(K,f) \quad = \quad \int_0^1 f(K,r)\, dW(r) - \left(\frac{\int_0^1 J_{fK}^c(r)\, dJ^c(r) + (1-\sigma_S^2/\sigma_L^2)/2}{\int_0^1 \{J_{fk}^c(r)\}^2\, dr} + c\right)
$$

$$\times \int_0^1 f(K,r) J^c(r)\, dr\,,$$

$$\Sigma(K,f) \;=\; I_K - \int_0^1 f(K,r) J^c(r)\, dr \int_0^1 f(K,r)' J^c(r)\, dr \Big/ \int_0^1 \{J^c(r)\}^2\, dr.$$

Theorem 4. *For the regression relation* (29) *it holds that, as* $T \to \infty$ *and then* $K \to \infty$,

$$ADF_\rho,\, Z_\rho \;\Rightarrow\; \mathrm{N}\left(-\frac{\pi^2 K}{2},\, \frac{\pi^4 K}{6}\right), \tag{34}$$

$$ADF_t,\, Z_t \;\Rightarrow\; \mathrm{N}\left(-\frac{\pi\sqrt{K}}{2},\, \frac{\pi^2}{24}\right), \tag{35}$$

$$\sqrt{T} c(K)'\hat{b}(K)/\sigma_S \;\Rightarrow\; \mathrm{N}\left(0,\, \frac{\pi^4}{4} K^2 \sum_{k=1}^{K} \frac{c_k^2}{\lambda_k}\right), \tag{36}$$

$$t_{c(K)'\hat{b}(K)} \;=\; O_p(\sqrt{K})\,. \tag{37}$$

It is seen that the T- and K-asymptotics in the near-integrated process are essentially the same as in the integrated process. Thus we can conclude that the regression of y_t on y_{t-1} and the deterministic trend $f(K, t/T)$ is valid even if $K \to \infty$. 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 (29) cannot be used as a model for tesing a unit root $H_0 : \rho = 1$ because the deterministic regressor $f(K, r)$ depends on the unknown parameter c. We should use the modol (10) discussed in Section 2 as a suitable model for this purpose, where we 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 (10) under H_1, that is, under the situation where the DGP is the near-integrated process (22). In that case, results on T-asymptotics can be obtained in the same way as in Theorem 3 by replacing $f_k(r)$ by $\phi_k(r)$. For instance, we have, as $T \to \infty$ under $\rho = 1 - (c/T)$,

$$ADF_\rho \;\Rightarrow\; \frac{\int_0^1 J^c_{\phi K}\, dJ^c(r)}{\int_0^1 \{J^c_{\phi K}(r)\}^2\, dr}\,.$$

It, however, turns out that results on K-asymptotics in the present case are not clear-cut because of the reason described below Theorem 2. We also mentioned there that replacing $f_k(r)$ by $\phi_k(r)$ affects K-asymptotics little. Thus it is expected that results similar to those in Theorem 4 hold true in this case. This last point will be examined by simulations in the next section.

4. Some Simulations

In this section we examine, by simulations, the finite sample performance of T- and K-asymptotics developed in previous sections. For simplicity we assume the DGP to be

$$y_t = \rho y_{t-1} + \varepsilon_t, \qquad y_0 = 0, \qquad \rho = 1 - (c/T), \qquad (t = 1, \ldots, T), \tag{38}$$

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

The regression relations considered are

$$y_t = \sum_{k=1}^{K} \tilde{b}_k g_k\left(\frac{t}{T}\right) + \tilde{\varepsilon}_t\,, \qquad (t = 2, \ldots, T)\,, \tag{39}$$

$$y_t = \hat{\rho} y_{t-1} + \sum_{k=1}^{K} \hat{b}_k g_k\left(\frac{t}{T}\right) + \hat{\varepsilon}_t\,, \qquad (t = 2, \ldots, T)\,, \tag{40}$$

where $g_k(r)$ is a deterministic function equal to $\phi_k(r)$ in (7) when $c = 0$ ($\rho = 1$) and equal to $f_k(r)$ in (26) when $c > 0$ ($\rho < 1$). Note that $f_k(r)$ cannot be defined explicitly so that it has to be obtained numerically (see (26)).

To compute $\hat{\rho}$ in a number of replications it is more convenient to run the regression

$$y_{t-1} = \sum_{k=1}^{K} \breve{b}_k g_k\left(\frac{t}{T}\right) + \breve{\varepsilon}_{t-1}, \qquad (t = 2, \ldots, T)\,, \tag{41}$$

to get

$$\hat{\rho} = \frac{\sum_{t=2}^{T} \breve{\varepsilon}_{t-1} \tilde{\varepsilon}_t}{\sum_{t=2}^{T} \breve{\varepsilon}_{t-1}^2} = \frac{\sum_{t=2}^{T} \breve{\varepsilon}_{t-1} y_t}{\sum_{t=2}^{T} \breve{\varepsilon}_{t-1}^2}\,. \tag{42}$$

We can also consider the estimator

$$\tilde{\rho} = \frac{\sum_{t=2}^{T} \tilde{\varepsilon}_{t-1} \tilde{\varepsilon}_t}{\sum_{t=2}^{T} \tilde{\varepsilon}_{t-1}^2}\,, \tag{43}$$

but we find that the two estimators $\hat{\rho}$ and $\tilde{\rho}$ are different in T-asymptotics, although they are equivalent in K-asymptotics. We state this fact as the following theorem.

Theorem 5. *For the DGP* (38), *it holds that, as* $T \to \infty$,

$$T(\hat{\rho} - 1) \quad \Rightarrow \quad \frac{\int_0^1 Y_{gK}(r)\, dY(r)}{\int_0^1 \{Y_{gK}(r)\}^2\, dr}\,, \tag{44}$$

$$T(\tilde{\rho}-1) \quad \Rightarrow \quad \frac{\int_0^1 Y_{gK}(r)\,dY(r)+H_{gK}}{\int_0^1 \{Y_{gK}(r)\}^2\,dr}\,, \tag{45}$$

and, as $K \to \infty$*,*

$$T(\hat{\rho}-1)\,, \quad T(\tilde{\rho}-1) \quad \Rightarrow \quad \mathrm{N}\left(-\frac{\pi^2 K}{2}, \frac{\pi^4 K}{6}\right), \tag{46}$$

where $Y(r) = W(r)$ *for* $c=0$ *and* $J^c(r)$ *for* $c>0$*, whereas* $Y_{gK}(r) = W_{\phi K}(r)$ *for* $c=0$ *and* $J^c_{fK}(r)$ *for* $c>0$*. Moreover*

$$H_{gK} = \int_0^1 Y(r)\left[g(r)'\int_0^1 g(s)g^{(1)}(s)'ds - g^{(1)}(r)'\right] dr \int_0^1 Y(r)g(r)dr\,,$$

where $g(r) = g(K,r)$ *and* $g^{(1)}(r) = \partial g(K,r)/\partial r$.

The estimator $\tilde{\rho}$ can be easily computed once the regression (39) is fitted, but, because of the reason described in Theorem 5, we use $\hat{\rho}$ throughout simulations.[4]

Table 1 is concerned with R^2 and DW for the model (39) 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 1 that the distribution of R^2 with K fixed depends little on T, as was described in Theorem 1, while R^2 tends to 1 as K becomes large, as Theorem 2 implies. 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 as well as R^2 increases so that the fitted model, although deviates from the DGP, becomes more plausible.

TABLE 1

Table 2 is concerned with the means and SDs of $T(\hat{\rho}-1)$ obtained from T=400 for ρ=1, 0.975, and 0.95, respectively. The entries in parentheses are the corresponding theoretical values derived from K-asymptotics described in (15). It is observed from Table 2 that, for each ρ, the distribution of $T(\hat{\rho}-1)$ is shifted to the left with the SDs increased as K becomes large. It is also observed that the distribution changes

[4] If we use polynomials as the deterministic trends, then it can be shown that the two estimators are equivalent in both T- and K-asymptotics.

litte with ρ close to 1 when K is moderately large ($K > 10$ in the present case). This last fact is a consequence of Theorem 4. Under T=400, Figure 1 draws the histogram of $T(\hat{\rho}-1)$ with ρ=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, although not terribly bad. Figures 2 and 3 draw the same graphs as in Figure 1, but, with ρ=1 and K=20 in the former, and with ρ=0.95 and K=20 in the latter. The normal distributions are N(-98.7, 324.7) in both figures. It is seen that the approximation is quite good in both cases. It is also seen that the two histograms are quite alike, which implies that the distribution of $T(\hat{\rho}-1)$ depends little on ρ close to 1 when K is reasonably large.

TABLE 2 FIGURE 1 FIGURE 2 FIGURE 3

We also conduct simulations in connection with the unit root test

$$H_0 : \rho = 1 \qquad \text{against} \qquad H_1 : \rho = 1 - (c/T)\,. \tag{47}$$

For this purpose we use the model (40) with $g_k(r) = \phi_k(r)$. The null distribution of the unit root statistic $T(\hat{\rho}-1)$ was discussed in Section 2, and is drawn in Figures 1 and 2. As for the distribution under H_1, we should distinguish it from the distribution with $g_k(r) = f_k(r)$ discussed in Section 3. Note that we cannot use $f_k(r)$ in stead of $\phi_k(r)$ because $f_k(r)$ depends on unknown c to be tested. One such histogram with $g_k(r) = f_k(r)$ was drawn in Figure 3. Figure 4 draws the histogram of the unit root statistic $T(\hat{\rho}-1)$ under c=20 when T=400 and K=20 so that ρ=0.95, together with N(-98.7,324.7) which is the distribution of the statistic in K-asymptotics with $\phi_k(r)$ replaced by $f_k(r)$. It is seen that the histogram in this figure is so close to that in Figure 3, which implies that the distribution of $T(\hat{\rho}-1)$ under ρ close to 1 depends little on whether $\phi_k(r)$ or $f_k(r)$ is used. Comparison of Figure 4 with Figure 2 leads us to conclude that the unit root test based on $T(\hat{\rho}-1)$ has no local power in K-asymptotics. The same is true, though not reported here, for the test based on the t-ratio statistic.

FIGURE 4

Finally we report some simulation results associated with the usual model for the unit root test, that is, the model (40) with $g_k(r)$ being the polynomial of degree k.

It should be mentioned here that use of ordinary polynomials of the form $g_k(r) = r^k$ causes trouble to compute $\hat{\rho}$ when K becomes as large as 10. This fact may be explained in terms of the condition number of the $K \times K$ matrix H_{KT} whose (j,k)th element $H_{KT}(j,k)$ is given by

$$H_{KT}(j,k) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^j \left(\frac{t}{T}\right)^k .$$

It holds that, as $T \to \infty$, $H_{KT}(j,k)$ converges to $H_K(j,k) = 1/(j+k+1)$. The corresponding limiting matrix H_K is known as the Hilbert matrix and it is found that the condition numbers of H_K are

1.24×10^5 $(K=4)$, 1.67×10^8 $(K=6)$, 2.05×10^{11} $(K=8)$, 2.42×10^{14} $(K=10)$.

To overcome the above difficulty we used Legendre's polynomials whose orthonormalized version of degree k is defined by

$$Q_k(r) = \sqrt{2k+1}\sum_{j=0}^{[k/2]}(-1)^j \frac{(2k-2j)!}{2^{2j}\, j!\,(k-j)!\,(k-2j)!}\left(r-\frac{1}{2}\right)^{k-2j} ,$$

where $Q_0(r) = 1$, $Q_1(r) = \sqrt{3}\,(2r-1)$, $Q_2(r) = \sqrt{5}\,(6r^2-6r+1)$, and so on. [5] It should be understood that the distribution of $T(\hat{\rho}-1)$ does not depend on the choice of polynomials, unlike in the case of trigonometric functions. Table 3 reports the means and SDs under T=400 for c=0 (ρ=1), 10 (ρ=0.975), 20 (ρ=0.95), and various values of K. The entries in parentheses are the corresponding theoretical values derived from K-asymptotics described in (20). The general feature is similar to that of Table 2. We should mention here that K-asymptotics in the near-integrated process remain to be established when polynomials are used as the deterministic trends, but it is conjectured that the distribution of $T(\hat{\rho}-1)$ depends little on ρ close to 1 as K becomes large with large T fixed. Figure 5 draws the histogram of $T(\hat{\rho}-1)$ with ρ=1 and K=20 under T=400, together with the density of N(-80, 17.89) derived from K-asymptotics, whereas Figure 6 the corresponding histogram with ρ=0.95 and K=20, together with the same density as in Figure 5. It is seen that the two histograms resemble closely, which supports the conjecture mentioned above. The normal approximation is also seen to be reasonably good. We can conclude that

[5] I am very grateful to Professor Seiji Nabeya for suggesting using and how to compute efficiently Legendre's polynomials.

the unit root test based on $T(\hat{\rho}-1)$ has also no local power in K-asymptotics when polynomials are used as the deterministic trends.

TABLE 3 FIGURE 5 FIGURE 6

5. Concluding Remarks

We have discussed K-asymptotics associated with deterministic trends fitted to the integrated and near-intgerated processes. The results obtained can be summarized into three respects as follows:

i) The vector of deterministic trends, $g(K,r) = (g_1(r), \ \ldots, \ g_K(r))'$, tends to explain fully the true process $\{y_t\}$ that contains purely stochastic trends in the sense that the regression of y_t on $g(K,r)$ 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_t is regressed on $g(K,r)$ as well as on y_{t-1}, although the DGP for y_t contains y_{t-1} only.

iii) The unit root test based on the regression of y_t on y_{t-1} and $g(K,r)$ loses its 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 under 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.

REFERENCES

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

Courant, R. and D. Hilbert (1953) *Methods of Mathematical Physics*, Vol. I, New York: Wiley.

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

Durlauf, S. N. and P. C. B. Phillips (1988) "Trends versus Random Walks in Time Series Analysis", *Econometrica*, Vol. 56, pp. 1333-1354.

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

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

Nabeya, S. (1999) "Unit Root Seasonal Autoregressive Models with a Polynomial Trend of Higher Degree", To appear in *Econometric Theory.*

—— and K. Tanaka (1988) "Asymptotic Theory of a Test for the Constancy of Regression Coefficients against The Random Walk Alternative", *Annals of Statistics*, Vol. 16, pp. 218-235.

—— and —— (1990a) "A General Approach to the Limiting Distribution for Estimators in Time Series Regression with Nonstable Autoregressive Errors", *Econometrica*, Vol. 58, pp. 145-163.

—— and —— (1990b) "Limiting Powers of Unit-Root Tests in Time-Series Regression", *Journal of Econometrics*, Vol. 46, pp. 247-271.

Phillips, P. C. B. (1987) "Time Series Regression with a Unit Root", *Econometrica*, Vol. 55, pp. 277-301.

—— (1998) "New Tools for Understanding Spurious Regressions", *Econometrica*, Vol. 66, pp. 1299-1325.

—— (1999) “New Unit Root Asymptotics in the Presence of Deterministic Trends”, Discussion Paper, Cowles Foundation for Research in Economics, Yale University.

Said, S. E. and D. A. Dickey (1984) “Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order”, *Biometrika*, Vol. 71, pp. 599-608.

Tanaka, K. (1990) “The Fredholm Approach to Asymptotic Inference on Nonstationary and Noninvertible Time Series Models”, *Econometric Theory*, Vol. 6, pp. 411-432.

—— (1993) “An Alternative Approach to the Asymptotic Theory of Spurious Regression, Cointegration, and Near Cointegration”, *Econometric Theory*, Vol. 9, pp. 36-61.

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

Whittaker, E. T. and G. N. Watson (1958) *A Course of Modern Analysis*, Fourth Edition, London: Cambridge University Press.

Appendix A: Brief Summary of the Fredholm Theory

Following Courant and Hilbert (1953) and Hochstadt (1973), we give a brief summary of the theory of Fredholm integral equations, whose knowledge is essential to the understanding of the discussions developed in Sections 2 and 3. A more detailed treatment of the theory in terms of the analysis of nonstationary time series can be found in Tanaka (1996), while various statistical applications are developed in Nabeya and Tanaka (1988, 1990a, 1990b), and Tanaka (1990, 1993).

Let us consider the following equation for λ and $f(r)$

$$f(r) = \lambda \int_0^1 K(r,s) f(s)\, ds\,, \tag{A1}$$

where $K(r,s)$, called the kernel, is a given, continuous and symmetric function on $[0,1]\times[0,1]$. This equation is called the homogeneous Fredholm integral equation of the second kind. A value λ for which this integral equation possesses a nonvanishing continuous solution $f(r)$ is called an eigenvalue of the kernel $K(r,s)$; the corresponding solution $f(r)$ is called an eigenfunction for the eigenvalue λ. The maximum number l of linearly independent solutions is called the multiplicity of λ. It is known that every eigenvalue is real because of the symmetry of the kernel and every multiplicity is finite. Moreover, we can assume, without any loss of generality, that the sequence of eigenfunctions $\{f_n(r)\}$ is orthonormal, where the eigenvalues are repeated as many times as their multiplicities.

The integral equation (A1) can be approximated by the algebraic system

$$f\left(\frac{j}{T}\right) = \frac{\lambda}{T}\sum_{k=1}^{T} K\left(\frac{j}{T},\frac{k}{T}\right) f\left(\frac{k}{T}\right), \quad (j = 1,\ \ldots,\ T)\,,$$

or, in matrix notation,

$$f_T = \frac{\lambda}{T} K_T\, f_T\,,$$

where $f_T = [(f(j/T))]$ is a $T\times 1$ vector and $K_T = [(K(j/T,k/T))]$ is a $T\times T$ matrix.

Let us consider

$$D_T(\lambda) = \left| I_T - \frac{\lambda}{T} K_T \right|\,,$$

where $D_T(\lambda) = 0$ is the characteristic equation that gives reciprocals of eigenvalues of the matrix K_T/T. Then it holds that

$$D(\lambda) = \lim_{T\to\infty} D_T(\lambda)$$
$$= \sum_{n=0}^{\infty} \frac{(-1)^n \lambda^n}{n!} \int_0^1 \cdots \int_0^1 \begin{vmatrix} K(t_1,t_1) & \cdot & \cdot & \cdot & K(t_n,t_n) \\ \vdots & & & & \vdots \\ K(t_l,t_1) & \cdot & \cdot & \cdot & K(t_n,t_n) \end{vmatrix} dt_1 \cdots dt_n . \qquad (A2)$$

The function $D(\lambda)$ is called the Fredholm determinant (FD) of the kernel $K(r,s)$. It holds that the series in (A2) converges for all λ so that $D(\lambda)$ is an entire or integral function with $D(0) = 1$.

The following theorem gives an important relationship between $D(\lambda)$ and eigenvalues.

Theorem A1. *Every solution to $D(\lambda) = 0$ is an eigenvalue of $K(r,s)$, and in turn every eigenvalue of $K(r,s)$ is a solution to $D(\lambda) = 0$.*

It follows that $\lambda = 0$ is never an eigenvalue since $D(0) = 1 \neq 0$. It is known that there exists at least one eigenvalue insofar as the symmmetric and continuous kernel $K(r,s)$ is not identically equal to zero. If there are an infinite number of eigenvalues, $K(r,s)$ is said to be nondegenerate; otherwise it is degenerate. When $K(r,s)$ is nondegenerate, $\lambda = \infty$ is the only accumulation point of zeros. If all the eigenvalues have the same sign, then $K(r,s)$ is said to be definite. Alternatively, $K(r,s)$ is positive (negative) definite if $\int_0^1 \int_0^1 K(r,s)g(r)g(s)\,dr\,ds$ is nonnegative (nonpositive) for any continuous function $g(r)$ on [0,1]. If all but a finite number of eigenvalues have the same sign, $K(r,s)$ is said to be nearly definite.

Two typical examples of kernels are

$$\mathrm{Cov}(W(r), W(s)) = \min(r,s), \quad \mathrm{Cov}(\tilde{W}(r), \tilde{W}(s)) = \min(r,s) - rs,$$

where $W(r)$ is the standard Brownian motion and $\tilde{W}(r)$ is the Brownian bridge. These kernels are evidently positive definite and nondegenerate. In fact, the FDs of $\min(r,s)$ and $\min(r,s) - rs$ are shown to be $\cos\sqrt{\lambda}$ and $\sin\sqrt{\lambda}/\sqrt{\lambda}$, respectively. Thus the eigenvalues are $(n-1/2)^2\pi^2$ and $n^2\pi^2$ for $n = 1, \ldots,$ respectively. A convenient method for computing FDs is demonstrated below to help understand the arguments in Section 3.

The following theorem, called Mercer's theorem, is specific to nearly definite kernels, and is an infinite-dimensional version of the eigenvalue decomposition of symmetric matrices.

Theorem A2. *Let $K(r,s)$ be continuous, symmetric and nearly definite on* $[0,1] \times [0,1]$. *Then it holds that*

$$K(r,s) = \sum_{n=1}^{\infty} \frac{1}{\lambda_n} f_n(r) f_n(s) ,$$

where $\{\lambda_n\}$ is a sequence of eigenvalues of $K(r,s)$ repeated as many times as their multiplicities, while $\{f_n(r)\}$ is an orthonormal sequence of eigenfunctions corresponding to eigenvalues λ_n and the series on the right side converges absolutely and uniformly to $K(r,s)$.

Mercer's theorem plays an important role in the analysis of nonstationary time series. For instance, if $\{Z(t)\}$ is a zero-mean Gaussian process on [0,1], then it admits the following series expansion

$$Z(t) = \sum_{n=1}^{\infty} \frac{f_n(r)}{\sqrt{\lambda_n}} \xi_n ,$$

where λ_n is the eigenvalue of the positive definite kernel $\mathrm{Cov}(Z(r), Z(s))$, $f_n(r)$ is the correponding orthonormal eigenfunction, and $\{\xi_n\} \sim$ NID(0,1). This fact can be easily shown to hold true by Mercer's theorem, and was used in (7) and (25).

The following theorem is also specific to nealy definite kernels, and will be effectively used to compute FDs.

Theorem A3. *Suppose that $K(r,s)$ is continuous, symmetric and nearly definite on* $[0,1] \times [0,1]$. *Then the FD of $K(r,s)$ can be expanded as*

$$D(\lambda) = \prod_{n=1}^{\infty} \left(1 - \frac{\lambda}{\lambda_n}\right)^{l_n} ,$$

where λ_n is the eigenvalue of $K(r,s)$ and l_n is the multiplicity of λ_n.

On the basis of this theorem together with Theorem A1, we can give a set of sufficient conditions for a function of λ to be the FD of a nearly definite kernel.

Theorem A4. *Let $K(r,s)$ be continuous, symmetric and nearly definite on $[0,1]\times[0,1]$ and $\{\lambda_n\}$ a sequence of eigenvalues of $K(r,s)$. Suppose that $\tilde{D}(\lambda)$ is an entire function of λ with $\tilde{D}(0)=1$. Then $\tilde{D}(\lambda)$ becomes the FD of $K(r,s)$ if*

i) *every zero of $\tilde{D}(\lambda)$ is an eigenvalue of $K(r,s)$, and in turn every eigenvalue of $K(r,s)$ is a zero of $\tilde{D}(\lambda)$;*

ii) *$\tilde{D}(\lambda)$ can be expanded as*

$$\tilde{D}(\lambda)=\prod_{n=1}^{\infty}\left(1-\frac{\lambda}{\lambda_n}\right)^{l_n},$$

where l_n is equal to the multiplicity of λ_n.

A word may be in order. If $\tilde{D}(\lambda)$ is an entire function with $\tilde{D}(0)=1$, so is $\tilde{D}^2(\lambda)$, for example. The zero of $\tilde{D}^2(\lambda)$ at λ_n, however, is of order $2l_n$, while the multiplicity of λ_n is l_n. Thus $\tilde{D}^2(\lambda)$ is not the FD of K.

To obtain a candidate $\tilde{D}(\lambda)$ for the FD of K we work with a differential equation with some boundary conditions equivalent to the integral equation (A1). As an illustration let us take up the positive definite kernel

$$K(r,s)=\mathrm{Cov}(J^c(r),J^c(s))=\frac{e^{-c|r-s|}-e^{-c(r+s)}}{2c}, \tag{A3}$$

where $J^c(r)$ is the O-U process defined in (24). Let us consider

$$\begin{aligned} f(r) &= \lambda\int_0^1 K(r,s)f(s)\,ds \\ &= \lambda\left[\int_0^r \frac{e^{-c(r-s)}}{2c}f(s)\,ds+\int_r^1 \frac{e^{-c(s-r)}}{2c}f(s)\,ds-\int_0^1 \frac{e^{-c(r+s)}}{2c}f(s)\,ds\right]. \qquad (A4)\end{aligned}$$

By differentiation we have

$$f''(r)+(\lambda-c^2)f(r)=0, \tag{A5}$$

and it can be shown that this differential equation together with the two boundary conditions $f(0)=0$ and $cf(1)+f'(1)=0$ is equivalent to the integral equation (A4).

The general solution to (A5) is given by

$$f(r)=a_1\cos\mu r+a_2\sin\mu r\,,\quad \mu=\sqrt{\lambda-c^2}\,, \tag{A6}$$

where a_1 and a_2 are arbitrary constants. From the boundary conditions $f(0) = 0$ and $cf(1) + f'(1) = 0$, we have the following homogeneous equation on $a = (a_1, a_2)'$:

$$\begin{pmatrix} 1 & 0 \\ c\cos\mu - \mu\sin\mu & c\sin\mu + \mu\cos\mu \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Leftrightarrow \quad M(\lambda)a = 0\,.$$

The eigenfunction $f(r)$ in (A6) must be nonvanishing, which occurs only when $a \neq 0$. Thus $\lambda(\neq 0)$ is an eigenvalue if and only if $|M(\lambda)| = c\sin\mu + \mu\cos\mu = 0$. We therefore obtain

$$\tilde{D}(\lambda) = \left[\cos\mu + \frac{c}{\mu}\sin\mu\right] \Big/ e^c, \quad \tilde{D}(0) = 1 \tag{A7}$$

as a candidate for the FD of $K(r,s)$ in (A3). Condition i) in Theorem A4 has now been established.

We proceed to establish ii) in the same theorem. From the boundary condition $a_1 = 0$ we have $f(r) = a_2 \sin\mu r$ with $a_2 \neq 0$, which yields the orthonormal eigenfunction given in (26). Thus the multiplicity of every eigenvalue is unity. We have only to show that $\tilde{D}(\lambda)$ admits an infinite product expansion as given in Theorem A4 with $l_n = 1$ for every n. For this purpose we have the following theorem [Whittaker and Watson (1958, p.137)].

Theorem A5. *Let $h(z)$ be an entire function with $h(0) = 1$ and have simple zeros at the points $a_1, a_2, ...$, where $\lim_{n\to\infty} |a_n| = \infty$. Suppose that there is a sequence $\{C_m\}$ of simple closed curves such that $h'(z)/h(z)$ is bounded on C_m as $m \to \infty$. Then $h(z)$ can be expanded as*

$$h(z) = \exp\{h'(0)z\} \prod_{n=1}^{\infty} \left\{\left(1 - \frac{z}{a_n}\right) \exp\left(\frac{z}{a_n}\right)\right\}.$$

Putting $z = \sqrt{\lambda - c^2}$, let us first show that $h(z) = [\cos z + c(\sin z)/z]/(1+c)$ satisfies the conditions in Theorem A5. It can be checked easily that $h(z)$ is an even entire function with $h(0)$=1 and $h'(0)$=0 whose zeros are all simple. Then we can define the zeros of $h(z)$ by $\pm a_1,\ \pm a_2, \cdots$, where $a_n = \sqrt{\lambda_n - c^2}\ \ (> 0)$ and $\lim_{n\to\infty} a_n = \infty$. Let C_m be the square in the complex plane with vertices $m\pi(\pm 1 \pm i)$, $m = 1, 2, \cdots$. Then it is seen that

$$\frac{h'(z)}{h(z)} = \frac{(c\ \cos z)/z - (c/z^2 + 1)\sin z}{\cos z + (c\ \sin z)/z}$$

is bounded on each side of C_m as $m \to \infty$.

It follows from Theorem A5 that we can expand the even function $h(z)$ with $h'(0) = 0$ as

$$\begin{aligned} h(z) &= \frac{1}{1+c}\left[\cos z + \frac{c}{z}\sin z\right] \\ &= \prod_{n=1}^{\infty}\left\{\left(1-\frac{z}{a_n}\right)\exp\left(\frac{z}{a_n}\right)\left(1+\frac{z}{a_n}\right)\exp\left(-\frac{z}{a_n}\right)\right\} \\ &= \prod_{n=1}^{\infty}\left(1-\frac{z^2}{a_n^2}\right) = \prod_{n=1}^{\infty}\left(1-\frac{\lambda-c^2}{\lambda_n-c^2}\right) = \prod_{n=1}^{\infty}\left\{\frac{\lambda_n}{\lambda_n-c^2}\left(1-\frac{\lambda}{\lambda_n}\right)\right\}. \end{aligned}$$

Then we have

$$\begin{aligned} \tilde{D}(\lambda) &= \frac{1+c}{e^c}h(z) = \frac{1+c}{e^c}\prod_{n=1}^{\infty}\left\{\frac{\lambda_n}{\lambda_n-c^2}\left(1-\frac{\lambda}{\lambda_n}\right)\right\} \\ &= \tilde{D}(0)\prod_{n=1}^{\infty}\left(1-\frac{\lambda}{\lambda_n}\right) = \prod_{n=1}^{\infty}\left(1-\frac{\lambda}{\lambda_n}\right), \end{aligned}$$

which establishes ii) in Theorem A4. Thus we have shown that $\tilde{D}(\lambda)$ in (A7) is really the FD of $K(r,s)$ in (A3) so that every eigenvalue is the solution to

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

This fact is effectively used in the discussions of K-asymptotics in Section 3.

Appendix B: Proofs of Theorems

Throughout this appendix we use, for notational simplicity,

$$f(t) = f(K, t/T),\ f^{(1)}(t) = \partial f(K, r)/\partial r|_{r=t/T},\ \triangle f(t) = f(K, t/T) - f(K, (t-1)/T).$$

Proof of Theorem 1

(a) It is easy to show, by the FCLT and the continuous mapping theorem (CMT), that

$$\begin{aligned} \frac{1}{\sqrt{T}}\tilde{b}(K) &= \left(\frac{1}{T}\sum f(t)f(t)'\right)^{-1}\frac{1}{T}\sum f(t)\frac{1}{\sqrt{T}}y_t \\ &\Rightarrow \sigma_L \int_0^1 f(K, r)J^c(r)\,dr\,, \end{aligned}$$

which leads us to the conclusion.

(b) This is an immediate consequence of (28) and the CMT.

(c) We deduce that

$$\begin{aligned} t_{c(K)'\tilde{b}(K)} \Big/ \sqrt{T} &= \frac{1}{\sqrt{T}}c(K)'\tilde{b}(K) \Big/ \sqrt{\frac{1}{T}\sum \tilde{u}_t^2\, c(K)'\left(\sum f(t)f(t)'\right)^{-1} c(K)} \\ &\Rightarrow c(K)'\int_0^1 f(K, r)J^c(r)\,dr \Big/ \sqrt{\int_0^1 \left\{J_{fK}^c(r)\right\}^2\,dr}\ , \end{aligned}$$

which yields the conclusion.

(d) This follows from (b) and the FCLT.

(e) We have $DW = N_T/D_T$, where $D_T = \sum \tilde{u}_t^2$ and

$$\begin{aligned} N_T &= \sum (\tilde{u}_t - \tilde{u}_{t-1})^2 \\ &= \sum \left\{y_t - \tilde{b}(K)'f(t) - (y_{t-1} - \tilde{b}(K)'f(t-1))\right\}^2 \\ &= \sum \triangle y_t^2 - 2\tilde{b}(K)'\sum \triangle f(t)\triangle y_t + \tilde{b}(K)'\sum \triangle f(t)\triangle f(t)'\tilde{b}(K)\,. \end{aligned}$$

Here it can be shown that

$$\sum \triangle f(t)\triangle y_t \cong \frac{1}{T}\sum f^{(1)}(t)\triangle y_t = O_p\left(\frac{1}{\sqrt{T}}\right),$$

$$\sum \triangle f(t)\triangle f(t)' \cong \frac{1}{T^2}\sum f^{(1)}(t)f^{(1)}(t)' = O_p\left(\frac{1}{T}\right),$$

so that

$$\frac{1}{T}N_T = \frac{1}{T}\sum u_t^2 + o_p(1) \to \sigma_S^2 \quad \text{in probability.}$$

Thus we have, from (b) and the CMT,

$$T \times DW \Rightarrow \frac{\sigma_S^2}{\sigma_L^2} \frac{1}{\int_0^1 \left\{ J_{fK}^c(r) \right\}^2 \, dr} \tag{B1}$$

which establishes (e).

Proof of Theorem 2

(a) We first have

$$\begin{aligned} J^c(r) &= \sum_{n=1}^{\infty} \frac{1}{\sqrt{\lambda_n}} f_n(r) \xi_n = \left(\sum_{n=1}^{K} + \sum_{n=K+1}^{\infty} \right) \frac{1}{\sqrt{\lambda_n}} f_n(r) \xi_n \\ &= f(K,r)' \Lambda_K^{-1/2} \xi(K) + f(\perp, r)' \Lambda_\perp^{-1/2} \xi(\perp) \,, \end{aligned} \tag{$B2$}$$

where

$$\Lambda_K = \operatorname{diag}\left(\lambda_1, \, \ldots, \, \lambda_K\right), \quad \Lambda_\perp = \operatorname{diag}\left(\lambda_{K+1}, \, \ldots, \, \right)$$
$$f(\perp, r)' = \left(f(K+1, r), \, \ldots, \, \right), \quad \xi(\perp)' = \left(\xi_{K+1}, \, \ldots, \, \right).$$

Then we have, from the orthonormality of $\{f_k(r)\}$,

$$\int_0^1 f(K,r) J^c(r) \, dr = \Lambda_K^{-1/2} \xi(K) \tag{B3}$$

so that

$$c(K)' \int_0^1 f(K,r) J^c(r) \, dr = c(K)' \Lambda_K^{-1/2} \xi(K) \Rightarrow \mathrm{N}\left(0, c(K)' \Lambda_K^{-1} c(K)\right) \,,$$

which yields the conclusion.

(b) It follows from the proof of (a) that

$$J_{fK}^c(r) = J^c(r) - \int_0^1 f(K,s)' J^c(s) \, ds \, f(K,r) = f(\perp, r)' \Lambda_\perp^{-1/2} \xi(\perp) \,,$$

which gives

$$Y(K) = \int_0^1 \left\{ J_{fK}^c(r) \right\}^2 \, dr = \xi(\perp)' \Lambda_\perp^{-1} \xi(\perp) = \sum_{n=K+1}^{\infty} \frac{1}{\lambda_n} \xi_n^2 \,. \tag{B4}$$

Since we have

$$\sum_{n=K+1}^{\infty} \frac{1}{\lambda_n} = \sum_{n=K+1}^{\infty} \frac{1}{\left((n-1/2)\pi + d_n \right)^2 + c^2} = \frac{1}{\pi^2 K} + o\left(\frac{1}{K} \right) \,,$$

where $0 < d_n < \pi/2$, we get the conclusion.

(c) This is an immediate consequence of (a) and (b) since

$$\int_0^1 f(K,r) J^c(r)\, dr = O_p(1)\,,$$
$$\int_0^1 \left\{ J^c_{fK}(r) \right\}^2 dr = O_p\left(\frac{1}{K}\right)\,.$$

(d) This immediately follows from (b).

(e) This follows from (B1) and (b).

Proof of Theorem 3

Here we consider, for simplicity, the case where the error term $\{u_t\}$ in the DGP (22) follows i.i.d.(0,σ^2). Then we have

$$\begin{pmatrix} \sum y_{t-1}^2 & \sum y_{t-1} f(t)' \\ \sum y_{t-1} f(t) & \sum f(t) f(t)' \end{pmatrix} \begin{pmatrix} \hat{\rho} \\ \hat{b}(K) \end{pmatrix} = \begin{pmatrix} \sum y_{t-1} y_t \\ \sum f(t) y_t \end{pmatrix}\,, \tag{B5}$$

so that $ADF_\rho = T(\hat{\rho}-1) = U_T/V_T$, where

$$\begin{aligned} U_T &= \frac{1}{T}\sum \left[y_{t-1} - \sum y_{s-1} f(s)' \left(\sum f(s) f(s)' \right)^{-1} f(t) \right] \triangle y_t \\ &\Rightarrow \sigma^2 \int_0^1 J^c_{fK}(r)\, dJ^c(r)\,, \\ V_T &= \frac{1}{T}\left[\sum y_{t-1}^2 - \sum y_{s-1} f(s)' \left(\sum f(s) f(s)' \right)^{-1} \sum y_{s-1} f(s) \right] \\ &\Rightarrow \sigma^2 \int_0^1 \left\{ J^c_{fK}(r) \right\}^2 dr\,. \end{aligned}$$

Since joint weak convergence of U_T and V_T is ensured, (30) is established.

We next consider

$$ADF_t = (\hat{\rho}-1)/SE(\hat{\rho}) = T(\hat{\rho}-1) \Big/ \left(\hat{\sigma}/\sqrt{V_T} \right)\,,$$

where

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{T-K-1}\sum \hat{u}_t^2 = \frac{1}{T-K-1}\sum \left(y_t - \hat{\rho} y_{t-1} - \hat{b}(K)' f(t) \right) y_t \\ &= \frac{1}{T-K-1}\sum \left[u_t^2 - (\hat{\rho}-\rho) y_{t-1} u_t - \hat{b}(K)' f(t) u_t \right]\,. \end{aligned}$$

It can be shown after some algebra that $\hat{\sigma}^2 \to \sigma^2$ in probability, which yields (31) because of (30) and the CMT.

It follows from (B5) that

$$\hat{b}(K) = \frac{1}{T}\sum f(t) \left\{ u_t - (\hat{\rho}-\rho) y_{t-1} \right\} \Big/ \left(\frac{1}{T}\sum f(t) f(t)' \right)^{-1}\,,$$

which evidently yields (32). As for (33), we have

$$t_{c(K)'\hat{b}(K)} = c(K)'\hat{b}(K) \Big/ \sqrt{c(K)'\hat{\Sigma}c(K)} \ ,$$

where

$$\begin{aligned}\hat{\Sigma} &= \frac{1}{T-K-1}\sum \hat{u}_t^2 \left(\sum f(t)f(t)' - \sum y_{t-1}f(t)\sum y_{t-1}f(t)'/\sum y_{t-1}^2\right)^{-1} \\ &\Rightarrow \sigma^2\Sigma(K,f)^{-1},\end{aligned}$$

with $\Sigma(K,f)$ defined in the theorem, which leads us to (33).

Proof of Theorem 4

We continue to assume that the error term $\{u_t\}$ in (22) follows i.i.d.$(0,\sigma^2)$. We first have that $Y(K)$ in (B4) tends to $\mathrm{N}(m_K,\sigma_K^2)$, where $m_K = 1/(\pi^2 K) + o(1/K)$ and

$$\begin{aligned}\sigma_K^2 &= \sum_{n=K+1}^{\infty}\frac{2}{\lambda_n^2} = \sum_{n=K+1}^{\infty}\frac{2}{(((n-1/2)\pi + d_n)^2 + c^2)^2} \\ &= \frac{2}{3\pi^4 K^3} + o\left(\frac{1}{K^3}\right).\end{aligned}$$

We next obtain, after some algebra,

$$\begin{aligned}\int_0^1 J_{fK}^c(r)\,dJ^c(r) &= \int_0^1 \left[J^c(r) - \int_0^1 f(K,s)'J^c(s)\,ds\, f(K,r)\right] dJ^c(r) \\ &= \frac{1}{2}\left((J^c(1))^2 - 1\right) - \xi(K)'\Lambda_K^{-1/2}\left[f(K,1)J^c(1) - \int_0^1 f^{(1)}(K,r)J^c(r)\,dr\right] \\ &= -\frac{1}{2} + \frac{1}{2}G(K) + H(K),\end{aligned}$$

where

$$G(K) = \xi(\perp)'\Lambda_{\perp}^{-1/2}f(\perp,1)f(\perp,1)'\Lambda_{\perp}^{-1/2}\xi(\perp),$$

$$H(K) = \xi(K)'\Lambda_K^{-1/2}\int_0^1 f^{(1)}(K,r)f(\perp,r)'\,dr\Lambda_{\perp}^{-1/2}\xi(\perp). \qquad \text{(B6)}$$

Here we have $E(H(K)) = 0$ and

$$\begin{aligned}E(G(K)) &= f(\perp,1)'\Lambda_{\perp}^{-1}f(\perp,1) = \sum_{n=K+1}^{\infty}\frac{1}{\lambda_n}f_n^2(1) \\ &< \sum_{n=K+1}^{\infty}\frac{1}{(n-1/2)^2\pi^2} = O\left(\frac{1}{K}\right), \\ V(H(K)) &= tr\left[\Lambda_K^{-1}\int_0^1 f^{(1)}(K,r)f(\perp,r)'\,dr\Lambda_{\perp}^{-1}\int_0^1 f(\perp,r)f^{(1)}(K,r)'\,dr\right].\end{aligned}$$

To evaluate $V(H(K))$, let us consider, for $1 \le j \le K$ and $k \ge K+1$,

$$
\begin{aligned}
a_{jk} &= \int_0^1 f_j^{(1)}(r) f_k(r)\, dr \\
&= \frac{\mu_j}{M_j\, M_k} \int_0^1 \cos\mu_j r\, \sin\mu_k r\, dr \\
&= \frac{\mu_j}{2M_j\, M_k} \left[\frac{1-\cos(\mu_j+\mu_k)}{\mu_j+\mu_k} + \frac{1-\cos(\mu_k-\mu_j)}{\mu_k-\mu_j} \right],
\end{aligned}
$$

where $\mu_j = (j-1/2)\pi + d_j$ with $0 < d_j < \pi/2$ and $0 < M_j < 1$. Then we have

$$
\begin{aligned}
V(H(K)) &= \sum_{j=1}^{K} \sum_{k=K+1}^{\infty} \frac{a_{jk}^2}{\lambda_j \lambda_k} \\
&= \frac{1}{4} \sum_{j=1}^{K} \sum_{k=K+1}^{\infty} \frac{\mu_j^2}{(\mu_j^2+c^2)(\mu_k^2+c^2) M_j^2\, M_k^2} \left[\frac{1-\cos(\mu_j+\mu_k)}{\mu_j+\mu_k} + \frac{1-\cos(\mu_k-\mu_j)}{\mu_k-\mu_j} \right]^2 \\
&< \text{const.} \times \left[\sum_{j=1}^{K} \sum_{k=K+1}^{\infty} \left\{ \frac{1}{(k-1/2)^2 k^2} + \frac{1}{k^2(k^2-j^2)} + \frac{1}{\mu_k^2(\mu_k-\mu_j)^2} \right\} \right] \\
&= O\left(\frac{\log K}{K^2} \right).
\end{aligned}
$$

Thus we obtain

$$
\begin{aligned}
\frac{\int_0^1 J_{fK}^c(r)\, dJ^c(r)}{\int_0^1 \left\{ J_{fK}^c(r) \right\}^2 dr} &= \frac{-1/2 + O_p(\sqrt{\log K}/K)}{1/(\pi^2 K) + \sqrt{2} Z/(\pi^2 K \sqrt{3K}) + o_p(1/(K\sqrt{K}))} \\
&= \pi^2 K \left(-\frac{1}{2} - \frac{Z}{\sqrt{6K}} + o_p\left(\frac{1}{\sqrt{K}} \right) \right),
\end{aligned}
$$

where $Z \sim \mathrm{N}(0,1)$, and arrive at (34).

Similarly we have

$$
\begin{aligned}
\frac{\int_0^1 J_{fK}^c(r)\, dJ^c(r)}{\left(\int_0^1 \left\{ J_{fK}^c(r) \right\}^2 dr \right)^{1/2}} &= \frac{-1/2 + O_p(\sqrt{\log K}/K)}{\left(1/(\pi^2 K) + \sqrt{2} Z/(\pi^2 K \sqrt{3K}) + o_p(1/(K\sqrt{K})) \right)^{1/2}} \\
&= \pi \sqrt{K} \left(-\frac{1}{2} + \frac{Z}{2\sqrt{6K}} + o_p\left(\frac{1}{\sqrt{K}} \right) \right),
\end{aligned}
$$

which yields (35).

We next deduce that

$$
c(K)' Y(K,f) = \frac{\pi^2 K}{2}\, c(K)' \Lambda_K^{-1/2} \xi(K) + o_p(K) \tag{B7}
$$

which proves (36). Finally we have

$$
\Sigma(K,f) = I_K - \Lambda_K^{-1/2} \xi(K) \xi(K)' \Lambda_K^{-1/2} \Big/ \int_0^1 \{J^c(r)\}^2\, dr
$$

so that, using the orthogonal matrix P such that

$$P\Lambda_K^{-1/2}\xi(K)\xi(K)'\Lambda_K^{-1/2}P' = \text{diag}\left(\xi(K)'\Lambda_K^{-1/2}\xi(K),\ 0,\ \ldots,\ 0\right),$$

we obtain

$$\begin{aligned} c(K)'\Sigma(K,f)^{-1}c(K) &= \frac{\beta_1^2}{1-\xi(K)'\Lambda_K^{-1}\xi(K)\Big/\int_0^1\{J^c(r)\}^2\,dr} + \sum_{j=2}^{K}\beta_j^2 \\ &= \frac{\beta_1^2\int_0^1\{J^c(r)\}^2\,dr}{\xi(\perp)'\Lambda_\perp^{-1}\xi(\perp)} + \sum_{j=2}^{K}\beta_j^2 \\ &= O_p(K)\,, \end{aligned}$$

where $\beta = (\beta_1,\ \ldots,\ \beta_K)' = Pc(K)$. Thus (37) follows from this and (B7).

Proof of Theorem 5

Suppose that $c > 0$ so that $g_k(r) = f_k(r)$ with $f_k(r)$ defined in (26). Then we obtain (44) and (46) concerning the limiting distribution of $T(\hat{\rho}-1)$. Let us consider $T(\tilde{\rho}-1) = A_T/B_T$, where $B_T = \sum\tilde{\varepsilon}_{t-1}^2/T^2$ and

$$\begin{aligned} A_T &= \frac{1}{T}\sum\tilde{\varepsilon}_{t-1}\left(\tilde{\varepsilon}_t-\tilde{\varepsilon}_{t-1}\right) \\ &= \frac{1}{T}\sum\left[y_{t-1}-f(t-1)'\left(\sum f(t)f(t)'\right)^{-1}\sum f(t)y_t\right] \\ &\qquad\times\left[\triangle y_t-\triangle f(t)'\left(\sum f(t)f(t)'\right)^{-1}\sum f(t)y_t\right] \\ &\Rightarrow \sigma^2\left[\int_0^1\left\{J^c(r)-f(K,r)'\int_0^1 f(K,s)J^c(s)\,ds\right\}\,dJ^c(r)\right. \\ &\qquad -\int_0^1 J^c(r)f^{(1)}(K,r)'\,dr\int_0^1 f(K,r)J^c(r)\,dr \\ &\qquad \left.+\int_0^1 f(K,r)'J^c(r)\,dr\int_0^1 f(K,r)f^{(1)}(K,r)'\,dr\int_0^1 f(K,r)J^c(r)\,dr\right] \\ &= \int_0^1 J_{fK}^c(r)\,dJ^c(r) + H_{fK}\,, \end{aligned}$$

where H_{fK} is defined in the theorem. Thus we obtain (45) by joint weak convergence of A_T and B_T, and the CMT. As for K-asymptotics, substituting (B2) and (B3) into H_{fK}, we have by partial integration

$$\begin{aligned} H_{fK} &= \frac{1}{2}\int_0^1 J^c(r)f(K,r)'\,dr f(K,1)f(K,1)'\int_0^1 f(K,r)J^c(r)\,dr \\ &\qquad -\int_0^1 J^c(r)f^{(1)}(K,r)'\,dr\int_0^1 f(K,r)J^c(r)\,dr \\ &= \xi(K)'\Lambda_K^{-1/2}\left[\frac{1}{2}f(K,1)f(K,1)'-\int_0^1 f(K,r)f^{(1)}(K,r)'\,dr\right]\Lambda_K^{-1/2}\xi(K) \end{aligned}$$

$$
\begin{aligned}
& -\xi(\perp)'\Lambda_{\perp}^{-1/2}\int_0^1 f(\perp,r)f^{(1)}(K,r)'\,dr\Lambda_K^{-1/2}\xi(K) \\
= \;& -H(K)\,,
\end{aligned}
$$

where $H(K)$ is defined in (B6), which is $O_p(\sqrt{\log K}/K)$. Thus we arrive at (46) because of Theorem 4.

Table 1

R^2 and DW Statistics for Model (39) 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 2

Distributions of $T(\hat{\rho}-1)$ for Model (40) with $g_k(r)=\phi_k(r)$ or $f_k(r)$

	$K=1$	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

Table 3

Distributions of $T(\hat{\rho}-1)$ for Model (40) with $g_k(r) = Q_k(r)$

	$K = 1$	2	5	10	15	20
			$\rho = 1$			
Mean	−10.39	−14.84	−26.69	−46.05	−65.12	−82.71
	(−4)	(−8)	(−20)	(−40)	(−60)	(−80)
SD	5.94	7.44	9.89	12.49	14.56	16.16
	(4.00)	(5.66)	(8.94)	(12.65)	(15.49)	(17.89)
			$\rho = 0.975$			
Mean	−17.34	−20.47	−30.09	−48.00	−66.41	−83.75
SD	7.13	8.15	10.01	12.54	14.62	16.23
			$\rho = 0.95$			
Mean	−26.65	−29.22	−37.32	−53.13	−70.12	−86.66
SD	8.33	9.00	10.32	12.63	14.72	16.29

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

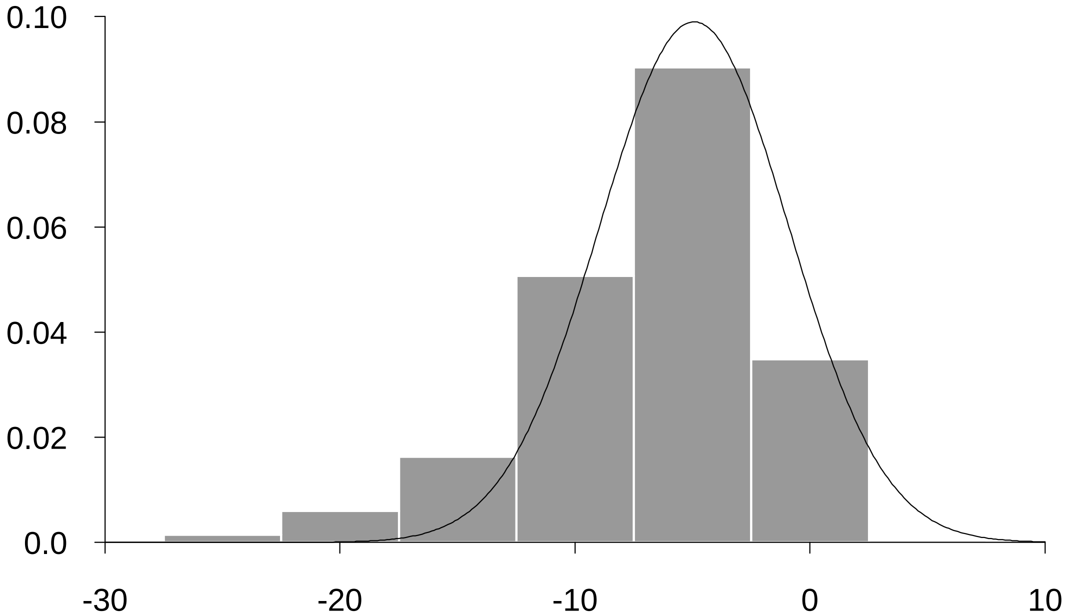

Figure 2. Distributions of $T(\hat{\rho}-1)$ with $T=400$, $K=20$ and $\rho=1$

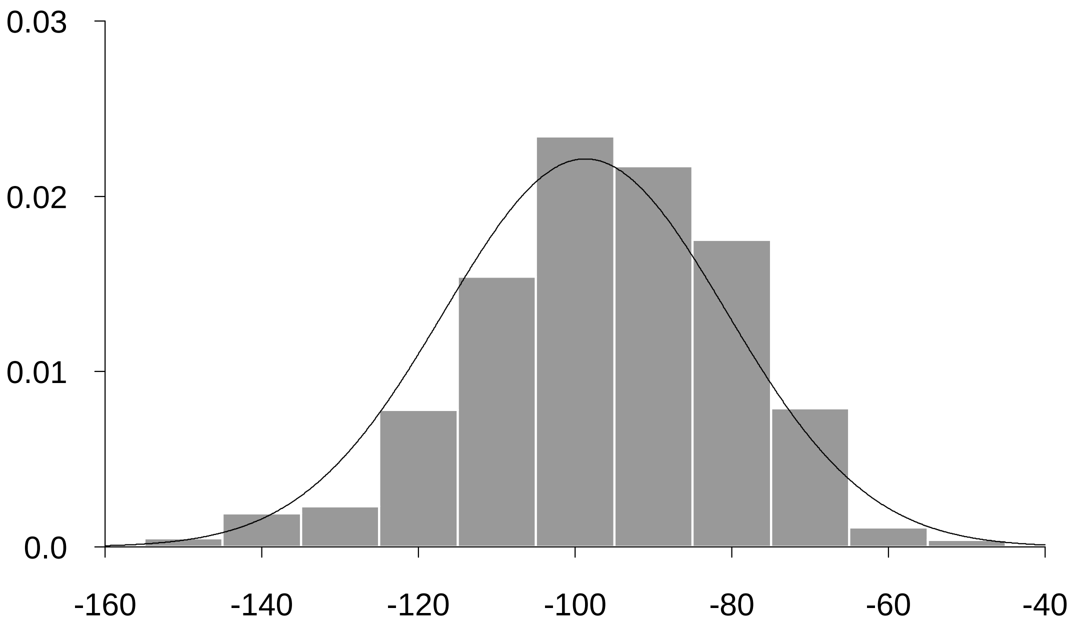

Figure 3. Distributions of $T(\hat{\rho}-1)$ with $T=400$, $K=20$ and $\rho=0.95$
(Regression with $f(K,r)$ as deterministic trends)

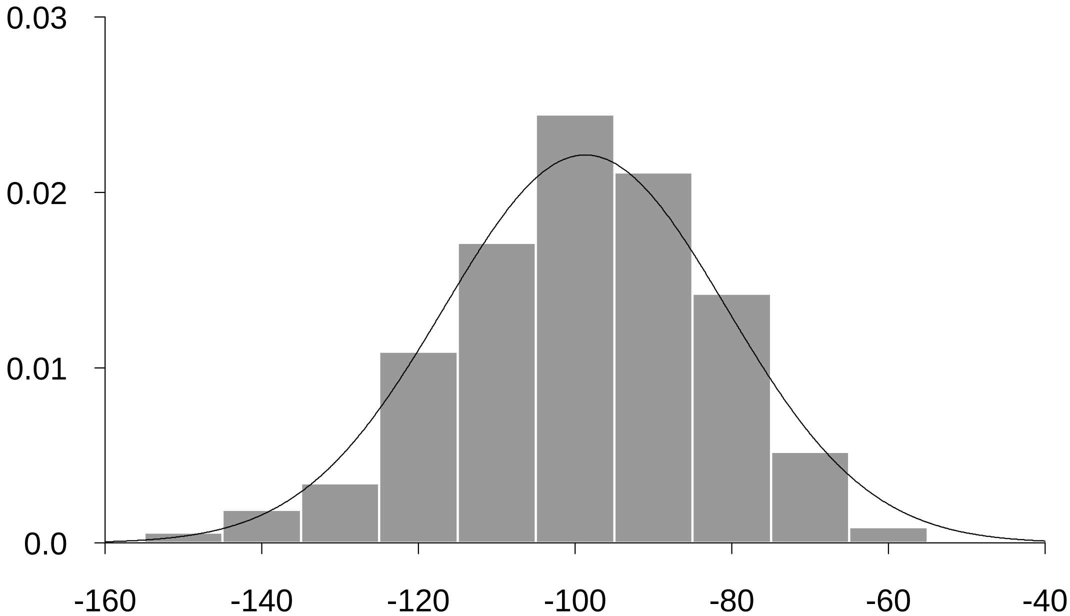

Figure 4. Distributions of $T(\hat{\rho}-1)$ with $T=400$, $K=20$ and $\rho=0.95$
(Regression with $\phi(K,r)$ as deterministic trends)

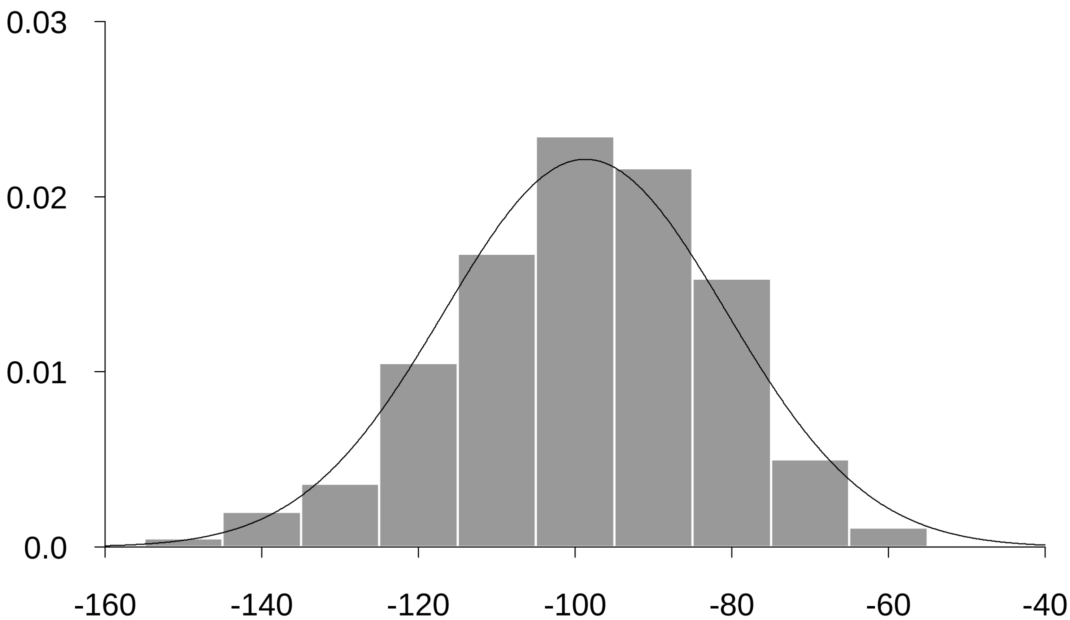

Figure 5. Distributions of $T(\hat{\rho}-1)$ with $T=400$, $K=20$ and $\rho=1$
(Regression with polynomial trends)

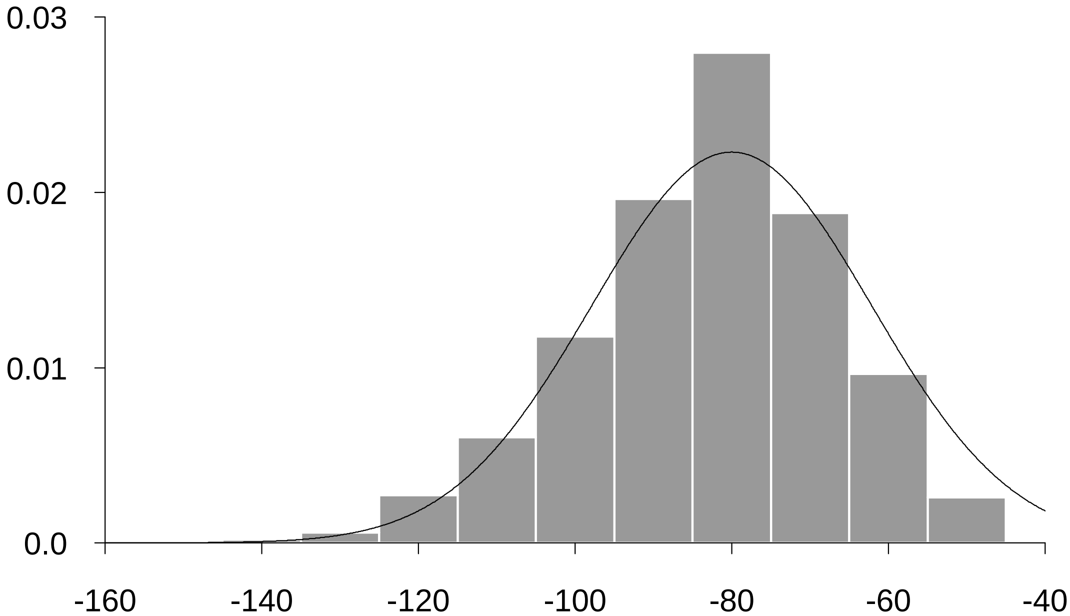

Figure 6. Distributions of $T(\hat{\rho}-1)$ with $T=400$, $K=20$ and $\rho=0.95$
(Regression with polynomial trends)