

UNIT ROOT TESTING FOR FUNCTIONALS OF LINEAR PROCESSES

WEI BIAO WU
University of Chicago

We consider the unit root testing problem with errors being nonlinear transforms of linear processes. When the linear processes are long-range dependent, the asymptotic distributions in the unit root testing problem are shown to be functionals of Hermite processes. Functional limit theorems for nonlinear transforms of linear processes are established. The obtained results differ sharply from the classical cases where asymptotic distributions are functionals of Brownian motions.

1. INTRODUCTION

The unit root testing problem has been extensively studied in the econometrics literature. In the paper we consider the following model:

$$y_t = \rho y_{t-1} + u_t, \quad \text{with } \rho = 1, \quad y_0 = 0, \quad (1)$$

and

$$u_t = K(x_t), \quad x_t = \sum_{i=0}^{\infty} a_i \varepsilon_{t-i}, \quad (2)$$

where $(\varepsilon_t)_{t \in \mathbb{Z}}$ are independent and identically distributed (i.i.d.) random variables with mean zero and finite variance, the real coefficients $(a_i)_{i=0}^{\infty}$ are square summable, and K is a measurable function such that $E[K(x_t)] = 0$ and $E[K^2(x_t)] < \infty$. Here (x_t) can be autoregressive moving average (ARMA) or fractional autoregressive integrated moving average (ARIMA) processes and K can be nonlinear functions. See Granger and Joyeux (1980) and Hosking (1981) for an introduction to fractionally integrated processes. Dittmann and Granger (2002) recently discussed nonlinear transforms of fractionally integrated processes and found some very interesting properties. We are interested in testing the hypothesis $H_0: \rho = 1$ versus the alternative $H_A: \rho \neq 1$. Given the observations y_0, \dots, y_n , the least squares slope estimate has the form

The author thanks the referee and Professor B. Hansen for their valuable suggestions. The work is supported in part by NSF grant DMS-04478704. Address correspondence to Wei Biao Wu, Department of Statistics, University of Chicago, 5734 S. University Avenue, Chicago, IL 60637, USA; e-mail: wbwu@galton.uchicago.edu.

$$\hat{\rho}_n = \frac{\sum_{t=1}^n y_t y_{t-1}}{\sum_{t=1}^n y_{t-1}^2}. \quad (3)$$

In the classical unit root testing problem, u_t are often assumed to be i.i.d.; see, for example, Dickey and Fuller (1979, 1981). In this case, it can be shown that

$$n(\hat{\rho}_n - 1) \Rightarrow \frac{\mathbb{B}^2(1) - 1}{2 \int_0^1 \mathbb{B}^2(t) dt}, \quad (4)$$

where \mathbb{B} is a standard Brownian motion and \Rightarrow denotes convergence in distribution. The i.i.d. assumption is very restrictive in practice. Various generalizations to dependent random variables have been extensively pursued for processes with special dependence structures. For a partial list, see Phillips (1987) for strong mixing processes, Phillips and Xiao (1998) and Wang, Lin, and Gulati (2002) for linear processes in which $K(x) = x$, and Chan and Terrin (1995) for Gaussian processes. The limiting distributions can be similarly expressed as functionals of Brownian motions. In an important paper, Sowell (1990) considers the unit root testing problem for long-memory processes. In particular, Sowell considers the special case of (1) with the identity function $K(x) = x$ and the fractionally integrated series $u_t \sim I(d)$, where $-\frac{1}{2} < d < \frac{1}{2}$. Namely,

$$u_t = (1 - B)^{-d} \varepsilon_t = \sum_{i=0}^{\infty} a_i \varepsilon_{t-i}, \quad \text{where } a_k = \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)}. \quad (5)$$

If $0 < d < \frac{1}{2}$, then the covariance function of u_t satisfies $\gamma_u(k) = E(u_0 u_k) \sim k^{2d-1} c$ for some constant $c > 0$, which is not summable because $2d - 1 > -1$. Such property is usually referred to as *long-range dependence* or *long memory*. In this case, the limiting distribution is strikingly different from (4). Sowell (1990) proves that

$$n(\hat{\rho}_n - 1) \Rightarrow \frac{\mathbb{B}_H^2(1)}{2 \int_0^1 \mathbb{B}_H^2(t) dt}, \quad (6)$$

where \mathbb{B}_H is the fractional Brownian motion with Hurst index $H = d + \frac{1}{2}$. Namely, \mathbb{B}_H is a mean zero Gaussian process with covariance function $\text{cov}[\mathbb{B}_H(s), \mathbb{B}_H(t)] = (t^{2H} + s^{2H} - |t - s|^{2H})/2$, $s, t \geq 0$ (cf. Mandelbrot and Van Ness, 1969). See also Wang, Lin, and Gulati (2003) for some recent developments.

Long-memory processes have received considerable attention in the econometrics literature. It would be impossible to compile a complete list. See Baillie

(1996) for an excellent survey and Robinson (2003), Doukhan, Oppenheim, and Taquq (2003), and Caporale and Gil-Alana (2004), among others, for some recent contributions. In this paper we shall generalize previous results on unit root problems in two directions, namely, by allowing general nonlinear transforms K and general forms of linear processes that include ARMA and fractional ARIMA processes as special cases. It turns out that, because of nonlinearity, the asymptotic problem becomes considerably more challenging when K assumes a general form than that in special cases such as $K(w) = w$. On the other hand, the limiting distributions have more interesting structures that appear rather atypical in the sense that they may no longer be functionals of Brownian motions. Instead, under suitable conditions, the asymptotic distributions are functionals of Hermite processes. Dittmann and Granger (2002) discuss nonlinear transforms of fractionally integrated processes and show that the dependence structure of the transformed sequence $u_t = K(x_t)$ may be significantly different from that of the input sequence x_t .

The paper is structured as follows. Section 2 presents functional limit theorems for the partial sum process $S_n = u_1 + \dots + u_n$. The latter problem has a rich literature, and it plays an important role in unit root problems. Our asymptotic results go beyond existing ones by allowing non-Gaussian processes and general functionals K ; see the discussion in Section 2. Applications to the unit root testing problem (1) are made in Section 3. Proofs are given in the Appendix.

2. FUNCTIONAL LIMIT THEOREMS

Functional limit theorems are powerful tools for asymptotic distributions of $\hat{\rho}_n$ (Phillips, 1987). In this section we shall present a functional limit theory for $S_n = S_n(K) = u_1 + \dots + u_n$. For $t \geq 0$ let $S_t = S_{[t]} + (t - [t])u_{[t]+1}$, where $[t]$ is the integer part of t . Then S_t is a continuous function in $t \geq 0$. Let $C[0,1]$ be the collection of all continuous functions defined on $[0,1]$ and define the metric $d(f, g) = \sup_{0 \leq t \leq 1} |f(t) - g(t)|$ for $f, g \in C[0,1]$. For $\xi_n, \xi \in C[0,1]$, denote by $\xi_n(t) \Rightarrow \xi(t)$ the weak convergence of ξ_n to ξ in the space $C[0,1]$ under the metric d . See Billingsley (1968) for an extensive treatment of the weak convergence theory in $C[0,1]$. Under mild conditions, the limiting distributions of $\{S_{nu}, 0 \leq u \leq 1\}$ under proper scaling are shown to be either Hermite processes or Brownian motions, depending on whether the process is long- or short-memory and a quantity (power rank) related to K .

Let $\|\xi\| = [E(\xi^2)]^{1/2}$ be the L^2 norm of the random variable ξ . Define the shift process $\mathcal{F}_t = (\dots, \varepsilon_{t-1}, \varepsilon_t)$ and the truncated processes $x_{t,k} = E(x_t | \mathcal{F}_k)$. Then for $n \geq 1$, $x_{n,0} = \sum_{i=n}^{\infty} a_i \varepsilon_{n-i}$ is independent of $x_n - x_{n,0} = \sum_{i=0}^{n-1} a_i \varepsilon_{n-i}$. Now define functions

$$K_n(w) = E[K(w + x_n - x_{n,0})] \quad \text{and} \quad K_{\infty}(w) = E[K(w + x_n)]. \quad (7)$$

Write κ_r for the r th derivative $K_{\infty}^{(r)}(0)$ if it exists. We say that K has *power rank* p if $\kappa_p \neq 0$ and $\kappa_r = 0$ for $1 \leq r \leq p - 1$ (Ho and Hsing, 1997). In the

case of Gaussian processes, p is the Hermite rank (Taqqu, 1975, 1979; Dittmann and Granger, 2002). For a function g let $g(w; \lambda) = \sup_{|y| \leq \lambda} |g(w + y)|$ be the local maximal function. Let $\mathbb{C}^p(\mathbb{R})$ ($p \geq 0$) be the collection of functions f with p th-order partial derivatives. To state our main results, we need the following condition.

Condition 1. Let $\mathbb{E}(|\varepsilon_1|^q) < \infty$ for some $2 < q \leq 4$ and $K_n \in \mathbb{C}^{p+1}(\mathbb{R})$ for all large n . Assume that for some $\lambda > 0$,

$$\sum_{\alpha=0}^{p+1} \|K_{n-1}^{(\alpha)}(x_{n,0}; \lambda)\| + \sum_{\alpha=0}^{p-1} \| |\varepsilon_1|^{q/2} K_{n-1}^{(\alpha)}(x_{n,1}) \| + \|\varepsilon_1 K_{n-1}^{(p)}(x_{n,1})\| = \mathcal{O}(1). \quad (8)$$

Condition 1 is actually quite mild; see Remark 1 in the Appendix and Example 1 in Section 3 for more discussion. This condition and relations (12) and (14), which follow, impose certain smoothness requirements on K_{n-1} . They are easily verifiable.

Let $\{B(u), u \in \mathbb{R}\}$ be a standard two-sided Brownian motion; let the simplex $\mathcal{S}_t = \{(u_1, \dots, u_r) \in \mathbb{R}^r: -\infty < u_1 < \dots < u_r < t\}$. For $\frac{1}{2} < \beta < \frac{1}{2} + 1/(2r)$, define the *Hermite process* (cf. Surgailis, 1982; Avram and Taqqu, 1987)

$$Z_{r,\beta}(t) = \int_{\mathcal{S}_t} \int_0^t \prod_{i=1}^r [\max(v - u_i, 0)]^{-\beta} dv dB(u_1) \dots dB(u_r). \quad (9)$$

When $r = 1$, $Z_{r,\beta}(t)$ is the fractional Brownian motion with Hurst index $\frac{3}{2} - \beta$; $Z_{r,\beta}(1)$ is called the *multiple Wiener–Ito integral*. Throughout the paper the notation $\ell(n)$ denotes a slowly varying function, namely, $\lim_{n \rightarrow \infty} \ell(\lambda n)/\ell(n) = 1$ for all $\lambda > 0$ (Bingham, Goldie, and Teugels, 1987).

THEOREM 1. *Assume that Condition 1 holds with $q = 4$ and that K has power rank $p \geq 1$. Let $a_n = n^{-\beta} \ell(n)$ with $\frac{1}{2} < \beta < 1$, $n \geq 1$. (i) If $p(2\beta - 1) < 1$, then*

$$\{S_{n_t}/\sigma_{n,p}, 0 \leq t \leq 1\} \Rightarrow \{\kappa_p Z_{p,\beta}(t), 0 \leq t \leq 1\} \quad (10)$$

in the space $C[0,1]$, where $\sigma_{n,p} = n^{1-p(\beta-1/2)} \ell^p(n)$. (ii) If $p(2\beta - 1) > 1$ or $p(2\beta - 1) = 1$ and $\sum_{n=1}^{\infty} |\ell^p(n)|/n < \infty$, then

$$\{S_{n_t}/\sqrt{n}, 0 \leq t \leq 1\} \Rightarrow \{\sigma B(t), 0 \leq t \leq 1\} \quad (11)$$

in the space $C[0,1]$ for some $\sigma < \infty$.

THEOREM 2. *Assume $\sum_{k=1}^{\infty} |a_k| < \infty$ and*

$$\|K_{n-1}(x_{n,1}) - K_{n-1}(x_{n,0})\| = \mathcal{O}(|a_{n-1}|). \quad (12)$$

Then

$$\{S_{nt}/\sqrt{n}, 0 \leq t \leq 1\} \Rightarrow \{\sigma B(t), 0 \leq t \leq 1\} \quad (13)$$

in the space $C[0,1]$ for some $\sigma < \infty$. A sufficient condition for (12) is

$$\|K'_{n-1}(x_{n,0};1)\| + \|K_{n-1}(x_{n,1})\varepsilon_1\| = \mathcal{O}(1). \quad (14)$$

Theorems 1 and 2 improve previous results in several aspects. We shall compare our results with earlier ones that are based on strong mixing processes and near-epoch dependence (NED). The concept of strong mixing is proposed by Rosenblatt (1956); see the review by Bradley (1986) for various strong mixing conditions. Gallant and White (1988) apply NED to characterize weak dependence.

Functional central limit theorems for strongly mixing processes have been widely discussed in the literature; see Peligrad (1986) for an excellent survey. However, for linear processes very restrictive conditions on the decay rate of a_n are needed to ensure the strong mixing property; see Withers (1981), Andrews (1984), Pham and Tran (1985), Gorodetskii (1977), and Doukhan (1994) for more discussion about mixing properties of linear processes. Withers (1981) shows that, if $E(\varepsilon_i^2) < \infty$, then under certain regularity conditions on the density function of ε_i and the inequality

$$\sum_{n=1}^{\infty} \max(A_n^{1/3}, \sqrt{A_n |\log A_n|}) < \infty,$$

the process x_t is strong mixing. Here $A_n = \sum_{i=n}^{\infty} a_i^2$. In the case that $a_n = n^{-\delta}$, $n \geq 1$, the preceding inequality requires $\delta > 2$ and the strong mixing coefficients $\alpha_k = O[\sum_{n=k}^{\infty} \max(A_n^{1/3}, \sqrt{A_n |\log A_n|})] = O(k^{(4-2\delta)/3})$ (Withers, 1981). For strong mixing processes, the celebrated central limit theorem by Ibragimov and Linnik (1971) asserts that S_n/\sqrt{n} is asymptotically normal if $\sum_{k=1}^{\infty} \alpha_k^{\gamma/(2+\gamma)} < \infty$ and $E[|K(x_t)|^{2+\gamma}] < \infty$ hold for some $\gamma > 0$. See Theorem 18.5.3 in Ibragimov and Linnik (1971). Therefore, even under the stronger moment condition $E[|K(x_t)|^3] < \infty$ (namely, $\gamma = 1$), one needs to impose $[(4-2\delta)/3](\frac{1}{3}) < -1$, or $\delta > \frac{13}{2}$, to ensure the asymptotic normality of S_n/\sqrt{n} . In comparison, our natural summability condition $\sum_{k=1}^{\infty} |a_k| < \infty$ only needs $\delta > 1$.

We now compare our Theorems 1 and 2 with limit theorems for linear processes based on NED. De Jong and Davidson (2000) recently developed new conditions for functional limit theorems for near-epoch dependent sequences; see Theorem 3.1 therein. In particular, we consider the weak convergence of $\{S_{nu}/\sqrt{n}, 0 \leq u \leq 1\}$ for two special cases (i) $K(x) = x$ and (ii) $K(x) = x^2 - E(x^2)$.

Generally speaking, limit theorems for transforms of linear processes based on NED require stronger conditions on the decay rate of (a_n) toward 0, espe-

cially when the function K is nonlinear. In comparison, our results impose minimal conditions on a_n . A key condition in Theorem 3.1 of De Jong and Davidson (2000) is that $K(x_t)/\sqrt{n}$ is L_2 -NED of size $-\frac{1}{2}$ on ε_t ; namely, there exists an $\eta > 0$ such that

$$\|K(x_t)/\sqrt{n} - E[K(x_t)/\sqrt{n} | \varepsilon_{t-m}, \varepsilon_{t-m+1}, \dots, \varepsilon_{t+m}]\| = n^{-1/2}O(m^{-1/2-\eta}). \quad (15)$$

See Definition 1 and Assumption 1 in De Jong and Davidson (2000) for more details. Under (i), namely, $K(x) = x$, the NED condition (15) requires that $A_{m+1}^{1/2} = \sqrt{\sum_{i=m+1}^{\infty} a_i^2} = O(m^{-1/2-\eta})$. The latter relation implies $\sum_{k=1}^{\infty} |a_k| < \infty$ in view of

$$\begin{aligned} \sum_{k=2}^{\infty} |a_k| &= \sum_{j=0}^{\infty} \sum_{l=2^{j+1}}^{2^{j+1}-1} |a_j| \\ &\leq \sum_{j=0}^{\infty} 2^{j/2} \left(\sum_{l=2^{j+1}}^{2^{j+1}-1} a_j^2 \right)^{1/2} \\ &\leq \sum_{j=0}^{\infty} 2^{j/2} \sqrt{A_{2^{j+1}}} = \sum_{j=0}^{\infty} 2^{j/2} O(2^{j(-1/2-\eta)}) < \infty. \end{aligned}$$

It is easily seen that the summability condition $\sum_{k=1}^{\infty} |a_k| < \infty$ does not imply the NED condition $A_{m+1}^{1/2} = O(m^{-1/2-\eta})$. To see this, let $a_n = 1/(n \log^2 n)$, $n \geq 2$. Then the NED condition is violated, and our summability condition of Theorem 2 is weaker. On the other hand, De Jong and Davidson's result has its advantage in that it can be applied to nonstationary processes.

Consider case (ii). In this case K is an Appell polynomial. Let $E(\varepsilon_i^4) < \infty$, $E(\varepsilon_i^2) = 1$, and $a_n = n^{-\delta}$, $n \geq 1$ with some $\delta > \frac{1}{2}$. Then the NED condition (15) requires $\delta > 1$. To this end, let $y_m = \sum_{i=0}^m a_i \varepsilon_{m-i}$ and $z_m = x_m - y_m$. Then $E(x_m^2 | \varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2m}) = E(z_m^2) + y_m^2$ and $x_m^2 - E(x_m^2 | \varepsilon_0, \varepsilon_1, \dots, \varepsilon_{2m}) = z_m^2 - E(z_m^2) + 2y_m z_m$. Elementary calculations show that there is a $c > 0$ such that $\|z_m^2 - E(z_m^2) + 2y_m z_m\| \sim cm^{1/2-\delta}$ as $m \rightarrow \infty$, which by (15) implies $\delta \geq 1 + \eta > 1$. However, by (ii) of Theorem 1, the functional limit theorem (11) holds under the weaker condition $\delta > \frac{3}{4}$ because $K(x) = x^2 - E(x_t^2)$ has power rank 2 in view of $K_{\infty}(w) = E(w + x_t)^2 - E(x_t^2) = w^2$, $K_{\infty}'(0) = 0$, and $K_{\infty}''(0) \neq 0$. The condition $\delta > \frac{3}{4}$ is much weaker, and it allows some long-range dependent sequences. It is optimal in the sense that the limiting distribution is the Rosenblatt process if $\frac{1}{2} < \delta < \frac{3}{4}$, as asserted by (i) of Theorem 1 or Avram and Taqqu (1987). Dittmann and Granger (2002) point out the similar phenomenon that the square of a Gaussian $I(d)$ process shows less dependence than the input process.

There is a substantial history regarding the asymptotic distributions of $S_n = S_n(K)$. Our results extend earlier ones in several aspects. Central and non-central limit theorems for $S_n(K)$ have been established for stationary Gaussian

processes (x_t) by Sun (1963), Taquq (1975, 1979), and Breuer and Major (1983), among others. In particular, Taquq (1975) established (10) for functionals of Gaussian processes. For non-Gaussian processes, the functional convergence (10) has been established for K with special forms. For example, Davydov (1970) considers $K(x) = x$, and Surgailis (1982) assumes that K is analytic and ε_0 has moments of all order. Appell polynomials are discussed in Avram and Taquq (1987) and Giraitis and Surgailis (1986). Surgailis (2000) considers the finite-dimensional convergence of (10) and assumes that either K is a polynomial or (x_t) is associated with some Gaussian processes. It is a difficult problem to derive limit theorems for S_n if K is not analytic and the linear process (x_t) is non-Gaussian. The difficulty is partly due to the fact that, in the non-Gaussian case, the associated Appell polynomials are no longer orthogonal. In the Gaussian case, they are Hermite polynomials and, hence, orthogonal (Taquq, 1975, 1979; Granger and Newbold, 1976; Giraitis and Surgailis, 1986). Ho and Hsing (1997) made a breakthrough and proved that $S_n/\sigma_{n,p} \Rightarrow \kappa_p Z_{p,\beta}(1)$, a marginal version of (10). See the latter paper for further references. The functional convergence is needed in unit root problems. Our Theorem 1 shows functional convergence for non-Gaussian processes under mild conditions on K . For other contributions see Wu (2002, 2003b), where noninstantaneous transforms and infinite variance linear processes are discussed.

3. ASYMPTOTICS OF THE UNIT ROOT STATISTICS

The functional limit theorems 1 and 2 easily lead to the asymptotic distributions of the unit root statistic $\hat{\rho}_n$, which are functionals of Hermite processes or Brownian motions as asserted by Theorems 3 and 4, respectively. We omit the proofs of the latter theorems because they routinely follow from the argument in Phillips (1987). Unfortunately, we know very little about the analytical properties of the limiting distribution in (iii) of Theorem 3. In comparison, Dickey and Fuller (1979) show that (4) has a nice representation. For statistical testing, quantiles of the limiting distribution in (iii) of Theorem 3 can be obtained by extensive simulations.

THEOREM 3. *Under assumption (i) of Theorem 1, we have as $n \rightarrow \infty$ that*

- (i) $(n\sigma_{n,p}^2)^{-1} \sum_{i=1}^n y_{i-1}^2 \Rightarrow \kappa_p^2 \int_0^1 Z_{p,\beta}^2(u) du$;
- (ii) $\sigma_{n,p}^{-2} \sum_{i=1}^n u_i y_{i-1} \Rightarrow \frac{1}{2} \kappa_p^2 Z_{p,\beta}^2(1)$;
- (iii) $n(\hat{\rho}_n - 1) \Rightarrow \frac{1}{2} Z_{p,\beta}^2(1) / \int_0^1 Z_{p,\beta}^2(u) du$.

THEOREM 4. *Under the assumptions of Theorem 2 or (ii) of Theorem 1, we have*

- (i) $n^{-2} \sum_{i=1}^n y_{i-1}^2 \Rightarrow \sigma^2 \int_0^1 B^2(u) du$;
(ii) $n^{-1} \sum_{i=1}^n u_i y_{i-1} \Rightarrow \frac{1}{2} \sigma^2 [B^2(1) - \gamma]$, where $\gamma = E(u_0^2)/\sigma^2$;
(iii) $n(\hat{\rho}_n - 1) \Rightarrow \frac{1}{2} [B^2(1) - \gamma] / \int_0^1 B^2(u) du$.

If $\kappa_1 \neq 0$, then the power rank of K is 1. Hence the asymptotic distribution asserted by Theorem 3 with $p = 1$ is the same as Sowell's result if $\frac{1}{2} < \beta < 1$. In this case, the asymptotic distributions are expressed as functionals of fractional Brownian motions. An interesting phenomenon happens when $p \geq 2$, as shown by Example 1.

Example 1

Let $K(w) = |w| - E|x_t|$. Assume that the density function f of x_t is symmetric, namely, $f(x) = f(-x)$. Then

$$K_\infty(w) = \int_{\mathbb{R}} |w+x|f(x) dx - \int_{\mathbb{R}} |x|f(x) dx = \int_{\mathbb{R}} |x|[f(x-w) - f(x)] dx.$$

Assume that $\int_{\mathbb{R}} |xf'(x)|dx < \infty$, $\int_{\mathbb{R}} |xf''(x)|dx < \infty$, and $\kappa_2 = \int_{\mathbb{R}} |x|f''(x) dx \neq 0$. Then the power rank of K is 2 because $\kappa_1 = \int_{\mathbb{R}} |x|f'(x) dx = 0$ in view of the symmetry of K . If $\frac{1}{2} < \beta < \frac{3}{4}$, then Theorem 3 and (i) of Theorem 1 are applicable. In this case, the limiting distribution is called the Rosenblatt process (Taqqu, 1975). On the other hand, if $\frac{3}{4} < \beta < 1$, then as in the classical cases, the limiting distributions are functionals of Brownian motions (Theorem 4).

REFERENCES

- Andrews, D.W.K. (1984) Non-strong mixing autoregressive processes. *Journal of Applied Probability* 21, 930–934.
- Avram, F. & M. Taqqu (1987) Noncentral limit theorems and Appell polynomials. *Annals of Probability* 15, 767–775.
- Baillie, R.T. (1996) Long memory processes and fractional integration in econometrics. *Journal of Econometrics* 73, 5–59.
- Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley.
- Bingham, N.H., C.M. Goldie, & J.L. Teugels (1987) *Regular Variation*. Cambridge University Press.
- Bradley, R. (1986) Basic properties of strong mixing conditions. In E. Eberlein & M.S. Taqqu (eds.), *Dependence in Probability and Statistics: A Survey of Recent Results*, pp. 165–192. Birkhauser.
- Breuer, P. & P. Major (1983) Central limit theorems for nonlinear functionals of Gaussian fields. *Journal of Multivariate Analysis* 13, 425–441.
- Caporale, G.M. & L.A. Gil-Alana (2004) Fractional co-integration and real exchange rates. *Review of Financial Economics* 13, 327–340.
- Chan, N.H. & N. Terrin (1995) Inference for unstable long-memory processes with applications to fractional unit root autoregressions. *Annals of Statistics* 20, 1662–1683.
- Davydov, Y.A. (1970) The invariance principle for stationary processes. *Theory of Probability and Its Applications* 15, 487–498.
- De Jong, R.M. & J. Davidson (2000) The functional central limit theorem and weak convergence to stochastic integrals, Part I: Weakly dependent processes. *Econometric Theory* 16, 621–642.

- Dickey, D.A. & W.A. Fuller (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–431.
- Dickey, D.A. & W.A. Fuller (1981) Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49, 1057–1072.
- Dittmann, I. & C.W.J. Granger (2002) Properties of nonlinear transformations of fractionally integrated processes. *Journal of Econometrics* 110, 113–133.
- Doukhan, P. (1994) *Mixing: Properties and Examples*. Springer.
- Doukhan, P., G. Oppenheim, & M.S. Taqqu (eds.) (2003) *Theory and Applications of Long-range Dependence*. Birkhauser.
- Gallant, A.R. & H. White (1988) *A Unified Theory for Estimation and Inference for Nonlinear Econometric Models*. Blackwell.
- Giraitis, L. & D. Surgailis (1986) Multivariate Appell polynomials and the central limit theorem. In E. Eberlein & M.S. Taqqu (eds.), *Dependence in Probability and Statistics: A Survey of Recent Results*, pp. 21–72. Birkhauser.
- Gorodetskii, V.V. (1977) On the strong mixing property for linear sequences. *Theory of Probability and Its Applications* 22, 411–412.
- Granger, C.W.J. & R. Joyeux (1980) An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis* 1, 15–30.
- Granger, C.W.J. & P. Newbold (1976) Forecasting transformed series. *Journal of the Royal Statistical Society, Series B* 38, 189–203.
- Hannan, E.J. (1979) The central limit theorem for time series regression. *Stochastic Processes and Their Applications* 9, 281–289.
- Ho, H.-C. & T. Hsing (1997) Limit theorems for functionals of moving averages. *Annals of Probability* 25, 1636–1669.
- Hosking, J. (1981) Fractional differencing. *Biometrika* 68, 165–176.
- Ibragimov, I.A. & Yu.V. Linnik (1971) *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff.
- Mandelbrot, B. & W.J. Van Ness (1969) Fractional Brownian motions, fractional noises and applications. *SIAM Review* 10, 422–437.
- Peligrad, M. (1986) Recent advances in the central limit theorem and its weak invariance principle for mixing sequences of random variables (A survey). In E. Eberlein & M.S. Taqqu (eds.), *Dependence in Probability and Statistics: A Survey of Recent Results*, pp. 193–224. Birkhauser.
- Pham, T.D. & L.T. Tran (1985) Some mixing properties of time series models. *Stochastic Processes and Their Applications* 19, 297–303.
- Phillips, P.C.B. (1987) Time series regression with a unit root. *Econometrica* 55, 277–301.
- Phillips, P.C.B. & Z. Xiao (1998) A primer on unit root testing. *Journal of Economic Surveys* 12, 423–469.
- Robinson, P.M. (2003) *Time Series with Long Memory*. Oxford University Press.
- Rosenblatt, M. (1956) A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the USA* 42, 43–47.
- Sowell, F. (1990) The fractional unit root distribution. *Econometrica* 58, 495–505.
- Sun, T.C. (1963) A central limit theorem for non-linear functions of a normal stationary process. *Journal of Mathematical Mechanics* 12, 945–978.
- Surgailis, D. (1982) Zones of attraction of self-similar multiple integrals. *Lithuanian Mathematical Journal* 22, 327–340.
- Surgailis, D. (2000) Long-range dependence and Appell rank. *Annals of Probability* 28, 478–497.
- Taqqu, M.S. (1975) Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 31, 287–302.
- Taqqu, M.S. (1979) Convergence of integrated processes of arbitrary Hermite rank. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 50, 53–83.
- Wang, Q.Y., Y.X. Lin, & C.M. Gulati (2002) The invariance principle for linear processes with applications. *Econometric Theory* 18, 119–139.

- Wang, Q.Y., Y.X. Lin, & C.M. Gulati (2003) Asymptotics for general fractionally integrated processes with applications to unit root tests. *Econometric Theory* 19, 143–164.
- Withers, C.S. (1981) Conditions for linear process to be strongly mixing. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 57, 477–480.
- Wu, W.B. (2002) Central limit theorems for functionals of linear processes and their applications. *Statistica Sinica* 12, 635–649.
- Wu, W.B. (2003a) Empirical processes of long-memory sequences. *Bernoulli* 9, 809–831.
- Wu, W.B. (2003b) Additive functionals of infinite-variance moving averages. *Statistica Sinica* 13, 1259–1267.

APPENDIX

We shall apply the central limit theory for Markov chains to prove our main results. Let $(\xi_n)_{n \in \mathbb{Z}}$ be a stationary and ergodic Markov chain; let $\mathcal{F}_i = (\dots, \xi_{i-1}, \xi_i)$ and define the projection operator

$$\mathcal{P}_i X = E(X|\mathcal{F}_i) - E(X|\mathcal{F}_{i-1}).$$

By the Markovian property, $\mathcal{P}_k g(\xi_i) = E[g(\xi_i)|\xi_k] - E[g(\xi_i)|\xi_{k-1}]$ if $k \leq i$.

LEMMA 1. *Assume that the Markov chain $(\xi_n)_{n \in \mathbb{Z}}$ is stationary and ergodic and the function g satisfies $E[g(\xi_1)] = 0$ and $E[g^2(\xi_1)] < \infty$. Let $S_n = \sum_{i=1}^n g(\xi_i)$. Further assume that*

$$\sigma_0 := \sum_{i=0}^{\infty} \|\mathcal{P}_0 g(\xi_i)\| < \infty. \tag{A.1}$$

Then $\|S_n\| \leq \sigma_0 \sqrt{n}$ and $\{S_{nu}/\sqrt{n}, 0 \leq u \leq 1\} \Rightarrow \{\sigma B(u), 0 \leq u \leq 1\}$ in $C[0,1]$, where $\sigma = \|\sum_{i=0}^{\infty} \mathcal{P}_0 g(\xi_i)\|$.

Lemma 1 is adapted from Hannan (1979), and it provides a useful tool for functional limit theorems for stationary processes. We shall apply it to $g(\xi_i) = u_i = K(x_i)$, where the shift process $\xi_i = \mathcal{F}_i = (\dots, \varepsilon_{j-1}, \varepsilon_j)$ is clearly a Markov chain. In such cases we are able to obtain bounds for $\|\mathcal{P}_0 g(\xi_i)\|$ (cf. Theorem 5, which follows), and thus condition (A.1) is verifiable.

Asymptotic Expansions. For $j \geq 2$ let $A_n(j) = \sum_{t=n}^{\infty} |a_t|^j$, $A_n = A_n(2)$, $\theta_n = |a_{n-1}|[|a_{n-1}| + A_n^{1/2}(q) + A_n^{p/2}]$, $\Theta_n = \sum_{k=1}^n \theta_k$, and $\Xi_{n,p} = n\Theta_n^2 + \sum_{i=1}^{\infty} (\Theta_{n+i} - \Theta_i)^2$. Define

$$S_n(L^{(p)}) = \sum_{i=1}^n L^{(p)}(\mathcal{F}_i),$$

where

$$L^{(p)}(\mathcal{F}_n) = K(x_n) - \sum_{r=0}^p \kappa_r U_{n,r} \quad \text{and} \quad U_{n,r} = \sum_{0 \leq j_1 < \dots < j_r < \infty} \prod_{s=1}^r a_{j_s} \varepsilon_{n-j_s}. \tag{A.2}$$

LEMMA 2. *Condition 1 implies that*

$$E[K_n^{(\alpha)}(x_{n,0})|\mathcal{F}_{-1}] = K_{n+1}^{(\alpha)}(x_{n,-1}) \quad \text{almost surely,} \quad \alpha = 0, 1, \dots, p; \quad (\text{A.3})$$

$$\|K_{n-1}^{(\alpha)}(x_{n,1}) - K_n^{(\alpha)}(x_{n,0}) - K_n^{(\alpha+1)}(x_{n,0})a_{n-1}\varepsilon_1\| = \mathcal{O}(|a_{n-1}|^{q/2}),$$

$$\alpha = 0, 1, \dots, p-1; \quad (\text{A.4})$$

and

$$\|K_n^{(p)}(x_{n,0}) - K_\infty^{(p)}(0)\|^2 = \mathcal{O}(A_n). \quad (\text{A.5})$$

Remark 1. If $\alpha = 0$, then (A.3) is valid by the smoothing property of conditional expectations. For larger α , differentiation under the expectation sign is required. A simple recursion yields that $E[K_n^{(\alpha)}(x_{n,0})|\mathcal{F}_{-i}] = K_{n+i}^{(\alpha)}(x_{n,-i})$ almost surely for all $i \geq 0$. Roughly speaking, (A.4) is a first-order Taylor's expansion.

Proof of Lemma 2. Assume without loss of generality that $\lambda = 1$. For $\alpha \leq p$ let $\delta = a_{n-1}\varepsilon_1$, $R_n^{(\alpha)} = K_{n-1}^{(\alpha)}(x_{n,1}) - K_{n-1}^{(\alpha)}(x_{n,0})$, and $T_n^{(\alpha)} = R_n^{(\alpha)} - K_{n-1}^{(1+\alpha)}(x_{n,0})\delta$. By Taylor's expansion, $|T_n^{(\alpha)}\mathbf{1}_{|\delta| \leq 1}| \leq \frac{1}{2}|\delta|^{q/2}K_{n-1}^{(\alpha+2)}(x_{n,0}; 1)$ because $\delta^2\mathbf{1}_{|\delta| \leq 1} \leq |\delta|^{q/2}$. On the other hand,

$$|T_n^{(\alpha)}\mathbf{1}_{|\delta| > 1}| \leq |R_n^{(\alpha)}\mathbf{1}_{|\delta| > 1}| + |K_{n-1}^{(1+\alpha)}(x_{n,0})\delta\mathbf{1}_{|\delta| > 1}| \leq |R_n^{(\alpha)}|\delta|^{q/2} + |K_{n-1}^{(1+\alpha)}(x_{n,0})|\delta|^{q/2}.$$

Hence $\|T_n^{(\alpha)}\| = \mathcal{O}(|a_{n-1}|^{q/2})$ by (8) and the independence between ε_1 and \mathcal{F}_0 , and $\|R_n^{(\alpha)}\| = \mathcal{O}(|a_{n-1}|)$. For (A.3), it suffices to show that

$$E[K_{n-1}^{(\alpha)}(x_{n,1})|\mathcal{F}_0] = K_n^{(\alpha)}(x_{n,0}) \quad (\text{A.6})$$

holds for $\alpha \leq p$. We shall use an induction argument. The case in which $\alpha = 0$ trivially follows. By letting $w \rightarrow 0$ in the identity

$$\frac{1}{w} E[K_{n-1}(w + x_{n,1}) - K_{n-1}(x_{n,1})|\mathcal{F}_0] = \frac{1}{w} [K_n(w + x_{n,0}) - K_n(x_{n,0})],$$

we have (A.6) with $\alpha = 1$ by the first term of (8) and the Lebesgue dominated convergence theorem. The general case $\alpha \geq 2$ follows recursively. Observe that for $\alpha < p$, by (A.6),

$$\|K_{n-1}^{(\alpha)}(x_{n,0}) - K_n^{(\alpha)}(x_{n,0})\| = \|E(T_n^{(\alpha)}|\mathcal{F}_0)\| \leq \|T_n^{(\alpha)}\| = \mathcal{O}(|a_{n-1}|^{q/2})$$

and similarly $\|K_{n-1}^{(p)}(x_{n,0}) - K_n^{(p)}(x_{n,0})\| = \mathcal{O}(|a_{n-1}|)$. Relation (A.4) follows from $\|T_n^{(\alpha)}\| = \mathcal{O}(|a_{n-1}|^{q/2})$, as does $\|K_{n-1}^{(\alpha)}(x_{n,1}) - K_n^{(\alpha)}(x_{n,0})\| = \mathcal{O}(|a_{n-1}|)$. Finally,

$$\|K_n^{(\alpha)}(x_{n,0}) - K_\infty^{(\alpha)}(0)\|^2 = \sum_{i=-\infty}^0 \|K_{n-i}^{(\alpha)}(x_{n,i}) - K_{n-i+1}^{(\alpha)}(x_{n,i-1})\|^2 = \sum_{i=-\infty}^0 \mathcal{O}(a_{n-i}^2)$$

entails (A.5) via the stationarity of x_n . ■

THEOREM 5 (Reduction principle). *Suppose that Condition 1 holds with $p \geq 1$. Then*

$$\|\mathcal{P}_1 L^{(p)}(\mathcal{F}_n)\| = \mathcal{O}(\theta_n). \quad (\text{A.7})$$

Proof. Let $N_n^{(\alpha)}(\mathcal{F}_0) = K_\infty^{(\alpha)}(0) + \sum_{r=1}^{p-\alpha} K_\infty^{(r+\alpha)}(0)E(U_{n,r}|\mathcal{F}_0)$ and define $M_n^{(\alpha)}(\mathcal{F}_0) = K_n^{(\alpha)}(x_{n,0}) - N_n^{(\alpha)}(\mathcal{F}_0)$. Then $E[L^{(p)}(\mathcal{F}_n)|\mathcal{F}_0] = M_n^{(0)}(\mathcal{F}_0)$. Observe that for $i \leq -1$,

$$\mathcal{P}_{i+1} N_n^{(\alpha)}(\mathcal{F}_0) = K_\infty^{(\alpha+1)}(0)a_{n-i-1}\varepsilon_{i+1} + \sum_{r=2}^{p-\alpha} K_\infty^{(r+\alpha)}(0)E(U_{n,r-1}|\mathcal{F}_i)a_{n-i-1}\varepsilon_{i+1}, \quad (\text{A.8})$$

which has the same distribution as $N_{n-i}^{(\alpha+1)}(\mathcal{F}_0)a_{n-i-1}\varepsilon_1$ via stationarity of \mathcal{F}_n . Now we claim by backward induction that for all $p \geq \alpha \geq 0$,

$$\|M_n^{(\alpha)}(\mathcal{F}_0)\|^2 = \mathcal{O}[A_n(q) + A_n^{p-\alpha+1}]. \quad (\text{A.9})$$

When $\alpha = p$, (A.9) follows from (A.5) because $M_n^{(\alpha)}(\mathcal{F}_0) = K_n^{(\alpha)}(x_{n,0}) - K_\infty^{(\alpha)}(0)$. Suppose (A.9) holds for $\alpha = m$, where $1 \leq m \leq p$, and consider the case $\alpha = m - 1$. So for $i \leq -1$,

$$\|K_{n-i}^{(\alpha+1)}(x_{n,-i})a_{n-i-1}\varepsilon_{i+1} - \mathcal{P}_{i+1} N_n^{(\alpha)}(\mathcal{F}_0)\|^2 = a_{n-i-1}^2 \mathcal{O}[A_{n-i}(q) + A_{n-i}^{p-m+1}] \quad (\text{A.10})$$

by the induction hypothesis and (A.8). By Lemma 2,

$$\|\mathcal{P}_{i+1} K_n^{(\alpha)}(x_{n,0}) - K_{n-i}^{(\alpha+1)}(x_{n,-i})a_{n-i-1}\varepsilon_{i+1}\| = \mathcal{O}(|a_{n-i-1}|^{q/2}) \quad (\text{A.11})$$

because $E[K_n^{(\alpha)}(x_{n,0})|\mathcal{F}_i] = K_{n-i}^{(\alpha)}(x_{n,-i})$. By (A.10), (A.11), and the orthogonality of \mathcal{P}_i ,

$$\begin{aligned} \|M_n^{(\alpha)}(\mathcal{F}_0)\|^2 &= \sum_{i=-\infty}^{-1} \|\mathcal{P}_{i+1} M_n^{(\alpha)}(\mathcal{F}_0)\|^2 \\ &= \sum_{i=-\infty}^{-1} \mathcal{O}(|a_{n-i-1}|^q + a_{n-i-1}^2 [A_{n-i}(q) + A_{n-i}^{p-m+1}]) \\ &= \mathcal{O}[A_n(q) + A_n^{p-m+2}]. \end{aligned}$$

Thus the induction is finished, and (A.7) follows by setting $\alpha = 0$ and $i = -1$. \blacksquare

Central and Noncentral Limit Theorems

COROLLARY 1. *Suppose that Condition 1 holds with $p \geq 1$. Then*

$$\|S_n(L^{(p)})\|^2 = \mathcal{O}(\Xi_{n,p}). \quad (\text{A.12})$$

In particular, let $|a_n| = \mathcal{O}[n^{-\beta}\ell(n)]$ with $\frac{1}{2} < \beta < 1$ and $q = 4$. Then (i) $\Xi_{n,p} = \mathcal{O}(n)$ if $(p+1)(2\beta-1) > 1$; (ii) $\Xi_{n,p} = \mathcal{O}[n^{2+(p+1)(1-2\beta)}\ell^{2(p+1)}(n)]$ if $(p+1)(2\beta-1) < 1$; (iii) $\Xi_{n,p} = \mathcal{O}\{n[\sum_{m=1}^n \ell^{p+1}(m)/m]^2\}$ if $(p+1)(2\beta-1) = 1$.

Proof. Observe that $\|\mathcal{P}_{i+1}L^{(p)}(\mathcal{F}_j)\| = \|\mathcal{P}_1L^{(p)}(\mathcal{F}_{j-i})\|$, which is zero if $j \leq i$. Hence

$$\begin{aligned} \|S_n(L^{(p)})\|^2 &= \sum_{i=-\infty}^{n-1} \|\mathcal{P}_{i+1}S_n(L^{(p)})\|^2 \\ &\leq \sum_{i=-\infty}^{n-1} \left\{ \sum_{j=1+\max(i,0)}^n \|\mathcal{P}_{i+1}L^{(p)}(\mathcal{F}_j)\| \right\}^2, \end{aligned}$$

by the orthogonality of \mathcal{P}_k . So (A.12) follows from (A.7). If $|a_n| = \mathcal{O}[n^{-\beta}\ell(n)]$, then (i)–(iii) follow from Lemma 5 in Wu (2003a) as easy applications of Karamata's theorem. \blacksquare

Remark 2. Corollary 1 goes beyond the important results by Ho and Hsing (1997) in several aspects. In particular, the imposed condition (8) is weaker and a_n is allowed to have forms other than $n^{-\beta}\ell(n)$. Moreover, if $|a_n| = \mathcal{O}[n^{-\beta}\ell(n)]$, then Corollary 1 gives a sharper bound. Ho and Hsing obtain the bound $\max(n, n^{2-(p+1)(2\beta-1)+\zeta})$ for any $\zeta > 0$. At a technical level, our induction argument appears much simpler, and it can be easily generalized to multiple linear processes. In addition, (A.12) does not require the finiteness of the fourth moment of ε_1 , whereas $E(\varepsilon_1^8) < \infty$ is needed in Ho and Hsing.

Proof of Theorem 1. (i) We generically say that (10) is a noncentral limit theorem because the asymptotic distribution is non-Gaussian if $p \geq 2$ and because the norming sequence $\sigma_{n,p}$ grows faster than \sqrt{n} , the norming sequence used in the class central limit theorem for i.i.d. random variables with finite variances. To prove (10), let $W_m = \sum_{i=1}^m U_{i,p}$, where $U_{n,r}$ is defined in (A.2). Because $p(2\beta - 1) < 1$, by Lemma 5 in Surgailis (1982) or Theorem 2 in Avram and Taqqu (1987), $\{W_m/\sigma_{n,p}, 0 \leq t \leq 1\} \Rightarrow \{Z_{p,\beta}(t), 0 \leq t \leq 1\}$ in $C[0,1]$. Note that $S_m(L^{(p)}) = S_m(K) - \kappa_p W_m$. Recall Corollary 1 for the bound of $\Xi_{n,p}$. By considering three cases $(p+1)(2\beta-1) < 1$, $(p+1)(2\beta-1) > 1$, and $(p+1)(2\beta-1) = 1$ separately, it is easily seen that $\max_{k \leq n} \Xi_{k,p} = o(\sigma_{n,p}^2)$ because $p(2\beta-1) < 1$. Thus the finite-dimensional convergence of $S_m/\sigma_{n,p}$ to $K_\infty^{(p)}(0)Z_{p,\beta}(t)$ holds. It remains to verify the tightness. By Theorem 12.3 in Billingsley (1968), we need to show that there exist $C < \infty$ and $\gamma > 1$ such that

$$\frac{\|S_k\|^2}{\sigma_{n,p}^2} \leq C \frac{k^\gamma}{n^\gamma} \tag{A.13}$$

holds for all $n \geq 1$ and $1 \leq k \leq n$. Because $\Xi_{n,p} = o(\sigma_{n,p}^2)$, $\|S_m\| \sim |\kappa_m| \sigma_{m,p}$ as $m \rightarrow \infty$. Let $\gamma = \frac{3}{2} - p(\beta - \frac{1}{2}) > 1$. Then by elementary properties of slowly varying functions,

$$\lim_{n \rightarrow \infty} \max_{k \leq n} \frac{n^\gamma \sigma_{k,p}^2}{k^\gamma \sigma_{n,p}^2} = \lim_{n \rightarrow \infty} \max_{k \leq n} \frac{k^{\gamma-1} \ell^{2p}(k)}{n^{\gamma-1} \ell^{2p}(n)} = 1,$$

which entails (A.13) and completes the proof of part (i).

(ii) In this case it is interesting to observe that the limiting distribution in (11) is Brownian motion even though (x_t) is long-range dependent. To prove (11), by Lemma 1, it suffices to show that $\sum_{n=1}^{\infty} \|\mathcal{P}_0 L^{(p)}(\mathcal{F}_n)\| < \infty$ and $\sum_{n=1}^{\infty} \|\mathcal{P}_0 U_{n,p}\| < \infty$ because $g(x_n) = L^{(p)}(\mathcal{F}_n) + \kappa_p U_{n,p}$. The former easily follows from Theorem 5 under the pro-

posed conditions of (a_n) . As to the latter, observe that $\|\mathcal{P}_0 U_{n,p}\|^2 \leq a_n^2 A_n^{p-1} (E(\varepsilon_1^2))^p$. Hence $\|\mathcal{P}_0 U_{n,p}\| = \mathcal{O}[n^{-\beta+(1-2\beta)(p-1)/2} \ell^p(n)]$ is also summable over n . ■

Proof of Theorem 2. Let (ε'_n) be an i.i.d. copy of (ε_n) and $x'_{n,1} = x_{n,1} + a_{n-1}(\varepsilon'_1 - \varepsilon_1)$. Then $E[K_{n-1}(x'_{n,1})|\mathcal{F}_0] = K_n(x_{n,0})$ and (13) follows from Lemma 1 in view of

$$\begin{aligned} \|\mathcal{P}_1 K(x_n)\| &= \|K_{n-1}(x_{n,1}) - K_n(x_{n,0})\| = \|E[K_{n-1}(x_{n,1}) - K_{n-1}(x'_{n,1})|\mathcal{F}_0]\| \\ &\leq \|K_{n-1}(x_{n,1}) - K_{n-1}(x'_{n,1})\| \leq 2\|K_{n-1}(x_{n,1}) - K_{n-1}(x_{n,0})\| \\ &= \mathcal{O}(|a_{n-1}|). \end{aligned}$$

That (14) entails (12) follows from the argument in the proof of Lemma 2. ■