Weak Unit Roots[1]

Joon Y. Park
Department of Economics
Rice University

Abstract

This paper develops the large sample theory for econometric models with time series having roots in proximity of unity. In particular, a special attention is given to the time series with roots outside the n^{-1}-neighborhood of unity, called the *weak* unit roots. They are the processes with roots approaching to unity as sample size increases, but not too fastly. It is shown that the weak unit root processes yield the standard law of large numbers and central limit theorem-like results, and as a consequence, the usual large sample theory of inference based on normal asymptotics is applicable for models with weak unit root processes. This suggests that we may rely on the conventional statistical theory also for models with roots close to unity, as long as the roots are not too close to unity. In practice, it seems that we may safely use the standard normal theory, unless the roots are very close to one in a metric proportional to the magnitude of sample size. We consider a wide class of models including autoregressions and nonlinear, as well as linear, cointegrated models with weak unit roots.

First Draft: February 25, 2002
This version: August 1, 2003

Key words and phrases: Unit roots, large sample theory, Brownian motion, Ornstein-Uhlenbeck process, local time, autoregressions with unit roots, linear and nonlinear cointegrated models

[1] I am very grateful to Yoosoon Chang for many helpful discussions and constructive suggestions, and to Bruce Hansen, Bennoit Perron and the participants at the 2003 North American Econometric Society Summer Meeting, Evanston, for useful comments.

1. Introduction

As is well known, the asymptotic theories for models with unit root processes are quite distinct from those for models with stationary processes. The standard normal asymptotics apply for the stationary models, while the asymptotics for the nonstationary models are generally nonstandard and nonnormal. Therefore, the applied time series econometricians always have to determine which of the two sets of asymptotic theories is more appropriate to use for his or her dataset. They often find that it is not an easy task, since many economic and financial time series have roots in a quite close neighborhood of unity. This has been widely noted particularly in conjunction with the empirical researches on the predictability of stock returns and exchange rates. Indeed, most variables that appear to be potentially useful for the predictions of stock returns and exchange rates have roots close to one.

The sharp discontinuity in the statistical theories of stationary and unit root models, of course, is something that exists only in the asymptotics. In finite samples, the distributions change continuously as the roots approach to unity. Therefore, arises an important practical question: How close the roots should be to the unity for the unit root asymptotics to be more appropriate, or equivalently, how distant the roots need to be from the unity for the standard normal limit theory to be more applicable. Obviously, there cannot be any single good answer to the question. It should depend on, among many others, the data at hand, the models and the inferences to be made. However, there is one important factor which matters in all cases: the sample size. The sample size indeed plays a critical role here, as we demonstrate clearly below.

We may formally investigate the dependency on the sample size n of the relevant asymptotics by modelling the largest autoregressive root of the underlying time series as a function of n. When the root is set by $1 - cn^{-1}$ with some constant $c > 0$, the underlying time series has the so-called *near* unit root, and this has been studied earlier by many authors as the local alternatives to the exact unit root. In the paper, we focus on the case where the root approaches to unity at a rate slower than n^{-1}, and say that the resulting time series have *weak* unit roots or are *weakly* integrated. The weak unit root processes have the root outside the n^{-1}-neighborhood of unity, and are contrasted with the exact or near unit root processes that have the root inside or on the boundary of the n^{-1}-neighborhood of unity. The weak unit root processes are the processes with the root approaching to unity as the sample size increases, but not too fastly.

For the exactly and nearly integrated processes, the asymptotics are nonstandard and nonnormal. However, the weakly integrated processes yield asymptotics that are drastically different. They obey standard law of large numbers and central limit theorem-like results, and therefore, the usual large sample theory of inference based on normal asymptotics holds for them. This implies in particular that the size of the unit root neighborhoods governed by the nonstandard and nonnormal asymptotics shrinks as the sample size increases, and we have normal distribution theory applicable everywhere outside of those shrinking neighborhoods. Actually, our theory shows that the size of the neighborhoods on which the normal asymptotics fail to work decreases at n^{-1} rate. This suggests that we may rely on the conventional statistical theory also for models with roots close to unity, as long as the roots are not too close to unity in a metric proportional to the magnitude of sample size.

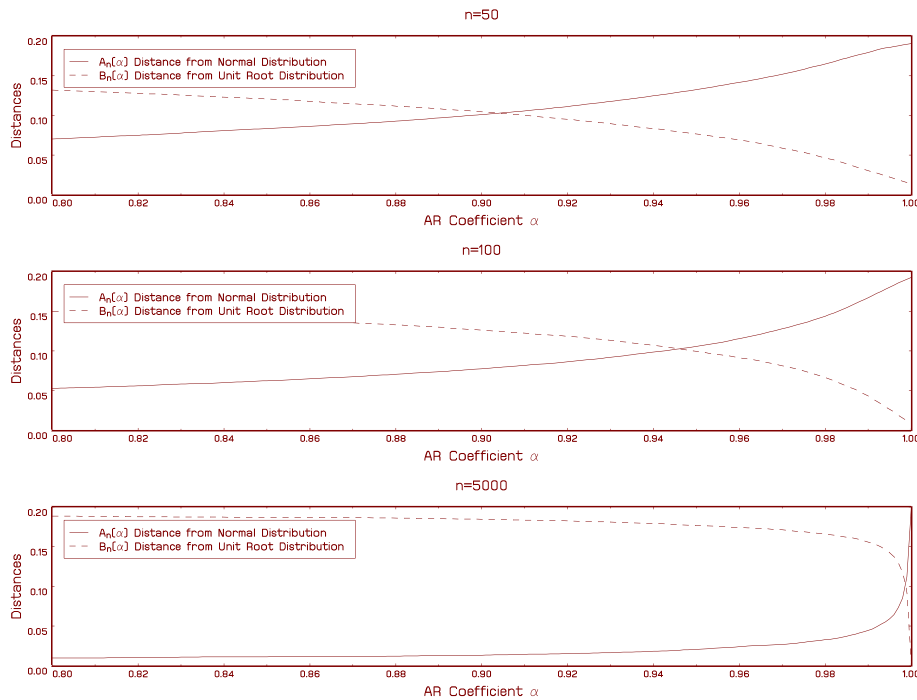

Figure 1: Distances from Normal and Unit Root Asymptotics of t-Tests in AR(1) Models

Figure 1 shows how distant are the finite sample distributions of the t-test in the Gaussian AR(1) models from the standard normal and also from the unit root distributions. For the samples of sizes $n = 50, 100$ and $5,000$ and for the autoregressive coefficient $\alpha \in [0.8, 1]$, the distances from the finite sample distribution function $F_n(\cdot, \alpha)$ of the t-statistic to the standard normal distribution function Φ and to the unit root distribution function Ψ are computed using the uniform metric. More precisely, their distances are defined by $A_n(\alpha) = \sup_{x \in \mathbf{R}} |F_n(x, \alpha) - \Phi(x)|$ and $B_n(\alpha) = \sup_{x \in \mathbf{R}} |F_n(x, \alpha) - \Psi(x)|$, and plotted as functions of α. As $n \to \infty$, we have $F_n(x, \alpha) \to \Phi(x)$ for all $|\alpha| < 1$ and $F_n(x, \alpha) \to \Psi(x)$ for $|\alpha| = 1$ uniformly in $x \in \mathbf{R}$, and consequently, it follows that $A_n(\alpha) \to 0$ for all $|\alpha| < 1$ and $B_n(\alpha) \to 0$ for $|\alpha| = 1$. The distance functions $A_n(\alpha)$ and $B_n(\alpha)$ are monotone in α for all n.

In Figure 1, we may see especially over what ranges of the α values which of the two competing asymptotics, the normal and the unit root asymptotics, is more appropriate to use. As is clearly demonstrated, the answer depends critically upon the size n of the samples. In general, the range of α's for which the normal asymptotics yield better approximations expands as n increases. When $n = 50$, the finite sample distributions of the t-tests are closer to the standard normal distribution until the values of α get roughly bigger than 0.904. The range, however, becomes much larger if the sample size gets as large as $n = 5,000$. In this case, the normal asymptotics work better than the unit root asymptotics for all values of α except for those belonging to an immediate neighborhood of unity, i.e., the standard normal distribution provides better approximations for all α's smaller than roughly 0.999.

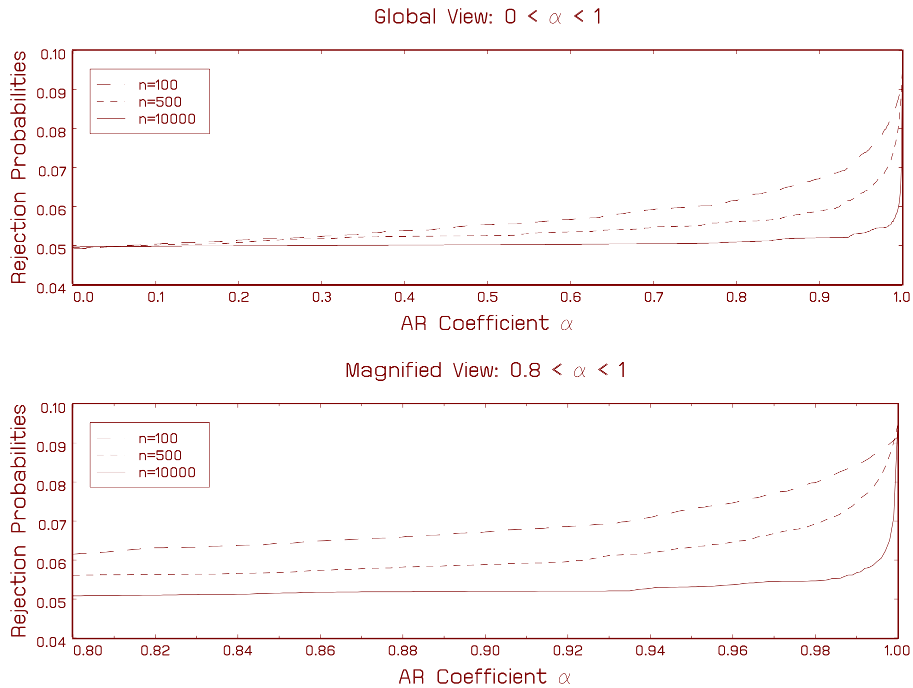

Figure 2: Rejection Probabilities of 5% t-Tests in AR(1) Models

Figure 2 presents the simulated rejection probabilities of the 5% t-tests once again in the Gaussian AR(1) models, for the samples of sizes $n = 100, 500$ and $10,000$. The critical value of -1.645 is used, which would yield the exact 5% rejection probability if the normal asymptotics work. As is well expected, the size distortions increase as the AR coefficient α approaches to unity. This is so for the samples of all sizes. However, the range of α's yielding noticeable distortions shrinks as the sample size increases. For $n = 100$, the actual rejection probabilities become 6% and 7% when α's are are approximately 0.745 and 0.935, respectively. However, for $n = 10,000$, they are below 6% all the way until α gets as big as 0.995, and become 7% only when α exceeds 0.999. For large samples, the normal asymptotics indeed work reasonably well, unless α is truly close to one.

Our findings in Figures 1 and 2 continue to hold for time series more general than AR(1). For the linear processes with the largest autoregressive root α close to unity, for instance, we may obtain the results quite similar qualitatively to those obtained for the simple AR(1). Moreover, the reported results for the simple AR(1) give some unambiguous clues to the asymptotics for more general regression models. The asymptotics for the regression models with unit root processes can in general be represented as the mixtures of the normal and unit root asymptotics, where the magnitude of the latter is proportional to the longrun correlation ρ between the innovations of the regressors and the regression errors. The autoregressions are governed solely by the unit root asymptotics, and this amounts to the case $|\rho| = 1$. For the case of the general regression models with $|\rho| < 1$, we may therefore well expect that the unit root distribution would be even less prevailing in their asymptotics.

The large sample theory established in the paper well explains our findings in Figures 1 and 2. We consider a wide class of econometric models including autoregressions and nonlinear regression models with weakly integrated processes. Though we only explicitly investigate their prototypical forms, our theory extends well, at least qualitatively, to other more general models. The asymptotics in the paper are derived in a quite sophisticated and unconventional manner. Yet, the main implication of our theoretical results is simple and straightforward: The usual inference based on the normal asymptotics continue to be valid even in the presence of the processes that are highly persistent, unless their largest autoregressive root is not too close to unity and the sample size is reasonably large. How close is 'not too close' and how large is 'reasonably large' should of course vary from one case to another, but our simulation results in Figures 1 and 2 give some useful general guidance.

The rest of the paper is organized as follows. Section 2 introduces the model and preliminaries. The weak unit root is formulated precisely, and the preliminary results that are necessary for the development of the subsequent theories are introduced. We present the basic asymptotic theories for the weak unit root processes in Section 3. Their asymptotics on nonlinear models are crucially dependent upon the specificity of the nonlinearity involved, and therefore, we introduce the classes of nonlinear functions that are used most frequently in practical applications. The asymptotic theories for models with weak unit root processes are developed in Section 4. In particular, we consider autoregressions and general nonlinear regressions explicitly there. Section 5 contains some concluding remarks, and the mathematical proofs for the theorems are given in Section 6.

2. The Model and Preliminaries

Consider the time series (x_t) generated as

$$x_t = \alpha x_{t-1} + v_t, \tag{1}$$

where (v_t) is a stationary linear process. We may allow the initial value x_0 of (x_t) to be any random variable as long as it is stochastically bounded. For expositional brevity, however, we assume that $x_0 = 0$ a.s. throughout the paper. We let

$$\alpha = 1 - \frac{m}{n}, \tag{2}$$

where n denotes the sample size and m is given as a function of n such that

$$\frac{m}{n} \to 0 \tag{3}$$

as $n \to \infty$.

We need to further specify m in our formulation of α in (2). Note that m is given as a function of n in (3).[2] It may therefore be more appropriately denoted by m_n, though we will not do so in the paper to simplify the exposition. Regardless of how we set m, α converges

[2]For example, we may set $m = \log n$ and our subsequent analysis is valid for such a specification of m.

to unity under the condition (3). However, the exact convergence rate is determined by the specification of m. In this paper, we will mainly look at the case

$$m \to \infty \tag{4}$$

as $n \to \infty$. Obviously, $\alpha = 1$ for all values of n if we set $m = 0$. On the other hand, if we let $m = c$ for some constant $c > 0$, then α converges to unity at n^{-1} rate. This specification yields what we usually refer to as the near unit root, or the root local-to-unity. It has widely been used to formulate the local altermatives to the unit root.[3] Clearly, α converges to unity at a rate slower than n^{-1} under the specification of m as in (4).

The time series (x_t) given by (1)–(3) represents a process with the root approaching to unity as the sample size gets large. With the additional specification of m in (4), it becomes a process with the root converging to unity more slowly than the process with the near unit root. Such a process is said to have the *weak* unit root, or to be *weakly* integrated. The condition (4) makes the process have a root outside the n^{-1}-neighborhood of unity, constrastingly with the exact and near unit root processes that have roots respectively inside and on the boundary of the n^{-1}-neighborhood of unity. The theories for weak unit root processes should thus be more appropriate in finite samples for models with roots close, but not too close, to unity. As we will show in the paper, the process with weak unit root has limit theories that are quite different from those for the processes with exact or near unit root. In particular, the latter has the nonstandard and nonnormal asymptotics, while the standard normal limit theories apply for the former.

We now define (v_t) more specifically as

$$v_t = \pi(L)\,\varepsilon_t = \sum_{k=0}^{\infty} \pi_k \varepsilon_{t-k}, \tag{5}$$

where (ε_t) is a sequence of independent and identically distributed random variables with mean zero, and $\pi(1) \neq 0$. Under specification (5), the time series (x_t) introduced in (1)–(3) becomes a general linear process with a weak unit root, including an autoregressive-weakly-integrated-moving-average, or ARWIMA, process as a special case. We assume

Assumption 2.1 $\sum_{k=0}^{\infty} k|\pi_k| < \infty$ and $\mathbf{E}|\varepsilon_t|^p < \infty$ for some $p > 2$.

For some of our subsequent results, we also need some additional assumptions on the distribution of (ε_t) as in

Assumption 2.2 The distribution of (ε_t) is absolutely continuous with respect to the Lebesgue measure, and has characteristic function φ for which $\lim_{t\to\infty} t^{\delta}\varphi(t) = 0$ for some $\delta > 0$.

The time series (v_t) has the longrun variance given by $\omega^2 = \pi(1)^2 \mathbf{E}\varepsilon_t^2$. Throughout the paper, we set this value to be unity, unless stated otherwise. This is to avoid unnecessary

[3] See, e.g., Phillips (1987) and Stock (1994) for more details on the motivation and analyses of nearly integrated processes.

complications in presenting our theoretical results. The longrun variance of (v_t) only has an unimportant scaling effect on our subsequent analysis.

We define

$$V_{mn}(r) = n^{-1/2} x_{[nr]+1} \tag{6}$$

for $r \in [0,1]$, where $[z]$ denotes the largest integer which does not exceed z. Moreover, we let

$$V_m(r) = \int_0^r \exp(-m(r-s))dV_0(s) \tag{7}$$

for $r \in [0,1]$, where V_0 is the standard Brownian motion. Then V_{mn} in (6) may be redefined, up to the distributional equivalence, on the same probability space as V_m in (7) so that V_{mn} and V_m are arbitrarily close for all large n in proability uniformly in $m \in \mathbf{R}_+$.[4] Indeed, we have

Lemma 2.3 Under Assumption 2.1, we may define V_{mn} and V_m on a common probability space so that

$$\sup_{0\leq r\leq 1} |V_{mn}(r) - V_m(r)| = o_p(n^{-1/2+1/p}) + O_p(mn^{-1})$$

for large n, uniformly in m such that $m/n \to 0$ as $n \to \infty$.

Due to Lemma 2.3, the sample moments of various functions of time series (x_t) can now be approximated, up to the distributional equivalence, by the integrals of the same functions of continuous process V_m. It is important to note that the approximations here can be made so that they have errors small uniformly in $m \in \mathbf{R}_+$ for all large n. Therefore, the required n-asymptotics for the sample moments of (x_t) under various transformations can be obtained directly from the m-asymptotics applied to the corresponding functionals of V_m. For each m fixed, V_m is an Ornstein-Uhlenbeck process that can be defined by the stochastic differential equation

$$dV_m(r) = -mV_m(r)\,dr + dW(r) \tag{8}$$

with the initial condition $V_m(0) = 0$, where W is the standard Brownian motion. In the paper, we refer to V_m as the Ornstein-Uhlenbeck process with parameter m.

To develop the m-asymptotics on the functionals of V_m, we need a proper normalization for V_m. Therefore, we introduce a normalized process V

$$V(r) = \sqrt{m}V_m\left(\frac{r}{m}\right) \tag{9}$$

which is defined from V_m by rescaling both its time and value. It follows immediately from (8) that the normalized process V is the Ornstein-Uhlenbeck process with unit parameter. It is well known that V has a stable stationary marginal distribution. In fact, the distribution

[4]Here and elsewhere in the paper, we do not distinguish two processes that are distributionally equivalent. Therefore, in general, equality means distributional equality, and both almost sure convergence and convergence in probability just imply convergence in distribution. They become convergence in probability only when limits are nonrandom, since then convergences in probability and in distribution become identical.

of $V(r)$ approaches to normal with mean 0 and variance 1/2 as $r \to \infty$, whose density we will denote by D throughout the paper.

Our asymptotics heavily rely on the local time of the normalized process V in (9), which is defined by

$$L(r,x) = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_0^r 1\{|V(s) - x| < \epsilon\}\, ds.$$

Roughly, $L(r,x)$ measures the rate of time spent by V, up to time r, in an immediate neighborhood of x. The concept of local time yields the so-called occupation times formula

$$\int_0^r T(V(s))\, ds = \int_{-\infty}^{\infty} T(x) L(r,x)\, dx \tag{10}$$

for any locally integrable function T on $\mathbf{R}$. The local time L is continuous with respect to both parameters x and r.

For the m-asymptotics, we eventually need to analyze the asymptotic behavior of $L(m,\cdot)$ as $m \to \infty$. For this purpose, we define

$$D_m(x) = \frac{L(m,x)}{m}.$$

Then it follows that

Lemma 2.4 We have

$$D_m(x) = D(x) + o(m^{-1/2} \log m \log\log m) \text{ a.s.}$$

uniformly over any compact interval, and for any $k > -1$

$$\int_{-\infty}^{\infty} |x|^k D_m(x)\, dx \to_{a.s.} \int_{-\infty}^{\infty} |x|^k D(x)\, dx$$

as $m \to \infty$.

Lemma 2.4 gives the asymptotics for the local time L of the Ornstein-Uhlenbeck process V with unit parameter. For such an Ornstein-Uhlenbeck process, the time average of local time approaches to its stable marginal density. This sharply contrasts with the asymptotic behavior of the local times of nonstationary processes. For instance, for Brownian motion, the local time is of stochastic order given by the square root of the progressing time, and if normalized, it remains to be random in the limit. For the local time of an Ornstein-Uhlenbeck process, we also have the central limit theorem as well as the law of large numbers given in Lemma 2.4. It is indeed established in Bosq (1999) that $\sqrt{m}(D_m(x) - D(x))$ converges in distribution to normal law as $m \to \infty$. Lemma 2.4 above gives the moment-type convergences.

3. Function Classes and Basic Asymptotics

In this section, we develop the basic asymptotics for weakly integrated processes under various transformations. Let (x_t) be a weakly integrated process defined in Section 2, and introduce an additional time series (u_t) satisfying

Assumption 3.1 Let (u_t) be a martingale difference sequence with respect to some filtration $(\mathcal{F}_t)$ such that
(a) (x_t) is adapted to $(\mathcal{F}_{t-1})$, and
(b) $\mathbf{E}(u_t^2|\mathcal{F}_{t-1}) = \sigma^2$ a.s. for all t, and $\sup_t \mathbf{E}(|u_t|^{2+\epsilon}|\mathcal{F}_{t-1}) < \infty$ a.s. for some $\epsilon > 0$.

For a general real-valued function F on $\mathbf{R}$, we will consider the asymptotics of

$$\sum_{t=1}^{n} F(x_t) \quad \text{and} \quad \sum_{t=1}^{n} F(x_t)\, u_t \tag{11}$$

upon appropriate normalizations. They are called the *mean* and *covariance* asymptotics, respectively. For the transformation function, we consider two classes of functions, integrable and asymptotically homogeneous functions, satifying some regularity conditions.[5]

3.1 Asymptotics for Integrable Transformations

To derive the mean and covariance asymptotics for the sample moments in (11), we need to introduce the regularity conditions for $T = F$ and F^2. Let T be integrable, and define

Definition 3.2 Let $T : \mathbf{R} \to \mathbf{R}$. We say that T is *regularly integrable* if
(a) T is piecewise Lipschitz, and
(b) $\int_{-\infty}^{\infty} |x|^q |T(x)|\, dx < \infty$ for some $q \geq (p-2)/6$.

Regularly integrable function T should therefore be piecewise Lipschitzian and have a fast enought decaying rate. Here we require that the tail of T decrease at a faster rate as the order of existing moments for the innovations increases. We may obtain finer asymptotics if the order of existing moments increases. To do so, however, the tail of T should decay accordingly at a faster rate.

The regularity conditions in Definition 3.2 allow us to develop the basic asymptotics for the integrable transformations of weakly integrated processes. It follows immediately from our definition of V_{mn} in (6) that

$$\frac{1}{\sqrt{nm}} \sum_{t=1}^{n} T(x_t) = \sqrt{\frac{n}{m}} \int_0^1 T(\sqrt{n} V_{mn}(r)) dr, \tag{12}$$

which we may further approximate as in

[5] These are roughly the same classes of functions considered in Park and Phillips (1999) for their study on the nonlinear models with exactly integrated processes. Our regularity conditions are, however, not directly comparable to theirs. On the one hand, we could relax some of their restrictive assumptions by refining their proofs. On the other hand, we need to impose somewhat stronger conditions to obtain the results that are applicable for models with weakly, as well as exactly, integrated processes.

Lemma 3.3 Let Assumptions 2.1 and 2.2 hold. If T is regularly integrable, then

$$\sqrt{\frac{n}{m}}\int_0^1 T(\sqrt{n}V_{mn}(r))dr = \sqrt{\frac{n}{m}}\int_0^1 T(\sqrt{n}V_m(r))dr + o_p(1)$$

for large n, uniformly in m such that $m = o\,(n^{1-2/p} \wedge n^{2/3})$.

Moreover, it follows from the successive applications of normalization (9), occupation times formular (10), change-of-variables for integrals, and Lemma 2.4 that

$$\begin{aligned}
\sqrt{\frac{n}{m}}\int_0^1 T\big(\sqrt{n}V_m(r)\big)\,dr &= \frac{1}{m}\sqrt{\frac{n}{m}}\int_0^m T\left(\sqrt{\frac{n}{m}}V(r)\right)dr \\
&= \frac{1}{m}\sqrt{\frac{n}{m}}\int_{-\infty}^{\infty} T\left(\sqrt{\frac{n}{m}}x\right) L(m,x)\,dx \\
&= \frac{1}{m}\int_{-\infty}^{\infty} T(x)L\left(m,\sqrt{\frac{m}{n}}x\right)dx \\
&\to_{a.s.} D(0)\int_{-\infty}^{\infty} T(x)\,dx \qquad (13)
\end{aligned}$$

as $m \to \infty$ and $m/n \to 0$. Consequently, we may easily deduce from (12) and (13), together with Lemma 3.3, that

$$\frac{1}{\sqrt{nm}}\sum_{t=1}^n T(x_t) \to_p D(0)\int_{-\infty}^{\infty} T(x)\,dx \qquad (14)$$

as $n \to \infty$, if $m = o\,(n^{1-2/p} \wedge n^{2/3})$ and $m \to \infty$ as $n \to \infty$.

The mean and covariance asymptotics for the sample moments in (11) under integrable transformations may now be easily developed if we apply the result in (14) for $T = F$ and $T = F^2$.

Theorem 3.4 Let Assumptions 2.1, 2.2 and 3.1 hold. If F is regularly integrable, and if $m = o\,(n^{1-2/p} \wedge n^{2/3})$ and $m \to \infty$ as $n \to \infty$, then we have

$$\begin{aligned}
\frac{1}{\sqrt{nm}}\sum_{t=1}^n F(x_t) &\to_p D(0)\int_{-\infty}^{\infty} F(x)\,dx, \\
\frac{1}{\sqrt[4]{nm}}\sum_{t=1}^n F(x_t)\,u_t &\to_d \mathbf{N}\left(0, \sigma^2 D(0)\int_{-\infty}^{\infty} F(x)^2 dx\right)
\end{aligned}$$

as $n \to \infty$.

If F is regularly integrable, then so is F^2. To see this, note in particular that a regularly integrable function is bounded.

Our results in (12) and Lemma 3.3 are applicable also for fixed m. If we let $m = c$ with some fixed constant c, we may therefore easily deduce that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} F(x_t) \to_d L_c(1,0)\int_{-\infty}^{\infty} F(x)\,dx, \tag{15}$$

$$\frac{1}{\sqrt[4]{n}}\sum_{t=1}^{n} F(x_t)\,u_t \to_d \mathbf{MN}\left(0, \sigma^2 L_c(1,0)\int_{-\infty}^{\infty} F(x)^2 dx\right) \tag{16}$$

as $n \to \infty$, where L_c is the local time of the Ornstein-Uhlenbeck process with parameter c. If $c > 0$, (15) and (16) provide the asymptotics for near unit root processes. If $c = 0$, they are reduced to the unit root asymptotics obtained earlier by Park and Phillips (1999). Recall that Ornstein-Uhlenbeck process V_c becomes Brownian motion if $c = 0$.

It is interesting to compare our asymptotics in Theorem 3.4 with those for the time series with exact and near unit roots given by (15) and (16). The mean asymptotics for weakly integrated processes involve nonrandom limits, and this contrasts with those for exactly and nearly integrated processes that are random. Moreover, the covariance asymptotics for weakly integrated processes have normal limiting distributions, whereas those for exactly and nearly integrated processes have mixed normal limiting distributions with the mixing variates given by the local times of Brownian motion and Ornstein-Uhlenbeck process. The mean and covariance asymptotics for weakly integrated processes respectively yield nonrandom probability limits and limiting normal distributions, analogously as LLN and CLT for the usual asymptotics for stationary processes.

3.2 Asymptotics for Asymptotically Homogeneous Transformations

The concept of asymptotically homogeneous function was first introduced by Park and Phillips (1999) in their development of the asymptotics for nonlinear transformations of exactly integrated time series. An asymptotically homogeneous function T behaves asymptotically like a homogeneous function, and can roughly be written for large λ as $T(\lambda\cdot) \approx \kappa(\lambda)S(\cdot)$ with S locally integrable. We call κ and S, respectively, the asymptotic order and the limit homogeneous function of T. We first introduce the required conditions for the class of locally integrable functions that define asymptotically homogeneous functions.

Definition 3.5 Let $S : \mathbf{R}\backslash\{0\} \to \mathbf{R}$. We say that S is *regular* if, for any $\epsilon > 0$ sufficiently small, it satisfies that

(a) for all $|x| \geq \epsilon$ and $|x - y| \leq \epsilon/2$, $|S(x) - S(y)| \leq K_{ab}(x)\,|x - y|$ with $K_{ab}(x) = K(1 + |x|^a)(1 + |x|^b)$, where $a > 0$, $b < 0$ and K are some constants not depending upon ϵ, and

(b) for all $|x| < \epsilon$, $|S(x)| \leq K|x|^c$ with some constants $c > -1$ and K independent of ϵ.

We say that S is regular in the *second-order* if both S and S^2 are regular.

A regular function may have a pole-type discontinuity at the origin. Loosely put, a function becomes regular if, near the origin, it is divergent at a slower rate than the reciprocal function and Lipschitz with Lipschitz constant increasing possibly only at polynomial rates near the origin and at infinity.

We now introduce the regularity conditions for the class of asymptotically homogeneous functions.

Definition 3.6 Let $T : \mathbf{R} \to \mathbf{R}$ be written as

$$T(\lambda x) = \kappa(\lambda) S(x) + R(x, \lambda).$$

We say that T is *regularly homogeneous* if

(a) S is regular in the second-order, and

(b) $|R(x, \lambda)| \leq \varpi(\lambda) Q(x)$ for all λ sufficiently large and for all x over any compact set, where $(\kappa^{-1}\varpi)(\lambda) \to 0$ as $\lambda \to \infty$ and Q is regular in the second-order.

For an asymptotically homogeneous function T with asymptotic order κ and limit homogeneous function S, it follows immediately from Definition 3.6 that

$$\frac{1}{n}\kappa\left(\sqrt{\frac{n}{m}}\right)^{-1} \sum_{t=1}^{n} T(x_t) \approx \frac{1}{n}\sum_{t=1}^{n} S\left(\sqrt{\frac{m}{n}} x_t\right). \tag{17}$$

The asymptotics of weakly integrated time series under asymptotically homogeneous transformations can therefore be easily obtained by analyzing the asymptotics of the normalized processes under their limit homogeneous transformations.

We have

$$\frac{1}{n}\sum_{t=1}^{n} S\left(\sqrt{\frac{m}{n}} x_t\right) = \int_0^1 S\left(\sqrt{m} V_{mn}(r)\right) dr \tag{18}$$

and

Lemma 3.7 Let Assumption 2.1 hold. If S is regular, then

$$\int_0^1 S(\sqrt{m} V_{mn}(r))\, dr = \int_0^1 S(\sqrt{m} V_m(r))\, dr + o_p(1)$$

for large n, uniformly in m such that $m = o\,(n^{1-2/p} \wedge n^{2/3})$.

Moreover, it follows directly from normalization (9), application of occupation time formula (10) and Lemma 2.4 that

$$\begin{aligned} \int_0^1 S\left(\sqrt{m} V_m(r)\right) dr &= \frac{1}{m}\int_0^m S(V(r))\, dr \\ &= \frac{1}{m}\int_{-\infty}^{\infty} S(x) L(m, x)\, dx \\ &\to_{a.s.} \int_{-\infty}^{\infty} (SD)(x)\, dx \end{aligned} \tag{19}$$

as $m \to \infty$. As a result, we may easily deduce from (18) and (19), together with Lemma 3.7, that

$$\frac{1}{n}\sum_{t=1}^{n} S\left(\sqrt{\frac{m}{n}}x_t\right) \to_p \int_{-\infty}^{\infty}(SD)(x)\,dx \tag{20}$$

as $n \to \infty$, if $m = o\,(n^{1-2/p} \wedge n^{2/3})$ and $m \to \infty$ as $n \to \infty$. The corresponding asymptotics for T follow immediately from (17).

Now it can be easily deduced that

Theorem 3.8 Let Assumptions 2.1 and 3.1 hold. If F is regularly homogeneous with asymptotic order κ and limit homogeneous function H, and if $m = o\,(n^{1-2/p} \wedge n^{2/3})$ and $m \to \infty$ as $n \to \infty$, then we have

$$\frac{1}{n}\kappa\left(\sqrt{\frac{n}{m}}\right)^{-1}\sum_{t=1}^{n} F(x_t) \to_p \int_{-\infty}^{\infty}(HD)(x)\,dx,$$

$$\frac{1}{\sqrt{n}}\kappa\left(\sqrt{\frac{n}{m}}\right)^{-1}\sum_{t=1}^{n} F(x_t)\,u_t \to_d \mathbf{N}\left(0, \sigma^2\int_{-\infty}^{\infty}(H^2D)(x)\,dx\right)$$

as $n \to \infty$.

Under asymptotically homogeneous transformations, the mean and covariance asymptotics for weakly integrated processes are quite distinct from those for exactly or nearly integrated processes. For the asymptotics for nearly integrated processes, we just let m be fixed at c, and deduce

$$\frac{1}{n}\kappa(\sqrt{n})^{-1}\sum_{t=1}^{n} F(x_t) \to_d \int_0^1 H(V_c(s))ds, \tag{21}$$

$$\frac{1}{\sqrt{n}}\kappa(\sqrt{n})^{-1}\sum_{t=1}^{n} F(x_t)\,u_t \to_d \int_0^1 H(V_c(s))dU(s), \tag{22}$$

where U is the limit Brownian motion for (u_t).[6] If we set $c = 0$, then the asymptotics in (21) and (22) become those for the unit root processes obtained earlier by Park and Phillips (1999).

For asymptotically homogeneous transformations, the differing characteristics of the asymptotics for weakly integrated processes become more obvious. The asymptotics for weakly integrated processes have LLN and CLT type results, i.e., nonrandom probability limits and normal limit distributions under asymptotically homogeneous transformations, just as in the case of integrable transformations. On the other hand, the asymptotics for exactly and nearly integrated processes under asymptotically homogeneous transformations are quite different from those for integrable transformations. For exactly and nearly integrated processes, the mean asymptotics under asymptotically homogeneous transformations

[6] If we define the partial sum process $U_n(r) = n^{-1/2}\sum_{t=1}^{[nr]} u_t$ jointly with $V_n(r) = n^{-1/2}x_{[nr]+1}$, then it follows that $(U_n, V_n) \to_d (U, V_c)$ under Assumptions 2.1 and 3.1.

yield random limits. Moreover, the covariance asymptotics for exactly and nearly integrated processes are given as stochastic integrals, which are generally non-Gaussian, under asymptotically homogeneous transformations. The stochastic integral in (22) is non-Gaussian, unless U is independent of V_c. As will be shown in the next section, this difference in the covariance asymptotics has an important consequence for inference.

4. Inference in Models with Weak Unit Roots

As noted in the previous section, the basic asymptotics for weakly integrated processes involve the usual LLN and CLT type results, under both integrable and asymptotically homogeneous transformations. That is, the mean asymptotics yield nonrandom probability limits and the covariance asymptotics are given by limiting normal distributions. Their mean and covariance asymptotics are entirely analogous to LLN and CLT, respectively. This makes the theories of inference for models with weakly integrated processes completely parallel to those for models with stationay and ergodic processes. Therefore, the standard asymptotics built upon LLN and CLT become valid for a wide class of models with weakly integrated processes. In this section, we explicitly consider two special classes of models that are of most interest for practical applications: autoregressions and nonlinear regressions.

For both classes of models, we only consider their prototypical forms. For the former, we restrict our attention to the univariate, finite-order autoregressive models with single weak unit roots. Likewise, for the latter, we look at the simple regression models which have single weakly integrated regressors and martingale difference regression errors. Neither of them is allowed to have deterministic trends. It must be emphasized here that the reason for this is purely expositional. Our theoretical results, at least qualitatively, extend well beyond the models that we explicitly consider here. In particular, the normal asymptotics established here continue to be applicable for VAR's and other general regression models with multiple regressors, possibly with various deterministic trends. The rigorous derivations of their asymptotics, however, require some lengthy developments of new frameworks and methodologies, and it appears to be more desirable to report them separately in a subsequent work.

4.1 Autoregressive Models

Consider an autoregressive model for (x_t) given by

$$x_t = \alpha_1 x_{t-1} + \cdots + \alpha_K x_{t-K} + \varepsilon_t, \tag{23}$$

where (ε_t) is assumed to be a sequence of independent and identically distributed random variables with mean zero. We let

$$\alpha(z) = z^K - \alpha_1 z^{K-1} - \cdots - \alpha_K$$

and assume

Assumption 4.1 Let $\mathbf{E}|\varepsilon_t|^p < \infty$ for some $p > 2$, and let $\alpha(z)$ have a root $z = 1 - m/n$, while all the other roots are inside the unit circle.

Under Assumption 4.1, the time series (x_t) has a weak unit root. We may indeed rewrite (x_t) as in (1), where (v_t) follows an invertible $(K-1)$-th order autoregression, and therefore, can be represented as a linear process given in (5). Clearly, Assumption 4.1 is sufficient to ensure that the summability and moment conditions in Assumption 2.1 hold for such a representation. The time series (v_t) has the longrun variance given by $\omega^2 = \mathbf{E}\varepsilon_t^2/\alpha(1)^2$.

Now we derive the asymptotics for the least squares estimators $\hat{\alpha}_1, \ldots, \hat{\alpha}_K$ of the autoregressive coefficients $\alpha_1, \ldots, \alpha_K$. To do so, it will be convenient to look at a transformed model

$$x_t = \alpha x_{t-1} + \sum_{k=1}^{K-1} \beta_k \left(x_{t-k} - \left(1 - \frac{m}{n}\right) x_{t-k-1} \right) + \varepsilon_t, \tag{24}$$

where α is as given in (2) and (β_k) are the autoregressive coefficients for (v_t), i.e.,

$$v_t = \beta_1 v_{t-1} + \cdots + \beta_{K-1} v_{t-K+1} + \varepsilon_t.$$

The parameters in (23) and (24) are related to each other by

$$\begin{aligned} \alpha_1 &= \alpha + \beta_1, \\ \alpha_k &= \beta_k - (1 - m/n)\beta_{k-1}, \ k = 2, \ldots, K-1, \\ \alpha_K &= -(1 - m/n)\beta_{K-1}. \end{aligned} \tag{25}$$

In our representation (24), α is the critical parameter determining the nonstationary character of the model. Note that $\alpha = \sum_{k=1}^{K}(1 - m/n)^{-(k-1)}\alpha_k$, as can be easily deduced from (25). As is well known, the asymptotics for $\hat{\alpha}_1, \ldots, \hat{\alpha}_K$ in (23) can be obtained directly from those of $\hat{\alpha}$ and $\hat{\beta}_1, \ldots, \hat{\beta}_{K-1}$, the least squares estimators of α and $\beta_1, \ldots, \beta_{K-1}$ in (24), using the relationships derived in (25).

Let

$$\theta_{TRS} = (\alpha, \beta_1, \ldots, \beta_{K-1})'$$

and denote respectively by $\hat{\theta}_{TRS}$ the least squares estimator of θ_{TRS}. Moreover, we define a diagonal matrix

$$C_{mn} = \mathrm{diag}\left(\frac{n}{\sqrt{m}}, \sqrt{n} I_{K-1} \right),$$

where I_{K-1} is the identity matrix of dimension $K-1$, and let $\Gamma = \mathbf{E} v_{tK} v_{tK}'$, where $v_{tK} = (v_{t-1}, \ldots, v_{t-K+1})'$.

Theorem 4.2 Let Assumption 4.1 hold. We have

$$C_{mn}\left(\hat{\theta}_{TRS} - \theta_{TRS}\right) \to_d \mathbf{N}(0, \Sigma),$$

where $\Sigma = \mathrm{diag}\left(2\alpha(1)^2, \Gamma^{-1}\right)$, as $n \to \infty$, if $m = o\left(n^{1-2/p} \wedge n^{2/3}\right)$ and $m \to \infty$ as $n \to \infty$.

Theorem 4.2 shows that the limiting distributions of all parameter estimates in the transformed model (24) are jointly normal. In particular, the estimator $\hat{\alpha}$ of the critical parameter α has limiting normal distribution. This is in contrast with the exact and near unit root models. In these models, the limiting distribution of $\hat{\alpha}$ is nonnormal. Indeed, if we set $m = c$ for some fixed c, then it follows that

$$n\left(\hat{\alpha}-\alpha\right) \to_d \left(\int_0^1 V_c(r)^2 dr\right)^{-1} \int_0^1 V_c(r) dV_0(r), \tag{26}$$

which reduces to the distribution tabulated in Dickey and Fuller (1996) when $c = 0$. The limiting distribution appearing in (26) is nonnormal for all values of c. The rate of covergence for $\hat{\alpha}$ in our weak unit root model is given by $n/\sqrt{m}$, an order of magnitude slower than the rate n for the exact and near unit root models. As m approaches to n, the convergence rate for $\hat{\alpha}$ in the weak unit root model gets closer to $\sqrt{n}$. The estimators for all the other parameters have $\sqrt{n}$ convergence rates and normal asymptotics in the weak unit root model, like in the exact and near unit root models.

The limiting distributions of the least squares estimators for the parameters in the original model (23) can now be obtained directly from Theorem 4.2 and the relationships in (25). If we define

$$\theta_{ORG} = (\alpha_1, \ldots, \alpha_K)'$$

and let $\hat{\theta}_{ORG}$ be the least squares estimator of θ_{ORG}, then we may easily deduce under the conditions in Theorem 4.2 that

$$\sqrt{n}\left(\hat{\theta}_{ORG} - \theta_{ORG}\right) \to_d \mathbf{N}\left(0, J\Gamma^{-1}J'\right),$$

where Γ is defined in the paragraph preceding Theorem 4.2 and

$$J = \begin{pmatrix} 1 & & & \\ -1 & 1 & & \\ & -1 & \ddots & \\ & & \ddots & 1 \\ & & & -1 \end{pmatrix},$$

and that

$$\frac{n}{\sqrt{m}}\,\iota'\left(\hat{\theta}_{ORG} - \theta_{ORG}\right) \to_d \mathbf{N}\left(0, 2\alpha(1)^2\right),$$

where ι is the K-dimensional vector of ones.

The limiting distribution of $(\hat{\alpha}_k)$ is jointly normal with the convergence rate $\sqrt{n}$ for our model with the weak unit root, and this is also precisely what we have for the exact and weak unit root models. The joint limit distribution of $(\hat{\alpha}_k)$, however, has a sigularity in the sum $\sum_{k=1}^K \hat{\alpha}_k$, and this is again true for all models with exact, near and weak unit roots. This common sigularity also brings out the important differences between the asymptotics for weak unit root models and those for exact and near unit root models. Although $\sum_{k=1}^K \hat{\alpha}_k$ converge at accelerated rates in all models, the rates are different in each model: It converges at the rate $n/\sqrt{m}$ for the weak unit root model but at n rate for the exact and near unit root models. More importantly, they yield different limiting distributions along the singularity: The distribution of $(\hat{\alpha}_k)$ along the singularity is again normal for the weak unit root model, whereas it is nonnormal for the exact and near unit root models.

In the weak unit root model, all linear combinations of $(\hat{\alpha}_k)$ have limiting normal distributions, including the one that is given by the direction which yields the asymptotic multicollinearity and an accelerated convergence rate. Qualitatively, the asymptotics for the weak unit root model are completely parallel to those for stationary invertible autoregressions. This has an obvious implication for inference. For the weak unit root model, the hypotheses on autoregressive coefficients can be tested using standard tests that rely on normal asymptotics. This, however, is not so for the exact and near unit root models. For those models, the standard t-statistic to test for a hypothesis on the sum of autoregressive coefficients, or the nonstationarity parameter, has nonnormal limiting distribution, and therefore, the usual t-test relying on normal critical values is not valid.

4.2 Nonlinear Regression Models

Next we look at the nonlinear regression model given by

$$y_t = f(x_t, \theta_0) + u_t, \tag{27}$$

where (x_t) is a weakly integrated regressor and (u_t) is the regression error that is assumed to be a martingale difference sequence. As usual, we let f be a known function and let θ_0 be the unknown parameter. The unknown parameter θ_0 is commonly estimated by the nonlinear least squares (NLS) estimator $\hat{\theta}$, which is defined by

$$\hat{\theta} = \underset{\theta\in\Theta}{\operatorname{argmin}} \sum_{t=1}^{n} (y_t - f(x_t, \theta))^2,$$

where Θ is the parameter set that is assumed to be compact.

To introduce the necessary regularity conditions for the regression function, we need to consider the class of vector-valued functions $F(\cdot, \pi)$ on $\mathbf{R}$ indexed by the parameter $\pi \in \Pi$, and require that the conditions in Definitions 3.2 and 3.5 hold uniformly for $\pi \in \Pi$ in an appropriate sense. Naturally, we say that the regularity conditions in Definitions 3.2 and 3.5 are satisfied for a vector-valued function if and only if they hold for each component of the function. For the required uniformity, we define more formally that

Definition 4.3 We say that $F(\cdot, \pi)$ satisfies the regularity conditions of Definitions 3.2 and 3.5 *uniformly* for $\pi \in \Pi$, if their Lipschitz conditions in part (a) hold with the Lipschitz constant independent of $\pi \in \Pi$, and their boundedness conditions in part (b) are met for $\sup_{\pi\in\Pi} |F(\cdot, \pi)|$.

We may now define the regularity conditions for the families $F(\cdot, \pi)$ of functions that are integrable and asymptotically homogeneous functions. We call them I- and H-regular, respectively, if they satisfy the required regularity conditions.

Definition 4.4 We say that F is I-*regular* on Π if

(a) $F(\cdot,\pi)$ satisfies the regularity conditions in Definition 3.2 uniformly for $\pi \in \Pi$, and

(b) for each $\pi_0 \in \Pi$, there exists a neighborhood N of π_0 and T bounded and regularly integrable such that $\|F(x,\pi) - F(x,\pi_0)\| \leq \|\pi - \pi_0\| T(x)$ for all $\pi \in N$.

Definition 4.5 Let

$$F(\lambda x, \pi) = \kappa(\lambda, \pi) H(x, \pi) + R(x, \lambda, \pi),$$

where κ is nonsingular. We say that F is H-*regular* on Π if

(a) $H(\cdot,\pi)$ satisfies the second-order regularity conditions in Definition 3.5 uniformly for $\pi \in \Pi$,

(b) $H(x,\cdot)$ is continuous on Π, and

(c) $R(x,\lambda,\pi) = \varpi(\lambda,\pi)Q(x)$, where $(\kappa^{-1}\varpi)(\lambda) \to 0$ as $\lambda \to \infty$ uniformly for $\pi \in \Pi$ and Q is regular in the second-order.

We call κ and H respectively the asymptotic order and limit homogeneous function of F. If F has the asymptotic order κ which does not depend upon π, we say that it is H_0-regular.

The regularity conditions in Definitions 4.4 and 4.5 are not very stringent, and are satisfied by virtually all nonlinear regression models used in practical applications.[7]

The asymptotic distributions of the NLS estimator $\hat{\theta}$ of θ_0 may now be developed under the regularity conditions introduced above. Define $\dot{f} = (\partial f/\partial\theta_i)$, $\ddot{f} = (\partial^2 f/\partial\theta_i\partial\theta_j)$ and $\dddot{f} = (\partial^3 f/\partial\theta_i\partial\theta_j\partial\theta_k)$ to be the vectors of partial derivatives of f with respect to θ, arranged by the lexicographic ordering of their indices. For the regressions with I-regular regression functions, we have

Theorem 4.6 Let Assumptions 2.1, 2.2 and 3.1 hold. Assume

(a) $f, \dot{f}$ and $\ddot{f}$ are I-regular on Θ,

(b) $\int_{-\infty}^{\infty}(f(x,\theta) - f(x,\theta_0))^2 dx > 0$ for all $\theta \neq \theta_0$, and

(c) $\int_{-\infty}^{\infty}(\dot{f}\dot{f}')(x,\theta_0)\, dx > 0$.

Then we have

$$\sqrt[4]{nm}\left(\hat{\theta} - \theta_0\right) \to_d \mathbf{N}\left(0, \sigma^2\left(D(0)\int_{-\infty}^{\infty}(\dot{f}\dot{f}')(x,\theta_0)\, dx\right)^{-1}\right)$$

as $n \to \infty$, if $m = o\left(n^{1-2/p} \wedge n^{2/3}\right)$ and $m \to \infty$ as $n \to \infty$.

Theorem 4.6 establishes the large sample theory for the NLS estimator $\hat{\theta}$ in the nonlinear regression with integrable regression function and weakly integrated regressor. Besides the technical regularity conditions, we only require some minimal conditions for identification, which are expected to hold for a wide range of integrable regression functions.

[7] The regularity conditions here are not directly comparable to those introduced by Park and Phillips (2001) to establish the asymptotic theory for the nonlinear regression models with exact integrated processes. See Footnote 5.

If the regressor is weakly integrated, the NLS estimator is asymptotically normal in the regression with integrable regression function. This is in contrast with the case where the regressor is exactly or nearly integrated. In this case, we have

$$\sqrt[4]{n}\left(\hat{\theta}-\theta_0\right) \to_d \mathbf{MN}\left(0, \sigma^2\left(L_c(1,0)\int_{-\infty}^{\infty}(\dot{f}\dot{f}')(x,\theta_0)\,dx\right)^{-1}\right), \tag{28}$$

where L_c is the local time of the Ornstein-Uhlenbeck process with parameter c. The asymptotics in (28) generalize those obtained in Park and Phillips (2001) for the regression with exactly integrated regressor. The limiting distribution in (28) is mixed normal, contrastingly with the case of weakly integrated regressor. The inference, however, can be based on the usual chi-square tests in all three cases where the regressor is exactly, nearly or weakly integrated. The convergence rate for the NLS estimator in the regression with weakly integrated regressor is $\sqrt[4]{nm}$, and faster than $\sqrt[4]{n}$ in the regression with exactly or nearly integrated regressor. It approaches to the usual $\sqrt{n}$ rate as m gets close to n.

We now consider the nonlinear regressions with H-regular regression functions. In what follows, we denote by (κ, h), $(\dot{\kappa}, \dot{h})$, $(\ddot{\kappa}, \ddot{h})$ and $(\dddot{\kappa}, \dddot{h})$, respectively, the asymptotic order and the limit homogeneous function of f, $\dot{f}$, $\ddot{f}$ and $\dddot{f}$ that are assumed to be asymptotically homogeneous. Moreover, we do not distinguish two functions on $\mathbf{R}$ that are identical a.e., i.e., identical except on a subset of $\mathbf{R}$ with Lebesgue measure zero. As an example, for two functions a and b on $\mathbf{R}$, $a \neq b$ implies that they disagree on a subset of $\mathbf{R}$ with nonzero Lebesgue measure.

Theorem 4.7 Let Assumptions 2.1 and 3.1 hold. Assume
(a) $f, \dot{f}$ and $\ddot{f}$ are H$_0$-regular on Θ,
(b) $\|\left((\dot{\kappa}\otimes\dot{\kappa})^{-1}\kappa\ddot{\kappa}\right)(\lambda)\| < \infty$ as $\lambda \to \infty$,
(c) $h(\cdot,\theta) \neq h(\cdot,\theta_0)$ for all $\theta \neq \theta_0$, and
(d) $\dot{h}(\cdot,\theta_0)$ is linearly independent.
Then we have

$$\sqrt{n}\dot{\kappa}\left(\sqrt{\frac{n}{m}}\right)'\left(\hat{\theta}-\theta_0\right) \to_d \mathbf{N}\left(0, \sigma^2\left(\int_{-\infty}^{\infty}(\dot{h}\dot{h}')(x,\theta_0)D(x)\,dx\right)^{-1}\right)$$

as $n \to \infty$, if $m = o\left(n^{1-2/p} \wedge n^{2/3}\right)$ and $m \to \infty$ as $n \to \infty$.

Theorem 4.8 Let Assumptions 2.1 and 3.1 hold. Assume
(a) $\dot{f}, \ddot{f}$ and $\dddot{f}$ are H-regular on Θ,
(b) there exists a neighborhood N of θ_0 such that we have as $\lambda \to \infty$

$$\lambda^{-1}\left\|\left((\dot{\kappa}\otimes\dot{\kappa})^{-1}\ddot{\kappa}\right)(\lambda,\theta_0)\right\| \to 0, \tag{29}$$

$$\lambda^{-1+\varepsilon}\left\|(\dot{\kappa}\otimes\dot{\kappa})^{-1}(\lambda,\theta_0)\left(\sup_{\theta\in N}\ddot{\kappa}(\lambda,\theta)\right)\right\| \to 0, \tag{30}$$

$$\lambda^{-1+\varepsilon}\left\|(\dot{\kappa}\otimes\dot{\kappa}\otimes\dot{\kappa})^{-1}(\lambda,\theta_0)\left(\sup_{\theta\in N}\dddot{\kappa}(\lambda,\theta)\right)\right\| \to 0, \tag{31}$$

for some $\epsilon > 0$, and

(c) $\dot{h}(\cdot, \theta_0)$ is linearly independent.

Then we have

$$\sqrt{n}\dot{\kappa}\left(\sqrt{\frac{n}{m}}, \theta_0\right)' \left(\hat{\theta} - \theta_0\right) \to_d \mathbf{N}\left(0, \sigma^2\left(\int_{-\infty}^{\infty} (\dot{h}\dot{h}')(x, \theta_0) D(x)\, dx\right)^{-1}\right)$$

as $n \to \infty$, if $m = o\left(n^{1-2/p} \wedge n^{2/3}\right)$ and $m \to \infty$ as $n \to \infty$.

Theorems 4.7 and 4.8 establish the large sample theory for the NLS estimator $\hat{\theta}$ in the nonlinear regression with asymptotically homogeneous regression function and weakly integrated regressor. The technical regularity conditions are not very restrictive, and the identifying restrictions are extremely mild. The required conditions are satisfied for a wide variety of asymptotically homogeneous regression functions, including all that are considered in Park and Phillips (2001).

If the regressor is weakly integrated, the NLS estimator is asymptotically normal also for the regression with asymptotically homogeneous regression function. As a natural consequence, the standard chi-square tests provide valid inferences. The asymptotics here are more drastically different from those for the regressions with exactly or nearly integrated regressors, in which case we have

$$\sqrt{n}\, \dot{\kappa}_n' \left(\hat{\theta} - \theta_0\right) \to_d \left(\int_0^1 (\dot{h}\dot{h}')(V_c(s), \theta_0) ds\right)^{-1} \int_0^1 \dot{h}(V_c(s), \theta_0) dU(s) \tag{32}$$

with $\dot{\kappa}_n = \dot{\kappa}(\sqrt{n})$ or $\dot{\kappa}_n(\sqrt{n}, \theta_0)$ correspondingly to the asymptotics in Theorem 4.7 or 4.8, where V_c is the Ornstein-Uhlenbeck process with parameter c and U is the limit Brownian motion for (u_t) as in (22). The asymptotics in (32) extend those in Park and Phillips (2001) to the case for nearly integrated regressors. The limiting distribution in (32) is nonnormal unless U is independent of V_c, and generally depends upon the correlation between U and V_c. This of course implies that the usual tests based on the NLS procedure are invalid, in sharp contrast to the regressions with weakly integrated regressors. The convergence rate for the NLS estimator in the regression with weakly integrated regressor is determined by the asymptotic order of the regression function, as in the case of the regression with exactly or nearly integrated regressor. However, it converges to the standard $\sqrt{n}$ rate as m approaches to n, irrespective of the regression function.

5. Concluding Remarks

In this paper, we consider the time series with roots given as functions of the sample size n, and approaching to unity at a rate slower than n^{-1}. The motivation is clear. We set the roots as functions of n, since their finite sample distributions in a neighborhood of the unit root depend crucially on n. Further, we concentrate on the roots converging slowly to unity, since this is the most relevant case for practical applications. Many economic and financial time series are indeed known to have roots in close, but not too close, proximity of unity. Our theories, though developed using abstract mathematics and sophisticated reasoning, have implications that are unambiguous and straightforward: The usual normal asymptotics are generally applicable in finite samples for the standard econometric models even if they include time series with roots close to one, as long as they are not too close to one and the sample size is reasonably large.

Our theories are developed for a wide class of econometric models that are frequently used in practical applications. However, we only explicitly consider their prototypical forms. In particular, our models only allow for a single weakly integrated process without any deterministic regressors and assume the absence of endogeneity. Needless to say, this is to focus on the important issues and effectively deliver the main messages. For the practical applications, the models considered in the paper may well be restrictive. The extensions to alleviate these restrictions, though possible, require some new tools and fundamental results that are not introduced in the paper. They are underway, and will be reported in subsequent works. Roughly, all of our qualitative results in the paper well extend to more general models with multiple regressors possibly including deterministic trends, and to a certain degree, also to models with endogeneity.

The normal asymptotics continue to hold for the regression models with multiple weakly integrated regressors. Our results in the paper are indeed naturally extended to them within a new framework allowing for multiple weakly integrated processes. We have similar normal limit theories for the regression models with deterministic trends, though the actual asymptotics are developed in a somewhat different way due to the presence of non-stochastic and trending regressors. The effect of endogeneity is more essential. The least squares estimators have the limiting distributions that are affected by the presence of endogeneity and depend upon the nuisance parameters characterizing the endogeneity. However, they are still consistent. This is because the stochastic orders of weakly integrated regressors are bigger than those of the regression errors that are assumed to be stationary, as in the theory of the usual cointegression models with exact unit roots.

6. Mathematical Proofs

Our n-asymptotics in the paper are developed under the condition $m = o\,(n^{1-2/p} \wedge n^{2/3})$ and $m \to \infty$ as $n \to \infty$. To avoid the repetition, however, we simply state in what follows "$n \to \infty$" and do not make it explicit that m is also given as a function of n satisfying the condition $m = o\,(n^{1-2/p} \wedge n^{2/3})$ and $m \to \infty$ as $n \to \infty$.

6.1 Useful Lemmas and Their Proofs

Lemma A1 Let V be the Ornstein-Uhlenbeck process with unit parameter, and let L be the local time of V. We have

(a) $\mathbf{E}\exp(cL(1,x)) < \infty$ for all c and x,

(b) $|L(1,a) - L(1,b)| \leq |a-b|^{1/2-\epsilon} Z$ for any $\epsilon > 0$, where $\mathbf{E}Z^k < \infty$ for all $k \geq 0$, and

(c) $\sup_{0\leq r\leq m} |V(r)| = O\left((\log m)^{1/2}\right)$ a.s.

Proof of Lemma A1 It follows from (8) with $m = 1$ and Tanaka's formula [see, e.g., Revuz and Yor (1994, Theorem 1.2, p213)] that

$$\begin{aligned} L(1,x) \,=\, & 2\left((V(1)-x)^+ - (V(0)-x)^-\right) \\ & - 2\left(\int_0^1 1\{V(r)>x\}\,dW(r) - \int_0^1 V(r)1\{V(r)>x\}\,dr\right), \end{aligned} \tag{33}$$

which will be used repeatedly below. It is well known that the stochastic differential equation (8) has a stationary solution for any $m > 0$, if the initial condition is given consistently. Therefore, if we let $V(0)$ to be a normal random variate with mean zero and variance 1/2, V becomes a stationary process. In our definition V is not strictly statonary, since $V(0)$ is assumed to be zero. This assumption, however, is made purely for expositional simplicity, and does not change any of our results in the paper that rely on the asymptotic behavior of V. Here and elsewhere, we will simply assume that V is a stationary process, with or without the convention $V(0) = 0$.

For the proof of part (a), we note that

$$\mathbf{E}\exp(cV(1)) < \infty \tag{34}$$

for any c, due to Gaussianity of $V(1)$. Moreover, we have

$$\int_0^1 1\{V(r)>x\}\,dW(r) \le \sup_{0\le r\le 1}|U(r)|,$$

where U is the DDS(Dambis-Dubins-Schwarz) Brownian motion [see, e.g., Revuz and Yor (1194, Theorem 1.6, p173)] of the martingale $\int_0^{\cdot} 1\{V(r)>x\}\,dW(r)$, the quadratic variation of which at time unity is given by $\int_0^1 1\{V(r)>x\}\,dr \le 1$. Consequently,

$$\mathbf{E}\exp\left(c\int_0^1 1\{V(r)>x\}\,dW(r)\right) < \infty \tag{35}$$

for any c. Finally, we have

$$\left|\int_0^1 V(r)1\{V(r)>x\}\,dr\right| \le \int_0^1 |V(r)|\,dr \le 2\sup_{0\le r\le 1}|W(r)|$$

since

$$V(r) = W(r) - \int_0^r e^{-(r-s)}W(s)\,ds$$

and, for all $r \in [0,1]$,

$$|V(r)| \le |W(r)| + \left(1-e^{-r}\right)\sup_{0\le s\le r}|W(s)| \le 2\sup_{0\le r\le 1}|W(r)|. \tag{36}$$

It therefore follows that

$$\mathbf{E}\exp\left(c\int_0^1 V(r)1\{V(r)>x\}\,dr\right) < \infty \tag{37}$$

for any c. Recall that the running maximum of Brownian motion has the same distribution as the modulus of Brownian motion. The result stated in part (a) can now be easily deduced from (34), (35) and (37).

For the proof of part (b), it suffices to show that

$$\mathbf{E}|L(1,a) - L(1,b)|^{2k} \leq c_k|a-b|^k \tag{38}$$

for all integer $k \geq 1$ and some constant c_k depending only upon k (which work in particular for all a and b), due to the extended version of the Kolmogorov criterion in Revuz and Yor (1994, Theorem 2.1, p25). Here and elsewhere in the proof, we use c_k to denote a generic constant depending upon k, which may vary from line to line. Note that

$$\begin{aligned}
\mathbf{E}\,L^k(1,x) \;&\leq\; c_k\mathbf{E}\left(|V(1)|^k + \left(\int_0^1 |V(r)|\,dr\right)^k + 1\right) \\
&\leq\; c_k\mathbf{E}\left(\sup_{0\leq r\leq 1}|W(r)|\right)^k,
\end{aligned}$$

which follows readily from (33), (36) and the inequality by BDG (Burkholder-Davis-Gundy) [see, e.g., Revuz and Yor (1994, Theorem 4.1, p153)]. We therefore have

$$\sup_{x\in\mathbf{R}} \mathbf{E}\,L^k(1,x) \leq c_k \tag{39}$$

for all $k \geq 1$.

Now define

$$M(x) = \int_0^1 1\{V(r) > x\}\,dW(r)$$

so that we have

$$M(a) - M(b) = \int_0^1 1\{a < V(r) \leq b\}\,dW(r).$$

Using BDG-inequality, occupation time formula, Hölder inequality and Fubini's theorem successively, we may deduce that

$$\begin{aligned}
\mathbf{E}|M(a) - M(b)|^{2k} \;&\leq\; c_k\mathbf{E}\left(\int_0^1 1\{a < V(r) \leq b\}\,dr\right)^k \\
&=\; c_k\mathbf{E}\left(\int_{-\infty}^{\infty} 1\{a < x \leq b\}L(1,x)\,dx\right)^k \\
&\leq\; c_k|a-b|^k\mathbf{E}\left(\frac{1}{b-a}\int_a^b L^k(1,x)\,dx\right) \\
&=\; c_k|a-b|^k\left(\sup_{x\in\mathbf{R}}\mathbf{E}\,L^k(1,x)\right)
\end{aligned} \tag{40}$$

for all $k \geq 1$.

Moreover, if we let

$$N(x) = \int_0^1 V(r)1\{V(r) > x\}\, dr$$

and

$$N(a) - N(b) = \int_0^1 V(r)1\{a < V(r) \le b\}\, dr,$$

then it follows from occupation times formula, Hölder inequality and Fubini's theorem that

$$\begin{aligned}
\mathbf{E}|N(a) - N(b)|^{2k} &= \mathbf{E}\left(\int_0^1 V(r)1\{a < V(r) \le b\}\, dr\right)^{2k} \\
&= \mathbf{E}\left(\int_{-\infty}^{\infty} x1\{a < x \le b\}L(1,x)\, dx\right)^{2k} \\
&\le \left(\frac{b^2 - a^2}{2}\right)^{2k} \mathbf{E}\left(\frac{2}{b^2 - a^2}\int_a^b xL^{2k}(1,x)\, dr\right) \\
&= \left(\frac{b^2 - a^2}{2}\right)^{2k} \left(\sup_{x\in\mathbf{R}} \mathbf{E}\, L^{2k}(1,x)\right)
\end{aligned}$$

for all $k \ge 1$. As a result, we may choose

$$|a - b| \le \max(|a| + |b|, 1)^{-2}$$

to deduce that

$$\mathbf{E}|N(a) - N(b)|^{2k} \le c_k|a - b|^k \left(\sup_{x\in\mathbf{R}} \mathbf{E}\, L^{2k}(1,x)\right) \tag{41}$$

for all $k \ge 1$. Due to (33), we may now easily derive (38) from (39)–(41), and therefore, the proof of part (b) is complete.

To prove part (c), we write

$$e^r V(r) = U\left(\frac{1}{2}(e^{2r} - 1)\right), \tag{42}$$

where U is the DDS Brownian motion of the martingale M

$$M(r) = e^r V(r) = \int_0^r e^{2s} dV_0(s).$$

Due to the law of iterated logarithm for Brownian motion, we have

$$(\log r)^{-1/2} e^{-r}\left|U\left(\frac{1}{2}(e^{2r} - 1)\right)\right| = (\log r)^{-1/2}|V(r)| = O_{a.s.}(1)$$

and it therefore follows from our representation V in (42) as a time changed Brownian motion that

$$\sup_{0\le r\le m} |V(r)| = O_{a.s.}((\log m)^{1/2})$$

as was to be shown.

Lemma A2 Let $\pi_0 \in \Pi$ be arbitrarily chosen and let N be any neighborhood of π_0.

(a) If $F(\cdot,\pi)$ satisfies the regularity conditions in Definition 3.2 uniformly in $\pi \in \Pi$, then $\sup_{\pi\in N} F(\cdot,\pi)$ and $\inf_{\pi\in N} F(\cdot,\pi)$ are both regularly integrable.

(b) If $F(\cdot,\pi)$ satisfies the regularity conditions in Definition 3.5 uniformly in $\pi \in \Pi$, then $\sup_{\pi\in N} F(\cdot,\pi)$ and $\inf_{\pi\in N} F(\cdot,\pi)$ are both regularly locally integrable.

Proof of Lemma A2 Note that

$$\left|\inf_{\pi\in N} F(x,\pi) - \inf_{\pi\in N} F(y,\pi)\right|, \ \left|\sup_{\pi\in N} F(x,\pi) - \sup_{\pi\in N} F(y,\pi)\right| \\ \leq \ \sup_{\pi\in N} |\, F(x,\pi) - F(y,\pi)\,|$$

and that

$$\left|\sup_{\pi\in N} F(\cdot,\pi)\right|, \ \left|\inf_{\pi\in N} F(\cdot,\pi)\right| \leq \sup_{\pi\in\Pi} |F(\cdot,\pi)|,$$

from which the stated results follow immediately.

Lemma A3 Let Assumptions 2.1, 2.2 and 3.1 hold. If F is I-regular on a compact set Π, then we have

$$\frac{1}{\sqrt{nm}}\sum_{t=1}^{n} F(x_t,\pi) \to_p D(0)\int_{-\infty}^{\infty} F(x,\pi)\,dx$$

and

$$\frac{1}{\sqrt{nm}}\sum_{t=1}^{n} F(x_t,\pi)\,u_t \to_p 0$$

as $n\to\infty$, uniformly in $\pi\in\Pi$.

Proof of Lemma A3 Fix $\pi_0 \in \Pi$ arbitrarily, and let N be any neighborhood of π_0. Due to I-regularity condition (a) and Lemma A2, $\sup_{\pi\in N} F(\cdot,\pi)$ and $\inf_{\pi\in N} F(\cdot,\pi)$ are regularly integrable. Therefore, we have

$$\frac{1}{\sqrt{nm}}\sum_{t=1}^{n} \sup_{\pi\in N} F(x_t,\pi) \ \to_p \ D(0)\int_{-\infty}^{\infty} \sup_{\pi\in N} F(x,\pi)\,dx, \tag{43}$$

$$\frac{1}{\sqrt{nm}}\sum_{t=1}^{n} \inf_{\pi\in N} F(x_t,\pi) \ \to_p \ D(0)\int_{-\infty}^{\infty} \inf_{\pi\in N} F(x,\pi)\,dx \tag{44}$$

from Theorem 3.4.

Let N_δ be the δ-neighborhood of π_0. By I-regularity condition (b), we have for all $x \in \mathbf{R}$

$$\sup_{\pi\in N_\delta} F(x,\pi) - \inf_{\pi\in N_\delta} F(x,\pi) \to 0$$

as $\delta \to 0$, and by dominated convergence,

$$\int_{-\infty}^{\infty} \sup_{\pi\in N_\delta} F(x,\pi)\,dx - \int_{-\infty}^{\infty} \inf_{\pi\in N_\delta} F(x,\pi)\,dx \to 0 \tag{45}$$

as $\delta \to 0$. We may now easily deduce from (43)–(45) that there exists a neighborhood of π_0 such that

$$\frac{1}{\sqrt{nm}}\sum_{t=1}^{n} F(x_t,\pi) \to_p D(0)\int_{-\infty}^{\infty} F(x,\pi)\,dx$$

uniformly in π. Since π_0 was chosen arbitrarily and Π is compact, the proof for mean asymptotics is complete.

For the proof of the result for covariance asymptotics, we also choose $\pi_0 \in \Pi$ arbitrarily. Due to the compactness of Π, it suffices to show that there exists a neighborhood N of π_0 such that

$$\sup_{\pi\in N}\left|\frac{1}{\sqrt{nm}}\sum_{t=1}^{n} F(x_t,\pi)\,u_t\right| = o_p(1) \tag{46}$$

as $n \to \infty$. Moreover, since it follows from Theorem 3.4 that

$$\frac{1}{\sqrt[4]{nm}}\sum_{t=1}^{n} F(x_t,\pi_0)\,u_t = O_p(1),$$

we only need to establish that

$$\sup_{\pi\in N}\left|\frac{1}{\sqrt{nm}}\sum_{t=1}^{n} (F(x_t,\pi)-F(x_t,\pi_0))\,u_t\right| = o_p(1) \tag{47}$$

to show (46).

However, it follows from I-regularity condition (b) that

$$\sum_{t=1}^{n} |F(x_t,\pi)-F(x_t,\pi_0)||u_t| \le \|\pi-\pi_0\|\left(\sigma\sum_{t=1}^{n}|T(x_t)| + \sum_{t=1}^{n}|T(x_t)|w_t\right), \tag{48}$$

where $w_t = |u_t| - \mathbf{E}(|u_t||\mathcal{F}_{t-1})$. Note that $\mathbf{E}(|u_t||\mathcal{F}_{t-1})^2 \le \sigma^2$ by Jensen's inequality. Furthermore, we have from Theorem 3.4 that

$$\frac{1}{\sqrt{nm}}\sum_{t=1}^{n}|T(x_t)|,\ \ \frac{1}{\sqrt[4]{nm}}\sum_{t=1}^{n}|T(x_t)|w_t = O_p(1),$$

since T is regularly integrable. It is therefore clear from (48) that we may choose a neighborhood N of π_0 such that (47) holds, and this completes the proof.

Lemma A4 Let Assumptions 2.1 and 3.1 hold. If F is H-regular with asymptotic order κ and limit homogeneous function H on a compact set Π, then we have

$$\frac{1}{n}\kappa\left(\sqrt{\frac{n}{m}},\pi\right)^{-1}\sum_{t=1}^{n} F(x_t,\pi) \to_p \int_{-\infty}^{\infty} H(x,\pi)D(x)\,dx$$

and

$$\frac{1}{n}\kappa\left(\sqrt{\frac{n}{m}},\pi\right)^{-1}\sum_{t=1}^{n} F(x_t,\pi)\,u_t \to_p 0$$

as $n \to \infty$, uniformly in $\pi \in \Pi$

Proof of Lemma A4 Due to H-regularity condition (c), we have

$$\frac{1}{n}\kappa\left(\sqrt{\frac{n}{m}},\pi\right)^{-1}\sum_{t=1}^{n}F(x_t,\pi)=\frac{1}{n}\sum_{t=1}^{n}H\left(\sqrt{\frac{n}{m}}x_t,\pi\right)+o_p(1) \tag{49}$$

uniformly in $\pi\in\Pi$, as $n\to\infty$. Moreover, if we let π_0 be chosen arbitrarily and let N be a neighborhood of π_0, then it follows directly from H-regularity condition (a) and Lemma A2 that $\sup_{\pi\in N}H(\cdot,\pi)$ and $\inf_{\pi\in N}F(\cdot,\pi)$ are regularly locally integrable. Therefore, we have from (20) that

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\pi\in N}H\left(\sqrt{\frac{n}{m}}x_t,\pi\right)\to_p\int_{-\infty}^{\infty}\sup_{\pi\in N}H(x,\pi)D(x)\,dx, \tag{50}$$

$$\frac{1}{n}\sum_{t=1}^{n}\inf_{\pi\in N}H\left(\sqrt{\frac{n}{m}}x_t,\pi\right)\to_p\int_{-\infty}^{\infty}\inf_{\pi\in N}H(x,\pi)D(x)\,dx \tag{51}$$

as $n\to\infty$.

Let N_δ be the δ-neighborhood of π_0. Then we have for every $x\in\mathbf{R}$

$$\sup_{\pi\in N_\delta}H(x,\pi)-\inf_{\pi\in N_\delta}H(x,\pi)\to 0$$

as $\delta\to 0$, due to the continuity of $H(x,\cdot)$ given by H-regularity condition (b). Moreover, it follows from dominated convergence that

$$\int_{-\infty}^{\infty}\left(\sup_{\pi\in N_\delta}H(x,\pi)-\inf_{\pi\in N_\delta}H(x,\pi)\right)D(x)\,dx\to 0 \tag{52}$$

as $\delta\to 0$. We may now easily deduce from (49)– (52) that there exists a neighborhood of π_0, where

$$\frac{1}{n}\kappa\left(\sqrt{\frac{n}{m}},\pi\right)^{-1}\sum_{t=1}^{n}F(x_t,\pi)\to_p\int_{-\infty}^{\infty}H(x,\pi)D(x)\,dx$$

holds uniformly in π. Since π_0 was chosen arbitrary and Π is compact, this proves the stated result for mean asymptotics.

For the proof of the result for covariance asymptotics, note that we have for any $\pi_0\in\Pi$

$$\frac{1}{\sqrt{n}}\kappa\left(\sqrt{\frac{n}{m}},\pi\right)^{-1}\sum_{t=1}^{n}F(x_t,\pi_0)\,u_t=O_p(1)$$

due to Theorem 3.8. Therefore, we only need to show that there exists a neighborhood N of π_0 such that

$$\sup_{\pi\in N}\left|\frac{1}{n}\kappa\left(\sqrt{\frac{n}{m}},\pi\right)^{-1}\sum_{t=1}^{n}\left(F(x_t,\pi)-F(x_t,\pi_0)\right)u_t\right|=o_p(1) \tag{53}$$

to establish that

$$\sup_{\pi\in N}\left|\frac{1}{n}\kappa\left(\sqrt{\frac{n}{m}},\pi\right)^{-1}\sum_{t=1}^{n}F(x_t,\pi)\,u_t\right| = o_p(1)$$

as $n\to\infty$. The stated result would then follow immediately from the compactness of Π.

It follows from Cauchy-Schwarz that

$$\begin{aligned}&\left|\frac{1}{n}\kappa\left(\sqrt{\frac{n}{m}},\pi\right)^{-1}\sum_{t=1}^{n}(F(x_t,\pi)-F(x_t,\pi_0))\,u_t\right|\\ &\quad\le\left(\frac{1}{n}\kappa\left(\sqrt{\frac{n}{m}},\pi\right)^{-2}\sum_{t=1}^{n}(F(x_t,\pi)-F(x_t,\pi_0))^2\right)^{1/2}\left(\frac{1}{n}\sum_{t=1}^{n}u_t^2\right)^{1/2}. \qquad (54)\end{aligned}$$

However, we may show as in the proof of part (a) that

$$\frac{1}{n}\kappa\left(\sqrt{\frac{n}{m}},\pi\right)^{-2}\sum_{t=1}^{n}(F(x_t,\pi)-F(x_t,\pi_0))^2\to_p\int_{-\infty}^{\infty}(H(x,\pi)-H(x,\pi_0)^2D(x)\,dx \qquad (55)$$

uniformly in $\pi\in\Pi$. Moreover, due to H-regularity condition (b), we have for all $x\in\mathbf{R}$

$$\sup_{\pi\in N_\delta}|H(x,\pi)-H(x,\pi_0)|\to 0$$

and it follows from the dominated convergence that

$$\int_{-\infty}^{\infty}(H(x,\pi)-H(x,\pi_0)^2D(x)\,dx\to 0 \qquad (56)$$

as $\delta\to 0$, where N_δ is the δ-neighborhood of π_0. It now follows readily from (54)–(56) that there exists a neighborhood N of π_0 such that (53) holds uniformly, and the proof is complete.

Lemma A5 Let Assumption 4.1 hold. Then we have for $k=1,\ldots,K-1$

$$\frac{1}{n}\sum_{t=1}^{n}x_t v_{t-k}=O_p(1)$$

as $n\to\infty$, if $m=o\left(n^{1-2/p}\wedge n^{2/3}\right)$ and $m\to\infty$ as $n\to\infty$.

Proof of Lemma A5 Notice that

$$\begin{aligned}\sum_{t=1}^{n}x_t v_{t-k} &= \left(1-\frac{m}{n}\right)^{k+1}\sum_{t=1}^{n}x_{t-k-1}v_{t-k}\\ &\quad+\sum_{t=1}^{n}v_t v_{t-k}+\left(1-\frac{m}{n}\right)\sum_{t=1}^{n}v_{t-1}v_{t-k}+\cdots+\left(1-\frac{m}{n}\right)^{k}\sum_{t=1}^{n}v_{t-k}v_{t-k}\\ &= \left(1+O\left(\frac{m}{n}\right)\right)\sum_{t=1}^{n}x_{t-1}v_t+O_p(n)\end{aligned}$$

uniformly for any finite number of k's. Now we write

$$\sum_{t=1}^{n} x_{t-1}v_t = \pi(1)\sum_{t=1}^{n} x_{t-1}\varepsilon_t + \sum_{t=1}^{n} x_{t-1}\left(\tilde{v}_{t-1} - \tilde{v}_t\right).$$

We have

$$\sum_{t=1}^{n} x_{t-1}\varepsilon_t = O_p\left(\frac{n}{\sqrt{m}}\right).$$

Moreover,

$$\begin{aligned}\sum_{t=1}^{n} x_{t-1}\left(\tilde{v}_{t-1} - \tilde{v}_t\right) &= \sum_{t=1}^{n}(x_t - x_{t-1})\tilde{v}_t + O_p(\sqrt{n}) \\ &= \sum_{t=1}^{n} v_t\tilde{v}_t - \frac{m}{n}\sum_{t=1}^{n} x_{t-1}\tilde{v}_t + O_p(\sqrt{n}) \\ &= -\frac{m}{n}\sum_{t=1}^{n} x_{t-1}\tilde{v}_t + O_p(n)\end{aligned}$$

and

$$\left|\sum_{t=1}^{n} x_{t-1}\tilde{v}_t\right| \le \left(\sum_{t=1}^{n} x_{t-1}^2\right)^{1/2}\left(\sum_{t=1}^{n} \tilde{v}_t^2\right)^{1/2} = O_p\left(\frac{n}{\sqrt{m}}\right)O_p(\sqrt{n}),$$

from which we may deduce

$$\sum_{t=1}^{n} x_{t-1}v_t = O_p\left(\frac{n}{\sqrt{m}}\right) + \frac{m}{n}O_p\left(\frac{n}{\sqrt{m}}\right)O_p(\sqrt{n}) + O_p(n) = O_p(n).$$

The proof is now complete. ∎

6.2 Proofs of Theorems

Proof of Lemma 2.3 It can be deduced after recursive substitution that

$$n^{-1/2}x_{i+1} = n^{-1/2}\sum_{j=1}^{i+1} v_j - (1-\alpha)\sum_{j=1}^{i}\alpha^{i-j}\left(n^{-1/2}\sum_{k=1}^{j} v_k\right). \tag{57}$$

See Stock (1994) for more detailed explanation for the derivation. We define

$$V_{0n}(r) = n^{-1/2}\sum_{i=1}^{[nr]+1} v_i \tag{58}$$

and note that

$$\alpha = \exp(-m/n) + O((m/n)^2). \tag{59}$$

Then we have for all i

$$
\begin{aligned}
& m\int_0^{i/n}\exp\left(-m\left(\frac{i}{n}-s\right)\right)V_{0n}(s)\,ds \\
& = \left(1-\exp\left(-\frac{m}{n}\right)\right)\sum_{j=1}^{i}\left(\exp\left(-\frac{m}{n}\right)\right)^{i-j}\left(n^{-1/2}\sum_{k=1}^{j}v_k\right) \qquad (60)
\end{aligned}
$$

and for any $i/n < r < (i{+}1)/n$

$$
\begin{aligned}
m\int_{i/n}^{r}\exp(-m(r-s))V_{0n}(s)\,ds &= V_{0n}\left(\frac{i}{n}\right)m\int_{i/n}^{r}\exp(-m(r-s))\,ds \\
&= V_{0n}\left(\frac{i}{n}\right)\left(1-\exp\left(-m\left(r-\frac{i}{n}\right)\right)\right) \\
&= O_p(m/n) \qquad (61)
\end{aligned}
$$

uniformly in i.

It follows from (57)–(61) that

$$
V_{mn}(r) = V_{0n}(r) - m\int_0^r \exp(-m(r-s))V_{0n}(s)\,ds + O_p(mn^{-1}). \qquad (62)
$$

However, we may have, by expanding the underlying probability space if necessary, Brownian motion V_0 such that

$$
\sup_{0\le r\le 1}|V_{0n}(r)-V_0(r)| = o_p(n^{-1/2+1/p}), \qquad (63)
$$

which follows from the strong approximation result for the linear process by e.g., Akonom (1993). We therefore have from (62) and (63)

$$
V_{mn}(r) = V_0(r) - m\int_0^r \exp(-m(r-s))V_0(s)\,ds + o_p(n^{-1/2+1/p}) + O_p(mn^{-1}). \qquad (64)
$$

The stated result now can easily be deduced from (64) upon noticing that

$$
\int_0^r \exp(-m(r-s))\,dV_0(s) = V_0(r) - m\int_0^r \exp(-m(r-s))V_0(s)\,ds,
$$

which is due to the integration by parts formula for stochastic integrals. ∎

Proof of Lemma 2.4 The first part follows from Bosq (1999, Theorem 6.11, p163). The conditions (i) and (iii) there are clearly satisfied, and the requirements (ii) and (iiii) are shown to hold in parts (a) and (b) of Lemma A1. The continuity of $\mathbf{E}L(T,\cdot)$, while it certainly holds in our case, is not required here, since D is Lipschitz over the entire $\mathbf{R}$.

To prove the second part, we let

$$
\Delta_m(x) = (D_m(x)-D(x))1\left\{|x|\le(\log m)^{1/2}\right\}.
$$

Then it follows that

$$|D_m(x) - D(x)| \leq |\Delta_m(x)| + D(x)1\left\{|x| > (\log m)^{1/2}\right\}$$

due to part (c) of Lemma A1. Obviously, we have

$$\int_{-\infty}^{\infty} |x|^k D(x)1\{|x| > (\log m)^{1/2}\}\, dx \to 0$$

as $m \to \infty$, and therefore, it suffices to show that

$$\int_{-\infty}^{\infty} |x|^k |\Delta_m(x)|\, dx \to 0 \tag{65}$$

as $m \to \infty$. As in Bosq (1999, p163), we may assume that m is an integer.

For the proof of (65), we first define

$$L_t(x) = L(t,x) - L(t-1,x) \tag{66}$$

and notice that $\mathbf{E}L_t(x) = D(x)$ for $1 \leq t \leq m$. Then it follows for $|x| \leq (\log m)^{1/2}$ that

$$\Delta_m(x) = \frac{1}{m}\sum_{i=1}^{m}(L_t(x) - \mathbf{E}L_t(x)).$$

Part (a) of Lemma A1 entails Cramer's conditions, and therefore, we have for $|x| \leq (\log m)^{1/2}$

$$\mathbf{P}\{c_m|\Delta_m(x)| > \eta\} \leq Km^{-\log\log m}, \tag{67}$$

where

$$c_m = m^{1/2}(\log m \log\log m)^{-1}$$

and K is some constant.

We now let

$$\delta_m = m^{-(1+2\epsilon)/(1-2\epsilon)}$$

for $\epsilon > 0$ given in part (b) of Lemma A1, and define $\delta_m(i) = [(i-1)\delta_m, i\delta_m]$ for the values of i's just enough to ensure $[-(\log m)^{1/2}, (\log m)^{1/2}] \subset \bigcup_i \delta_m(i)$. For $x \in \delta_m(i)$, we have

$$|\Delta_m(x)| \leq |D_m(x) - D_m(i\delta_m)| + |\Delta_m(i\delta_m)| + |D(x) - D(i\delta_m)|. \tag{68}$$

Let Z_t be the random variable associated with L_t in part (b) of Lemma A1. Then it follows that

$$|D_m(x) - D_m(i\delta_m)| \leq \frac{1}{m}\sum_{t=1}^{m}|L_t(x) - L_t(i\delta_m)| \leq \delta_m^{1/2-\epsilon}\frac{1}{m}\sum_{t=1}^{m} Z_t.$$

Remark that the bound does not depend upon i, and $\sum_{t=1}^m Z_t/m = O(1)$ a.s. by the ergodic theorem. Therefore, we have

$$c_m \max_i \sup_{x\in\delta_m(i)} |D_m(x) - D_m(i\delta_m)| \leq c_m\delta_m^{1/2-\epsilon} \to_{a.s.} 0 \tag{69}$$

as $m \to \infty$.

On the other hand, it follows from (67) that

$$\mathbf{P}\left\{\max_i c_m|\Delta_m(i\delta_m)| > \eta\right\} \le K\delta_m^{-1}m^{-\log\log m}\log m$$

and Borel-Cantelli lemma entails

$$\max_i c_m|\Delta_m(i\delta_m)| \to_{a.s.} 0 \tag{70}$$

as $m \to \infty$. Finally, we have

$$\max_i \sup_{x\in\delta_m(i)} |D(x) - D(i\delta_m)| \le \sqrt{2/\pi e}\,\delta_m$$

and

$$c_m \max_i \sup_{x\in\delta_m(i)} |D(x) - D(i\delta_m)| \le \sqrt{2/\pi e}\,c_m\delta_m \to_{a.s.} 0 \tag{71}$$

as $m \to \infty$.

Now it follows from (68)–(71) that

$$\sup_{x\in\mathbf{R}} |\Delta_m(x)| = O(m^{-1/2}\log m\log\log m) \text{ a.s.}$$

and we may easily deduce that

$$\begin{aligned}\int_{-\infty}^{\infty} |x|^k|\Delta_m(x)|\,dx &= \int_{|x|\le(\log m)^{1/2}} |x|^k|\Delta_m(x)|\,dx \\ &= O_{a.s.}\left(\frac{(\log m)^{(k+3)/2}\log\log m}{\sqrt{m}}\right).\end{aligned}$$

The result stated in the second part now follows immediately.

Proof of Lemma 3.3 We assume without loss of generality that $x_0 = 0$, and that the support of T is included on $\mathbf{R}_+$. Also, we assume, by taking piece by piece if necessary, that T satisfies the conditions in Definition 3.2 over its entire support. In our subsequent proof, $\epsilon > 0$ denotes an arbitrarily small number, which may vary from line to line. Denote by (κ_{mn}) and (δ_{mn}) the sequences of numbers satisfying

$$\delta_{mn} = \min(m^{-1}n^{-1/p-\epsilon}, m^{-1/2}n^{-1/6-2/3p}),\ \ \kappa_{mn}\delta_{mn} = n^{1/p+\epsilon} \tag{72}$$

if $n^{1/2+1/p} \ge m$, and

$$\delta_{mn} = \min(m^{-2}n^{1/2-\epsilon}, m^{-7/6}n^{1/6}),\ \ \kappa_{mn}\delta_{mn} = n^{1/p+\epsilon} \tag{73}$$

if $m \ge n^{1/2+1/p}$. Note that

$$m^{1/2}n^{1/2}\delta_{mn} \ge 1 \tag{74}$$

for both cases (72) and (73), given the assumption $m = o\left(n^{1-2/p} \wedge n^{2/3}\right)$. This is required in the subsequent proof. Note also that

$$m\kappa_{mn}\delta_{mn}^2 \leq 1 \tag{75}$$

for both cases (72) and (73).

We define

$$\Delta_{mn} = \sup_{0\leq r\leq 1} |V_{mn}(r) - V_m(r)|\,. \tag{76}$$

It follows that

$$\sqrt{n}\Delta_{mn} = o_p(n^{1/p}) + O_p(n^{-1/2}m) = o_p(\kappa_{mn}\delta_{mn})$$

and

$$\kappa_{mn}\delta_{mn} \pm 2\sqrt{n}\Delta_{mn} \geq \kappa_{mn}\delta_{mn}(1 + o_p(1)).$$

Moreover, we let

$$\begin{aligned} T_{mn}(x) &= T(x)1\{0 \leq x < \kappa_{mn}\delta_{nm}\}, \\ T_{nm}(x) &= \sum_{k=1}^{\kappa_{mn}} T(k\delta_{mn})1\{(k-1)\delta_{mn} \leq x < k\delta_{mn}\}. \end{aligned}$$

The function T_{mn} is a truncated version of T, and T_{nm} is a simple function approximating T_{mn}.

First, we show that

$$\begin{aligned} \frac{1}{\sqrt{nm}}\sum_{t=1}^{n} T(x_t) - \frac{1}{\sqrt{nm}}\sum_{t=1}^{n} T_{mn}(x_t) &=_d \sqrt{\frac{n}{m}}\int_0^1 (T - T_{mn})(\sqrt{n}V_{mn}(r))dr \\ &= o_p(n^{-1/6+1/3p+\epsilon}) \end{aligned} \tag{77}$$

uniformly in $m \in \mathbf{R}_+$. By taking n (and hence m) sufficiently large, we may assume that $T - T_{mn}$ is monotone decreasing on its support. Then it follows that

$$\begin{aligned} &\sqrt{\frac{n}{m}}\int_0^1 (T - T_{mn})(\sqrt{n}V_{mn}(r))dr \\ &\leq \sqrt{\frac{n}{m}}\int_0^1 T(\sqrt{n}(V_m(r) - \Delta_{mn}))1\{\sqrt{n}(V_m(r) + \Delta_{mn}) > \kappa_{mn}\delta_{mn}\}dr \\ &= \frac{1}{m}\sqrt{\frac{n}{m}}\int_0^1 T\left(\sqrt{\frac{n}{m}}V(r) - \sqrt{n}\Delta_{mn}\right)1\left\{\sqrt{\frac{n}{m}}V(r) + \sqrt{n}\Delta_{mn} > \kappa_{mn}\delta_{mn}\right\}dr \\ &= \frac{1}{m}\sqrt{\frac{n}{m}}\int_{-\infty}^{\infty} T\left(\sqrt{\frac{n}{m}}x - \sqrt{n}\Delta_{mn}\right)1\left\{\sqrt{\frac{n}{m}}x + \sqrt{n}\Delta_{mn} > \kappa_{mn}\delta_{mn}\right\}L(m,x)dx \\ &= \frac{1}{m}\int_{-\infty}^{\infty} T(x)1\{x > \kappa_{mn}\delta_{mn} - 2\sqrt{n}\Delta_{mn}\}L\left(m, \sqrt{\frac{m}{n}}x + \sqrt{m}\Delta_{mn}\right)dx \\ &\leq \int_{-\infty}^{\infty} T(x)1\{x > \kappa_{mn}\delta_{mn}(1 + o_p(1))\}dx \\ &\leq \int_{-\infty}^{\infty} |T(x)|1\{x > n^{1/p+\epsilon}(1 + o_p(1))\}dx \\ &= o_p(n^{-1/6+1/3p-\epsilon}) \end{aligned}$$

uniformly in $m \in \mathbf{R}_+$, for some small $\epsilon > 0$.

Second, we have from the Lipschitz condition for T that

$$\begin{aligned}
&\left| \frac{1}{\sqrt{nm}} \sum_{t=1}^n T_{mn}(x_t) - \frac{1}{\sqrt{nm}} \sum_{t=1}^n T_{nm}(x_t) \right| \\
&\quad \leq \; c\kappa_{mn}\delta_{mn}^2 \frac{1}{\kappa_{mn}\delta_{mn}\sqrt{nm}} \sum_{t=1}^n 1\{0 \leq x_t < \kappa_{mn}\delta_{mn}\} \\
&\quad = \; O_p(\kappa_{mn}\delta_{mn}^2) \\
&\quad = \; m^{-1/2}\left(o_p(n^{-1/6+1/3p+\epsilon}) + o_p(m^{1/3}n^{-1/3+\epsilon}) \right) \qquad (78)
\end{aligned}$$

uniformly in $m \in \mathbf{R}_+$ for some constant c, since we have in particular

$$\begin{aligned}
&\frac{1}{\kappa_{mn}\delta_{mn}\sqrt{nm}} \sum_{t=1}^n 1\{0 \leq x_t < \kappa_{mn}\delta_{mn}\} \\
&=_d \; \frac{1}{\kappa_{mn}\delta_{mn}} \sqrt{\frac{n}{m}} \int_0^1 1\{0 \leq \sqrt{n}V_{mn}(r) < \kappa_{mn}\delta_{mn}\} dr \\
&\leq \; \frac{1}{\kappa_{mn}\delta_{mn}} \sqrt{\frac{n}{m}} \int_0^1 1\{0 \leq \sqrt{n}V_m(r) < \kappa_{mn}\delta_{mn} + \sqrt{n}\Delta_{mn}\} dr \\
&= \; \frac{1}{m}\frac{1}{\kappa_{mn}\delta_{mn}} \sqrt{\frac{n}{m}} \int_0^m 1\left\{0 \leq \sqrt{\frac{n}{m}}V(r) < \kappa_{mn}\delta_{mn}(1+o_p(1))\right\} dr \\
&= \; \frac{1}{m}\frac{1}{\kappa_{mn}\delta_{mn}} \sqrt{\frac{n}{m}} \int_{-\infty}^\infty 1\left\{0 \leq \sqrt{\frac{n}{m}}x < \kappa_{mn}\delta_{mn}(1+o_p(1))\right\} L(m,x) dr \\
&= \; \frac{1}{m} \int_{-\infty}^\infty 1\{0 \leq x < 1 + o_p(1)\} L\left(m, \kappa_{mn}\delta_{mn}\sqrt{\frac{m}{n}}x\right) dx \\
&= \; O_p(1)
\end{aligned}$$

uniformly in $m \in \mathbf{R}_+$.

Third, we have

$$\begin{aligned}
\frac{1}{\sqrt{nm}} \sum_{t=1}^n T_{nm}(x_t) \;=\; & \left(\int_{-\infty}^\infty T(x)dx\right) \frac{1}{\delta_{mn}\sqrt{nm}} \sum_{t=1}^n 1\{0 \leq x_t < \delta_{mn}\} \\
& + o_p(n^{-1/6+1/3p+\epsilon}) + o_p(m^{1/4}n^{-1/4+1/2p+\epsilon}) \\
& + o_p(m^{1/3}n^{-1/3+\epsilon}) + o_p(m^{3/4}n^{-1/2+\epsilon}) \qquad (79)
\end{aligned}$$

uniformly in $m \in \mathbf{R}$, for any $\epsilon > 0$. To show (79), we first note that

$$\mathbf{E}\left(\sum_{t=1}^n 1\{0 \leq x_t < \delta\} - \sum_{t=1}^n 1\{(k-1)\delta \leq x_t < k\delta\} \right)^2 \leq c\, m^{1/2}n^{1/2}\delta(1 + mk\delta^2 \log n), \quad (80)$$

where c is some number, which is dependent only upon the distribution of (ε_t) and bounded by some absolute constant. We may deduce (80) exactly as in Akonom (1993) upon noticing that the density and its first derivative of (x_t) are bounded uniformly by the constant multiples of $(m/t)^{1/2}$ and m/t, respectively. The proof for (80) goes precisely as the proof of Lemma 6 in Akonom (1993) for simple random walks, and can be extended for more general linear processes following the arguments used in his proof of Lemma 13. Note that the condition (74) is required to obtain (80).

As we noted earlier, we have (75) given our choices of δ_{mn} and κ_{mn} in (72) and (73). Therefore, it follows that

$$\sum_{t=1}^{n} 1\{0 \leq x_t < \delta_{mn}\} = \sum_{t=1}^{n} 1\{(k-1)\delta_{mn} \leq x_t < k\delta_{mn}\} + o_p(m^{1/4}n^{1/4+\epsilon}\delta_{mn}^{1/2})$$

uniformly in $m \in \mathbf{R}_+$ and $k = 1, \ldots, \kappa_{mn}$. However, if $n^{1/2+1/p} \geq m$, then

$$m^{1/4}n^{1/4+\epsilon}\delta_{mn}^{1/2} = \delta_{mn}\sqrt{nm}\left(o(n^{-1/6+1/3p+\epsilon}) + o(m^{1/4}n^{-1/4+1/2p+\epsilon})\right)$$

and, if $m \geq n^{1/2+1/p}$, then

$$m^{1/4}n^{1/4+\epsilon}\delta_{mn}^{1/2} = \delta_{mn}\sqrt{nm}\left(o(m^{1/3}n^{-1/3+\epsilon}) + o(m^{3/4}n^{-1/2+\epsilon})\right).$$

Consequently, it follows that

$$\left(\int_{-\infty}^{\infty} T_{nm}(x)dx\right)\frac{1}{\delta_{mn}\sqrt{nm}}\sum_{t=1}^{n} 1\{0 \leq x_t < \delta_{mn}\} = \frac{1}{\sqrt{nm}}\sum_{t=1}^{n} T_{nm}(x_t)$$
$$+\, o_p(n^{-1/6+1/3p+\epsilon}) + o_p(m^{1/4}n^{-1/4+1/2p+\epsilon}) + o_p(m^{1/3}n^{-1/3+\epsilon}) + o_p(m^{3/4}n^{-1/2+\epsilon}).$$

Note that

$$\begin{aligned}\int_{-\infty}^{\infty} T_{nm}(x)dx &= \int_{-\infty}^{\infty} T_{mn}(x)dx + O(\kappa_{mn}\delta_{mn}^2)\\ &= \int_{-\infty}^{\infty} T(x)dx + m^{-1/2}\left(o(n^{-1/6+1/3p+\epsilon}) + o(m^{1/3}n^{-1/3+\epsilon})\right)\end{aligned}$$

from which we may easily derive (79).

Fourth, it can be deduced that

$$\begin{aligned}\frac{1}{\delta_{mn}\sqrt{nm}}\sum_{t=1}^{n} 1\{0 \leq x_t < \delta_{mn}\} &= \frac{1}{\delta_{mn}}\sqrt{\frac{n}{m}}\int_0^1 1\{0 \leq \sqrt{n}V_m(r) < \delta_{mn}\}dr\\ &\quad + o_p(n^{-1/6+1/3p+\epsilon}) + o_p(m^{1/3}n^{-1/3+\epsilon}). \qquad (81)\end{aligned}$$

To show (81), we may proceed as in the proof of Theorem 8 in Akonom (1993) and establish using (80) and Lemma 2.3 that the terms $A_1(n,m), A_2(n,m)$ and $A_3(n,m)$ corresponding to $A_1(n), A_2(n)$ and $A_3(n)$ defined in Akonom (1993) have orders given by

$$\begin{aligned}A_1(n,m), A_3(n,m) &= o_p(m^{1/4}n^{1/4+\epsilon}\delta^{1/2}(1+mk\delta^2)^{1/2}),\\ A_2(n,m) &= o_p(n^{1/2+1/p}) + O_p(n^{-1/2}m).\end{aligned}$$

When $n^{1/2+1/p} \geq m$, we may choose $k^{-1} = m^{1/2}n^{-1/6-2/3p}\delta$ to get

$$A_1(n,m), k^{-1}A_2(n,m), A_3(n,m) = o_p(m^{1/2}n^{1/3+1/3p+\epsilon}\delta) = \delta\sqrt{nm}o_p(n^{-1/6+1/3p+\epsilon}).$$

If, on the other hand, $m \geq n^{1/2+1/p}$, then we have

$$A_1(n,m), k^{-1}A_2(n,m), A_3(n,m) = o_p(m^{5/6}n^{1/6+\epsilon}\delta) = \delta\sqrt{nm}o_p(m^{1/3}n^{-1/3+\epsilon})$$

with the choice of $k^{-1} = (n/m)^{1/6}\delta$. The result in (81) now follows immediately.

Fifth, it follows from part (b) of Lemma A1 that

$$\frac{1}{m}|L(m,a) - L(m,b)| \leq \frac{1}{m}\sum_{t=1}^{m}|L_t(a) - L_t(b)| \leq |a-b|^{1/2-\epsilon}\frac{1}{m}\sum_{t=1}^{m}Z_t,$$

where L_t is defined in (66) and Z_t is the random variable associated with L_t, and therefore,

$$\frac{1}{m}|L(m,a) - L(m,b)| = O(|a-b|^{1/2-\epsilon}) \text{ a.s.} \tag{82}$$

We have

$$\begin{aligned}
&\frac{1}{\delta_{mn}}\sqrt{\frac{n}{m}}\int_0^1 1\{0 \leq \sqrt{n}V_m(r) < \delta_{mn}\}dr \\
&= \frac{1}{m\delta_{mn}}\sqrt{\frac{n}{m}}\int_0^m 1\left\{0 \leq \sqrt{\frac{n}{m}}V(r) < \delta_{mn}\right\}dr \\
&= \frac{1}{m\delta_{mn}}\sqrt{\frac{n}{m}}\int_{-\infty}^{\infty} 1\left\{0 \leq \sqrt{\frac{n}{m}}x < \delta_{mn}\right\}L(m,x)dr \\
&= \frac{1}{m}\int_{-\infty}^{\infty} 1\{0 \leq x < 1\}L\left(m, \delta_{mn}\sqrt{\frac{m}{n}}x\right)dx \\
&= D_m(0) + \frac{1}{m}\int_{-\infty}^{\infty} 1\{0 \leq x < 1\}\left[L\left(m, \delta_{mn}\sqrt{\frac{m}{n}}x\right) - L(m,0)\right]dx \\
&= D_m(0) + O(\delta_{mn}^{1/2}m^{1/4}n^{-1/4+\epsilon}) \text{ a.s.}
\end{aligned} \tag{83}$$

and

$$\begin{aligned}
\sqrt{\frac{n}{m}}\int_0^1 T(\sqrt{n}V_m(r))dr &= \frac{1}{m}\sqrt{\frac{n}{m}}\int_0^m T\left(\sqrt{\frac{n}{m}}V(r)\right)dr \\
&= \frac{1}{m}\sqrt{\frac{n}{m}}\int_{-\infty}^{\infty} T\left(\sqrt{\frac{n}{m}}x\right)L(m,x)dx \\
&= \frac{1}{m}\int_{-\infty}^{\infty} T(x)L\left(m, \sqrt{\frac{m}{n}}\right)dx \\
&= D_m(0)\int_{-\infty}^{\infty} T(x)dx \\
&\quad + \frac{1}{m}\int_{-\infty}^{\infty} T(x)\left[L\left(m, \sqrt{\frac{m}{n}}\right) - L(m,0)\right]dx \\
&= D_m(0)\int_{-\infty}^{\infty} T(x)dx + O(m^{1/4}n^{-1/4+\epsilon}) \text{ a.s.,}
\end{aligned} \tag{84}$$

which are due in particular to (82).

We may now easily deduce from (77), (78), (79), (81), (83) and (84) that

$$\sqrt{\frac{n}{m}}\int_0^1 T(\sqrt{n}V_{mn}(r))dr = \sqrt{\frac{n}{m}}\int_0^1 T(\sqrt{n}V_m(r))dr + R_{mn},$$

where

$$R_{mn} = o_p\left(\max\left(n^{-1/6+1/3p}, m^{1/4}n^{-1/4+1/2p}, m^{1/3}n^{-1/3}, m^{3/4}n^{-1/2}\right)^{1-\epsilon}\right)$$

for some $\epsilon > 0$. The proof is therefore complete.

Proof of Theorem 3.4 The mean asymptotics follow directly from (14). For the covariance asymptotics, we invoke the CLT for general martingale difference sequences in Hall and Hyde (1981, Corollary 3.1). We have for any $\epsilon > 0$

$$\begin{aligned}
&\sum_{t=1}^{n}\mathbf{E}\left(\left|\frac{1}{\sqrt[4]{nm}}F(x_t)\,u_t\right|^{2+\epsilon}\middle|\mathcal{F}_{t-1}\right) \\
&\quad\le (nm)^{-\epsilon/4}\left(\sup_{1\le t\le n}\mathbf{E}\left(|u_t|^{2+\epsilon}\,|\mathcal{F}_{t-1}\right)\right)\frac{1}{\sqrt{nm}}\sum_{t=1}^{n}|F(x_t)|^{2+\epsilon} \\
&\quad= O_p((nm)^{-\epsilon/4}),
\end{aligned}$$

where $(\mathcal{F}_t)$ is the filtration introduced in Assumption 3.1, and therefore, the conditional Lindeberg condition is satisfied. Remark that if F is regularly integrable, so is $|F|^{2+\epsilon}$ for any $\epsilon > 0$.

To obtain the stated result, we simply note that

$$\mathbf{E}\left((F(x_t)\,u_t)^2\middle|\,\mathcal{F}_{t-1}\right) = \sigma^2 F^2(x_t)$$

and apply (14) with $T = F^2$ to deduce

$$\frac{1}{\sqrt{nm}}\sum_{t=1}^{n}F^2(x_t) \to_p D(0)\int_{-\infty}^{\infty}F^2(x)\,dx$$

as $n \to \infty$. The stated covariance asymptotics now follow immediately from Hall and Heyde (1981, Corollary 3.1).

Proof of Lemma 3.7 Let Δ_{mn} be defined as in (76), and let (c_{mn}) be a random sequence such that $c_{mn} \ge \sqrt{m}\Delta_{mn}$ and $c_{mn} \to_p 0$ as $n \to 0$. We write

$$\left|S(\sqrt{m}V_{mn}(r)) - S(\sqrt{m}V_m(r))\right| \le A_{mn}(r) + B_{mn}(r) + C_{mn}(r) + D_{mn}(r),$$

where

$$\begin{aligned}
A_{mn}(r) &= \left|S(\sqrt{m}V_{mn}(r)) - S(\sqrt{m}V_m(r))\right| 1\left\{\sqrt{m}|V_{mn}(r)| \geq c_{mn}\right\} 1\left\{\sqrt{m}|V_m(r)| \geq c_{mn}\right\}, \\
B_{mn}(r) &= \left|S(\sqrt{m}V_{mn}(r)) - S(\sqrt{m}V_m(r))\right| 1\left\{\sqrt{m}|V_{mn}(r)| < c_{mn}\right\} 1\left\{\sqrt{m}|V_m(r)| \geq c_{mn}\right\}, \\
C_{mn}(r) &= \left|S(\sqrt{m}V_{mn}(r)) - S(\sqrt{m}V_m(r))\right| 1\left\{\sqrt{m}|V_{mn}(r)| \geq c_{mn}\right\} 1\left\{\sqrt{m}|V_m(r)| < c_{mn}\right\}, \\
D_{mn}(r) &= \left|S(\sqrt{m}V_{mn}(r)) - S(\sqrt{m}V_m(r))\right| 1\left\{\sqrt{m}|V_{mn}(r)| < c_{mn}\right\} 1\left\{\sqrt{m}|V_m(r)| < c_{mn}\right\}
\end{aligned}$$

for every $r \in [0, 1]$.

Due to the condition in part (a) of Definition 3.5, we have

$$A_{mn}(r) \leq K\left(1 + \left|\sqrt{m}V_m(r)\right|^a\right)(1 + c_{mn}^b)\sqrt{m}\,|V_{mn}(r) - V_m(r)|$$

for all large n. However,

$$\begin{aligned}
\int_0^1 \left(1 + \left|\sqrt{m}V_m(r)\right|^a\right) dr &= \frac{1}{m}\int_0^m (1 + |V(r)|^a)\, dr \\
&= \frac{1}{m}\int_{-\infty}^{\infty} (1 + |x|^a)\, L(m, x)\, dr \\
&\to_{a.s.} \int_{-\infty}^{\infty} (1 + |x|^a)\, D(x)\, dr
\end{aligned}$$

as $m \to \infty$, and it therefore follows that

$$\int_0^1 A_{mn}(r)\, dr = O_p\left(\sqrt{m}(1 + c_{mn}^b)\Delta_{mn}\right) \tag{85}$$

for all large n.

Since for all n sufficiently large

$$\left\{\left|\sqrt{m}V_m(r)\right| \geq c_{mn}\right\} \subset \left\{\left|\sqrt{m}V_{mn}(r)\right| \geq c_{mn} - \Delta_{mn}\right\}$$

and

$$\left\{\left|\sqrt{m}V_{mn}(r)\right| \geq c_{mn}\right\} \subset \left\{\left|\sqrt{m}V_m(r)\right| \geq c_{mn} - \Delta_{mn}\right\},$$

we may also easily deduce that

$$\int_0^1 B_{mn}(r)\, dr,\ \int_0^1 C_{mn}(r)\, dr = O_p\left(\sqrt{m}(1 + c_{mn}^b)\Delta_{mn}\right) \tag{86}$$

for all large n. Remark that $c_{mn} \geq \sqrt{m}\Delta_{mn}$.

Finally, we have

$$\int_0^1 D_{mn}(r)\, dr \leq 2(1 + o_p(1))\int_0^1 \left|S\left(\sqrt{m}V_m(r)\right)\right| 1\left\{\sqrt{m}|V_m(r)| < c_{mn}\right\} dr$$

for large n, since, in particular,

$$\begin{aligned}
&\int_0^1 \left|S\left(\sqrt{m}V_{mn}(r)\right)\right| 1\left\{\sqrt{m}|V_{mn}(r)| < c_{mn}\right\} dr \\
&\quad = (1 + o_p(1))\int_0^1 \left|S\left(\sqrt{m}V_m(r)\right)\right| 1\left\{\sqrt{m}|V_m(r)| < c_{mn}\right\} dr
\end{aligned}$$

for large n. Moreover, due to the condition in part (b) of Definition 3.5,

$$\left|S\left(\sqrt{m}V_m(r)\right)\right| 1\left\{\sqrt{m}|V_m(r)| < c_{mn}\right\} \leq \left|\sqrt{m}V_m(r)\right|^c 1\left\{\sqrt{m}|V_m(r)| < c_{mn}\right\}$$

for all $r \in [0,1]$ and for all n sufficiently large, and we have

$$\begin{aligned}\int_0^1 \left|\sqrt{m}V_m(r)\right|^c 1\left\{\sqrt{m}|V_m(r)| < c_{mn}\right\} dr &= \frac{1}{m}\int_0^m |V(r)|^c 1\{|V(r)| < c_{mn}\} \\ &= \frac{1}{m}\int_{-\infty}^{\infty} |x|^c 1\{|x| < c_{mn}\} L(m,x)\, dx \\ &\leq \left(\sup_{x\in\mathbf{R}} \frac{L(m,x)}{m}\right)\int_{-\infty}^{\infty} |x|^c 1\{|x| < c_{mn}\}\, dx \\ &= (D(0) + o(1))\,\frac{c_{mn}^{1+c}}{1+c} \text{ a.s.,}\end{aligned}$$

due in particular to Lemma 2.4. It therefore follows that

$$\int_0^1 D_{mn}(r)\, dr = O_p(c_{mn}^{1+c}) \tag{87}$$

for large n.

We now have from (85)–(87)

$$\int_0^1 S(\sqrt{m}V_{mn}(r)\, dr = \int_0^1 S(\sqrt{m}V_m(r)\, dr + R_{mn},$$

where

$$R_{mn} = O_p\left(\sqrt{m}(1 + c_{mn}^b)\Delta_{mn}\right) + O_p\left(c_{mn}^{1+c}\right).$$

For $c \geq 0$ and $b \geq 0$, we may choose $c_{mn} = \sqrt{m}\Delta_{mn}$ to show that $R_{mn} = O_p(\sqrt{m}\Delta_{mn})$. If $c \leq b$ and $-1 < c < 0$, then the same choice of c_{mn} gives the optimal rate $R_{mn} = O_p((\sqrt{m}\Delta_{mn})^{1+c})$. Finally, when $c > -1$, $c > b$ and $b < 0$, the optimal choice of c_{mn} reduces to $c_{mn} = (\sqrt{m}\Delta_{mn})^{1/(1+c-b)}$, which yields $R_{mn} = O_p((\sqrt{m}\Delta_{mn})^{(1+c)/(1+c-b)})$. To complete the proof, note that

$$\sqrt{m}\Delta_{mn} = o_p(m^{1/2}n^{-1/2+1/p}) + O_p(m^{3/2}n^{-1}),$$

which becomes of order $o_p(1)$ if $m = o\,(n^{1-2/p} \wedge n^{2/3})$ as we assume here.

Proof of Theorem 3.8 The mean asymptotics follow immediately from (17) and (20). The covariance asymptotics can be obtained using the CLT for general martingale sequences used in the proof of Theorem 3.4. To establish the required conditional Linderberg condition, we first note that if F^2 is regularly homogeneous, then so is $|F|^{2+\epsilon}$ for $\epsilon > 0$ sufficiently small. Then it follows that

$$\begin{aligned}
&\sum_{t=1}^{n} \mathbf{E}\left(\left| \frac{1}{\sqrt{n}} \kappa\left(\sqrt{\frac{n}{m}}\right)^{-1} F(x_t)\, u_t \right|^{2+\epsilon} \middle| \mathcal{F}_{t-1} \right) \\
&\quad \le n^{-\epsilon/2} \left(\sup_{1\le t\le n} \mathbf{E}\left(|u_t|^{2+\epsilon} \, |\mathcal{F}_{t-1} \right) \right) \frac{1}{n} \kappa\left(\sqrt{\frac{n}{m}}\right)^{-2-\epsilon} \sum_{t=1}^{n} |F(x_t)|^{2+\epsilon} \\
&\quad = O_p(n^{-\epsilon/2}),
\end{aligned}$$

where $(\mathcal{F}_t)$ is the filtration introduced in Assumption 3.1.

To deduce the stated covariance asymptotics, we note

$$\mathbf{E}\left((F(x_t)\, u_t)^2 \middle| \mathcal{F}_{t-1} \right) = \sigma^2 F^2(x_t)$$

and use the results in (17) and (20) with $T = F^2$ and $S = H^2$ to derive

$$\frac{1}{n} \kappa\left(\sqrt{\frac{n}{m}}\right)^{-2} \sum_{t=1}^{n} F^2(x_t) \to_p \int_{-\infty}^{\infty} (HD)^2(x)\, dx$$

as $n \to \infty$. Remark that if F is regularly homogeneous with asymptotic order κ and limit homogeneous function H, and if F^2 is regularly homogeneous, then the asymptotic order and limit homogeneous function of F^2 are given respectively by κ^2 and H^2. The covariance asymptotics for regularly homogeneous functions can now be established directly from Hall and Heyde (1981, Corollary 3.1).

Proof of Theorem 4.2 Let $\sigma^2 = \mathbf{E}\varepsilon_t^2$. As is well known, we have

$$\frac{1}{n} \sum_{t=1}^{n} v_{tK} v_{tK}' \to_p \Gamma \tag{88}$$

and

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} v_{tK} \varepsilon_t \to_d \mathbf{N}(0, \sigma^2 \Gamma) \tag{89}$$

as $n \to \infty$.

We now apply the mean asymptotics in Theorem 3.8 for (x_t/ω) (that has the unit longrun variance) with $F(x) = x^2$, which has asymptotic order $\kappa(\lambda) = \lambda^2$ and limit homogeneous function $H(x) = x^2$, to get

$$\frac{m}{n^2} \sum_{t=1}^{n} x_{t-1}^2 \to_p \frac{\omega^2}{2} \tag{90}$$

as $n \to \infty$. Moreover, if applied for (x_t/ω) and $(u_t) = (\varepsilon_t)$ with $F(x) = x$ that has asymptotic order $\kappa(\lambda) = \lambda$ and limit homogeneous function $H(x) = x$, the covariance asymptotics in Theorem 3.8 yield

$$\frac{\sqrt{m}}{n} \sum_{t=1}^{n} x_{t-1} \varepsilon_t \to_d \mathbf{N}\left(0, \sigma^2 \frac{\omega^2}{2}\right) \tag{91}$$

as $n \to \infty$. The stated result now follows readily from (88)–(91) and Lemma A5.

Proof of Theorem 4.6 Given Lemma A3, the stated result can easily be obtained as in the proofs of Theorem 4.1 and 5.1 in Park and Phillips (2001). The details are therefore omitted.

Proof of Theorem 4.7 Given Lemma A4, the stated result follows essentially as in the proofs of Theorem 4.2 and 5.2 in Park and Phillips (2001). Our conditions here are, however, slightly different from those used by them. First, they assume that κ is bounded away from zero in the proof of Theorem 4.2. This is not assumed here, due to our extended asymptotics that are also applicable for functions having poles at the origin. Second, for the identification condition, we only require that $h(\cdot,\theta) \neq h(\cdot,\theta_0)$ for every $\theta \neq \theta_0$. This is in contrast with their unit root asymptotics, where it is necessary to have for all $\delta > 0$ that $\int_{|x|\leq\delta} h(x,\theta)\,dx \neq \int_{|x|\leq\delta} h(x,\theta_0)$ for every $\theta \neq \theta_0$. This is because the spatial distribution of the exact or near unit root processes is random and compactly supported a.s., while that of the weak unit root process is nonrandom and supported over the entire real line. For the same reason, our asymptotics are valid as long as $\dot{h}(\cdot,\theta_0)$ is linearly independent over the entire real line, while the unit root asymptotics require that $\int_{|x|\leq\delta} \dot{h}(x,\theta_0)\dot{h}(x,\theta_0)\,dx > 0$ for all $\delta > 0$.

Proof of Theorem 4.8 Our proof here is analogous to the proof of Theorem 5.3 in Park and Phillips (2001), which will be referred to as PP henceforth. Define

$$\dot{\kappa}_{mn}(\theta) = \dot{\kappa}\left(\sqrt{\frac{n}{m}},\theta\right),\ \ddot{\kappa}_{mn}(\theta) = \ddot{\kappa}\left(\sqrt{\frac{n}{m}},\theta\right),\ \dddot{\kappa}_{mn}(\theta) = \dddot{\kappa}\left(\sqrt{\frac{n}{m}},\theta\right).$$

Also, we let $\nu_{mn} = n^{1/2}\dot{\kappa}_{mn}(\theta_0)$ and define, for δ such that $0 < \delta < \epsilon/3$, $\mu_{mn} = n^{1/2-\delta}\dot{\kappa}_{mn}(\theta_0)$. Moreover, we let $(\dot{Q}_n, \ddot{Q}_n)$ (and $\ddot{Q}_n^\circ$) and $(\ddot{D}_{in})$ be defined as in PP. Subsequently, we also define

$$\varpi^2_{imn}(\theta) = \left\| \mu_{mn}^{-1} \ddot{D}_{in} \mu_{mn} \right\|$$

similarly as in PP.

To deduce the stated result, it suffices to show that the conditions AD1–AD4 and AD7 in PP hold with their ν_n and μ_n replaced by our ν_{mn} and μ_{mn}, respectively. The conditions AD1 and AD3 are immediate from Theorem 3.8. Also, since we have as $n \to \infty$

$$\begin{aligned}
&\left\| (\nu_{mn} \otimes \nu_{mn})^{-1} \sum_{t=1}^n \ddot{f}(x_t,\theta_0)\,u_t \right\| \\
&\quad \leq n^{-1/2} \left\| \left((\dot{\kappa}_{mn} \otimes \dot{\kappa}_{mn})^{-1} \ddot{\kappa}_{mn} \right)(\theta_0) \right\| \left\| \frac{1}{\sqrt{n}} \ddot{\kappa}_{mn}^{-1}(\theta_0) \sum_{t=1}^n \ddot{f}(x_t,\theta_0)\,u_t \right\| \to_p 0,
\end{aligned}$$

and $n^{-1/2} \leq (n/m)^{-1/2}$, the condition AD2 follows readily from (29) in (b). Furthermore, the conditon AD4 holds due to our identifiability assumption in (c).

To establish the condition AD7, we need to show that

$$\sup_{\theta \in N_{mn}} \varpi^2_{imn}(\theta) = o_p(1) \tag{92}$$

as $n \to \infty$, where N_{mn} is the neighborhood of θ_0 given by $N_{mn} = \{\|\mu'_{mn}(\theta - \theta_0)\| \leq 1\}$. However, we have for all $\theta \in N_{mn}$ that

$$\sum_{t=1}^{n} \left\| (\mu_{mn} \otimes \mu_{mn})^{-1}(\theta_0) \ddot{f}(x_t, \theta) \right\|^2 \leq n^{-1+4\delta} \left\| (\dot{\kappa}_{mn} \otimes \dot{\kappa}_{mn})^{-1}(\theta_0) \left(\sup_{\theta \in N_{mn}} \ddot{\kappa}_{mn}(\theta) \right) \right\| \frac{1}{n} \sum_{t=1}^{n} \left\| \ddot{\kappa}^{-1}(\theta) \ddot{f}(x_t, \theta) \right\|^2 \tag{93}$$

and

$$\sum_{t=1}^{n} \left\| (\mu_{mn} \otimes \mu_{mn} \otimes \mu_{mn})^{-1}(\theta_0) \dddot{f}(x_t, \theta) \right\|^2 \leq n^{-2+6\delta} \left\| (\dot{\kappa}_{mn} \otimes \dot{\kappa}_{mn} \otimes \dot{\kappa}_{mn})^{-1}(\theta_0) \left(\sup_{\theta \in N_{mn}} \dddot{\kappa}_{mn}(\theta) \right) \right\| \frac{1}{n} \sum_{t=1}^{n} \left\| \dddot{\kappa}^{-1}(\theta) \dddot{f}(x_t, \theta) \right\|^2 \tag{94}$$

and it follows from Lemma A4 that

$$\frac{1}{n} \sum_{t=1}^{n} \left\| \ddot{\kappa}^{-1}(\theta) \ddot{f}(x_t, \theta) \right\|^2, \ \frac{1}{n} \sum_{t=1}^{n} \left\| \dddot{\kappa}^{-1}(\theta) \dddot{f}(x_t, \theta) \right\|^2 = O_p(1) \tag{95}$$

as $n \to \infty$, uniformly in a neighborhood of θ_0. Now, under (30) and (31), (92) follows straightforwardly from (93)–(95) if we apply Cauch-Schwarz inequality to the results (58)–(61) in the proof of Theorem 5.3 in PP. Note that $n^{-1/2+3\delta} \leq (n/m)^{-1/2+\epsilon}$ if $m \to \infty$ as $n \to \infty$ and $0 < \delta < \epsilon/3$, as we assume here. The proof is therefore complete.

References

Akonom, J. (1993). "Comportement asymptotique du temps d'occupation du processus des sommes partielles," *Annales de l'Institut Henri Poincaré*, 29, 57–81.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes*, 2nd ed. Springer, New York.

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

Park, J.Y. and P.C.B. Phillips (1999). "Asymptotics for nonlinear transformations of integrated time series," *Econometric Theory*, 15, 269-298.

Park, J.Y. and P.C.B. Phillips (2001). "Nonlinear regressions with integrated time series," *Econometrica*, 69, 1452-1498.

Phillips, P.C.B. (1987). "Towards a unified asymptotic theory for autoregression," *Biometrika*, 74, 535-547.

Revuz, D. and M. Yor (1994). *Continuous Martingale and Brownian Motion.* New York: Springer–Verlag.

Stock, J.H. (1994). "Unit roots and structural breaks," In R.F. Engle and D. McFadden, eds., *Handbook of Econometrics*, Vol. 4, 2739-2841, Elsevier: Amsterdam.