NONLINEAR REGRESSIONS WITH INTEGRATED TIME SERIES

BY

JOON Y. PARK and PETER C.B. PHILLIPS

COWLES FOUNDATION PAPER NO. 1016

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2001

http://cowles.econ.yale.edu/

Econometrica, Vol. 69, No. 1 (January, 2001), 117–161

NONLINEAR REGRESSIONS WITH INTEGRATED TIME SERIES

BY JOON Y. PARK AND PETER C. B. PHILLIPS[1]

An asymptotic theory is developed for nonlinear regression with integrated processes. The models allow for nonlinear effects from unit root time series and therefore deal with the case of parametric nonlinear cointegration. The theory covers integrable and asymptotically homogeneous functions. Sufficient conditions for weak consistency are given and a limit distribution theory is provided. The rates of convergence depend on the properties of the nonlinear regression function, and are shown to be as slow as $n^{1/4}$ for integrable functions, and to be generally polynomial in $n^{1/2}$ for homogeneous functions. For regressions with integrable functions, the limiting distribution theory is mixed normal with mixing variates that depend on the sojourn time of the limiting Brownian motion of the integrated process.

KEYWORDS: Functionals of Brownian motion, integrated process, local time, mixed normal limit theory, nonlinear regression, occupation density.

1. INTRODUCTION AND HEURISTIC IDEAS

THE ASYMPTOTIC THEORY OF NONLINEAR REGRESSION plays a central role in econometrics, underlying models as diverse as simultaneous equations systems and discrete choice. In the context of time series applications, a longstanding restriction on the range of potential applications has been the availability of suitable strong laws or central limit theorems, effectively restricting attention to models with stationary or weakly dependent data. While it is well known (e.g., Wu (1981)) that consistent estimation does not rely on assumptions of stationarity or weak dependence, the development of a limit distribution theory has been hamstrung by such restrictions for a very long time.

Two examples in econometrics where these restrictions are important are GMM estimation and nonlinear cointegration. GMM limit theory was originally developed for ergodic and strictly stationary time series (Hansen (1982)) for which all measurable functions are stationary and ergodic, so that applications of strong laws and central limit theory (CLT) are straightforward. Although some attempts have been made to extend the theory to models with deterministically trending data (e.g., Wooldridge (1994), Andrews and McDermott (1995)), traditional CLT approaches have still been used and no significant progress has

[1] The authors thank a co-editor and three referees for helpful comments on earlier versions of the paper and Yoosoon Chang for many helpful discussions on the subject matter of the paper. The research reported here was begun in 1994 and the first version of the paper was completed in January 1998. Park thanks the Department of Economics at Rice University, where he is an Adjunct Professor, for its continuing hospitality, and the Cowles Foundation for support during several visits over the period 1994–1998. Park acknowledges research support from the Korea Research Foundation. Phillips thanks the NSF for support under Grant No. SBR 94-22922 and SBR 97-30295. The paper was typed by the authors in SW2.5.

been made. Nonlinear cointegrating models also seem important in a range of applications (e.g., Granger (1995)) and models with nonlinear attractor sets have been popular in economics for many years. Yet, the statistical analysis of such models with trending data has been effectively restricted to models that are linear in variables and nonlinear in parameters (Phillips (1991); Saikonnen (1995)). In fact, in such models not only is a limit theory undeveloped, but rates of convergence are also generally unknown and this inhibits the use of the traditional machinery of asymptotic analysis. Saikonnen put it this way in the conclusion of his recent paper (1995):

> "...the limiting distribution of a consistent (nonlinear) ML estimator may not be obtained in the conventional way unless something is known about the order of consistency."

Indeed, traditional methods limit our understanding somewhat further than this statement indicates, because they do not provide a basis even for the asymptotic analysis of sample moments of nonlinear functions of nonstationary data, quantities that are fundamental to our understanding of the simplest of regressions.

The purpose of the present paper is to introduce new machinery for the asymptotic analysis of nonlinear nonstationary systems. The mechanism for the asymptotic analysis of linear systems of integrated time series that was introduced in Phillips (1986, 1987) and Phillips and Durlauf (1986) relied on weak convergence in function spaces, the use of the continuous mapping theorem, and weak convergence of martingales to stochastic integrals. These methods have been in popular use ever since and play a major role in nonstationary time series analysis. However, they are unequal to the task of analyzing even simple nonlinear functionals, as the following example makes clear.

Let x_t be a standard Gaussian random walk with zero initialization, and define a process $X_n(r) = n^{-1/2}x_{[nr]}$ for $0 \leq r \leq 1$ with $[\cdot]$ denoting the greatest integer part. Then, $X_n \to_d W$, where W is standard Brownian motion. The nonlinear function $f(x) = 1/(1+x^2)$ is everywhere continuous and is well behaved at the limits of the domain of definition of x_t. What then is the limit behavior of the sample sum $\sum_{t=1}^{n} f(x_t)$? The standard approach outlined in the previous paragraph suggests the following analysis:

$$(1) \qquad \sum_{t=1}^{n} \frac{1}{1+x_t^2} = \frac{1}{n}\sum_{t=1}^{n} \frac{1}{\frac{1}{n} + \left(\frac{x_t}{\sqrt{n}}\right)^2} \sim \frac{1}{n}\sum_{t=1}^{n} \frac{1}{\left(\frac{x_t}{\sqrt{n}}\right)^2} \sim \int_0^1 \frac{dr}{X_n(r)^2}$$

$$\to_d \int_0^1 \frac{dr}{W(r)^2}.$$

However, while this approach looks convincing, it fails to deliver a useful result because the limit is undefined. Indeed, the behavior of the integral is dominated by the local behavior of the Brownian motion $W(r)$ in the vicinity of the origin and it is well known (e.g., Shorack and Wellner (1986)) that $W(r)$ satisfies a local law of the iterated logarithm at the origin, so that

$$\limsup_{r \to 0+} \frac{W(r)}{\sqrt{2r \log\log\left(\frac{1}{r}\right)}} = 1,$$

and hence

$$\int_0^{\varepsilon} \frac{1}{W(r)^2}\, dr \geq \int_0^{\varepsilon} \frac{1}{2r \log\log\left(\frac{1}{r}\right)}\, dr = \infty \qquad \text{a.s.}$$

for $\varepsilon > 0$. Thus, $\int_0^1 dr/W(r)^2 = \infty$ a.s. and all we have shown in (1) is that $\sum_{t=1}^n 1/(1+x_t^2)$ diverges as $n \to \infty$. Thus, as intimated earlier, traditional methods including those of functional weak convergence, which have become very popular in time series econometrics, fail to reveal even the order of magnitude, not to mention the limiting form, of sample moments of nonlinear functions of nonstationary series.

How then do we analyze the limit behavior of this apparently simple function? Our new approach is conceptually very easy. The key notion is to transport the sample function into a spatial function that relies on the good behavior of the function itself. In essence, we replace the sample sum by a spatial sum and treat it as a location problem in which we use the average time spent by the process in the vicinity of spatial point s, i.e., $(1/2\delta) \times \text{time}(x_t \in [s-\delta, s+\delta];\ t = 1, \ldots, n)$. Noting that $x_{[n\cdot]}$ is of stochastic order $O_p(\sqrt{n})$ we set $\delta = \sqrt{n}\,\varepsilon$ for some small $\varepsilon > 0$. The heuristic development that follows outlines the essential ideas. These are made rigorous in the rest of the paper.

We start by writing

$$\begin{aligned}
(2) \qquad \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{1}{1+x_t^2} &\sim \frac{1}{\sqrt{n}} \sum_{\min_{t \leq n}(x_t)}^{\max_{t \leq n}(x_t)} \frac{2\delta}{1+s^2} \times \frac{1}{2\delta} \\
&\qquad \times \text{time}(x_t \in [s-\delta, s+\delta];\ t = 1, \ldots, n) \\
&= \sum_{\min_{t \leq n}(x_t)}^{\max_{t \leq n}(x_t)} \frac{2\delta}{1+s^2} \times \frac{1}{n} \frac{1}{2\varepsilon} \\
&\qquad \times \text{time}\left(\frac{x_t}{\sqrt{n}} \in \left[\frac{s}{\sqrt{n}} - \varepsilon, \frac{s}{\sqrt{n}} + \varepsilon\right];\ t = 1, \ldots, n\right) \\
&= \sum_{\min_{t \leq n}(x_t)}^{\max_{t \leq n}(x_t)} \frac{2\delta}{1+s^2} \times \frac{1}{n} \frac{1}{2\varepsilon} \\
&\qquad \times \sum_{t=1}^{n} 1\left\{\frac{x_t}{\sqrt{n}} \in \left[\frac{s}{\sqrt{n}} - \varepsilon, \frac{s}{\sqrt{n}} + \varepsilon\right]\right\},
\end{aligned}$$

where here and elsewhere in the paper we denote the indicator function by $1\{\cdot\}$. The essential simplification that is involved in the transition to the form (2) is that it converts a nonlinear function of x_t into a formulation that involves x_t linearly, inside the indicator function, so that it can be readily standardized by $\sqrt{n}$.

Now, as $n \to \infty$, we note that $\max_{t \le n}(x_t) \to \infty$, $\min_{t \le n}(x_t) \to -\infty$, so that for large n and small δ we have the spatial approximation

$$\sum_{\min_{t \le n}(x_t)}^{\max_{t \le n}(x_t)} \frac{2\delta}{1+s^2} \sim \int_{-\infty}^{\infty} \frac{ds}{1+s^2}.$$

Also for all finite s we have

$$\begin{aligned} &\frac{1}{n}\frac{1}{2\varepsilon} \times \sum_{t=1}^{n} 1\left\{\frac{x_t}{\sqrt{n}} \in \left[\frac{s}{\sqrt{n}} - \varepsilon, \frac{s}{\sqrt{n}} + \varepsilon\right]\right\} \\ &\quad \sim \frac{1}{n}\frac{1}{2\varepsilon} \times \sum_{t=1}^{n} 1\left\{\frac{x_t}{\sqrt{n}} \in [-\varepsilon, \varepsilon]\right\} \\ &\quad \sim \frac{1}{2\varepsilon}\int_0^1 1\{|X_n(r)| \le \varepsilon\}\, dr \\ &\quad \sim \frac{1}{2\varepsilon}\int_0^1 1\{|W(r)| \le \varepsilon\}\, dr. \end{aligned}$$

From these heuristics we get the approximate expression

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{1}{1+x_t^2} \sim \left(\int_{-\infty}^{\infty} \frac{ds}{1+s^2}\right)\left(\frac{1}{2\varepsilon}\int_0^1 1(|W(r)| \le \varepsilon)\, dr\right), \tag{3}$$

which is given in terms of the product of a spatial integral and a functional of the limiting Brownian motion process. Note that the resulting formula is free of the sample size, so that the order of the magnitude of the sample function is now properly determined, as distinct from (1).

The final step in these heuristics is to simplify (3). Noting that ε was arbitrary, we can let $\varepsilon \to 0$ in (3). In fact, the final expression has a natural limit as $\varepsilon \to 0$ that measures the spatial density of Brownian motion over the time interval $[0,1]$. Specifically, the limit

$$L_W(1,0) = \lim_{\varepsilon \to 0+} \frac{1}{2\varepsilon}\int_0^1 1(|W(r)| \le \varepsilon)\, dr \tag{4}$$

is well defined and is known as the local time of standard Brownian motion at the origin (cf. equation (15) below). It is analogous to a probability density, but is a random process rather than a deterministic function. Local time is a very useful process associated with Brownian motion and it will be used extensively in the development of our theory, so more exposition and discussion of its properties is provided in Section 2 of the paper. For the moment, we are content to note that using (3) and (4), our heuristics lead us to the following asymptotic behavior as $n \to \infty$:

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{1}{1+x_t^2} \to_d \left(\int_{-\infty}^{\infty} \frac{ds}{1+s^2} \right) L_W(1,0). \tag{5}$$

Clearly, this limit expression is very different from the usual limit formulae for sample moments of linear functionals of integrated processes, yet it is simple and neat. Obviously, the heuristic argument that leads to (5) could have been used to obtain the limit behavior of the sample mean of an arbitrary integrable function f, specifically

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} f(x_t) \to_d \left(\int_{-\infty}^{\infty} f(s)\, ds \right) L_W(1,0), \tag{6}$$

a formula that we will derive rigorously in an extended form in Theorem 3.2 of Section 3.

In addition to studying the limit behavior of sample means like (6), a theory of regression also requires that we be able to analyze sample covariances between nonstationary and stationary random elements. In the simple case of linear regression, this amounts to studying the sample covariance between an integrated process and a stationary process. It was shown in Phillips (1987) and Phillips and Durlauf (1986) that this can be accomplished by directly establishing weak convergence of the sample moment to a stochastic integral. Unfortunately, the sample covariance of a nonlinear nonstationary random process and a stationary process is as resistant to analysis by these methods as the sample mean function considered above. To illustrate, consider the sample covariance $\sum_{t=1}^{n} f(x_t)u_t$ between the integrable nonlinear function $f(x_t) = 1/(1+x_t^2)$ and an iid(0,1) process, u_t. To fix ideas, we take x_t to be $\mathscr{F}_{t-1}$-measurable and u_t to be $\mathscr{F}_t$-measurable with respect to some filtration $\mathscr{F} = \{\mathscr{F}_t\}$, and suppose $Y_n(r) = n^{-\frac{1}{2}}\sum_{t=1}^{[nr]} u_t \to_d U(r)$, a standard Brownian motion. If we let $v_t = \Delta x_t$ and allow $\sigma_{uv} = E(u_t v_{t+1}) \neq 0$, then U and W are correlated Brownian motions with covariance σ_{uv}. If the conventional approach were attempted, we would be tempted to write

$$\sqrt{n} \sum_{t=1}^{n} \frac{1}{1+x_t^2} u_t = \sqrt{n} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{1}{\frac{1}{n} + \left(\frac{x_t}{\sqrt{n}} \right)^2} \frac{u_t}{\sqrt{n}} \right) \tag{7}$$

$$\sim \sum_{t=1}^{n} \frac{1}{\left(\frac{x_t}{\sqrt{n}} \right)^2} \frac{u_t}{\sqrt{n}} \sim \int_0^1 \frac{dY_n(r)}{X_n(r)^2} \sim_d \int_0^1 \frac{dU(r)}{W(r)^2},$$

which is again undefined because the integrand $W(r)^{-2}$ is not square integrable. Thus, direct analysis fails, as before, revealing only that the sample covariance (7) diverges.

Taking an alternate approach, we note that $\Sigma_{t=1}^{n} f(x_t)u_t$ is a martingale whose conditional variance is the sample mean function $\Sigma_{t=1}^{n} f(x_t)^2$, which may be treated in the same spatial manner as (6) above, leading to the following limit:

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} f(x_t)^2 \to_d \left(\int_{-\infty}^{\infty} f(s)^2 \, ds \right) L_W(1,0). \tag{8}$$

This result indicates that $n^{-\frac{1}{4}}$ scaling is appropriate for the martingale $\Sigma_{t=1}^{n} f(x_t)u_t$. Further, observe that the conditional covariance between the martingales $\Sigma_{t=1}^{n} f(x_t)u_t$ and $x_n = \Sigma_{t=1}^{n} v_t$ is $\Sigma_{t=1}^{n-1} f(x_t)\sigma_{uv}$, which, for the scaled quantities $n^{-\frac{1}{4}}\Sigma_{t=1}^{n} f(x_t)u_t$ and $n^{-\frac{1}{2}}x_n$, produces

$$\frac{1}{n^{\frac{3}{4}}} \sum_{t=1}^{n-1} f(x_t)\sigma_{uv} = O_p\left(\frac{1}{n^{\frac{1}{4}}} \right) \to_p 0, \tag{9}$$

indicating that the two scaled martingales have uncorrelated limits. A technical embedding argument then allows us to apply the martingale central limit theorem to $n^{-\frac{1}{4}}\Sigma_{t=1}^{n} f(x_t)u_t$, showing that this standardized sum converges to a scaled normal variate of the form

$$\left[\left(\int_{-\infty}^{\infty} f(s)^2 \, ds \right) L_W(1,0) \right]^{\frac{1}{2}} \xi, \qquad \xi =_d N(0,1),$$

i.e., a mixed normal limiting distribution, with a mixing variate that is given by the limit of the conditional variance (8). It is the $\sqrt{n}$ convergence rate of this conditional variance that determines the $n^{\frac{1}{4}}$ convergence factor in the sample covariance $n^{-\frac{1}{4}}\Sigma_{t=1}^{n} f(x_t)u_t$. Interestingly, in the present example, the limit Brownian motions U and W (arising from the two shocks u_t and v_t) are correlated, but the mixing variate (8) (whose randomness depends on W through the local time L_W) is independent of ξ. This independence arises because of the zero conditional covariance (9) between the limiting martingales. It is, in fact, the presence of the integrable nonlinear function $f(x_t)$ in the martingale $\Sigma_{t=1}^{n} f(x_t)u_t$ that secures this independence. In effect, the full sample path of x_t is no longer needed in the limit theory, unlike the linear case which inevitably relies on functionals of the limiting trajectory, $W(r)$, of $n^{-1/2}x_{[nr]}$, a trajectory that is correlated with $U(r)$ when $\sigma_{uv} \neq 0$. Indeed, the integrable nonlinear function $f(x_t)$ influences the martingale $\Sigma_{t=1}^{n} f(x_t)u_t$ in such a way that only contributions from x_t in the neighborhood of the origin are important in the limit (as is apparent from (8), the conditional variance limit). Independence is then a consequence of the fact that u_t is orthogonal to x_t for all t, including of course, those t that most matter in the sum $\Sigma_{t=1}^{n} f(x_t)u_t$. This heuristic argument reveals the major role that nonlinearity can play in influencing the asymptotics of nonstationary series. It also reveals the importance of the requirement that the innovation, u_t, be a martingale difference.

With these heuristics behind us, the plan of the rest of the paper is as follows. Section 2 outlines the model we will be using, the assumptions needed, and gives some preliminary discussion of Brownian local time and some of its properties that we utilize in our development. Section 3 provides an asymptotic theory for certain families of nonlinear functions. This section builds on some earlier work by the authors in Park and Phillips (1999)—hereafter, P^2—and makes rigorous the heuristic ideas described above. Consistency in nonlinear regression is proved in Section 4 and the limit distribution theory is developed in Section 5. Section 6 concludes. A technical Appendix is provided and is divided into two parts. Some useful technical lemmas are given in Section 7 and proofs of the theorems in the paper are given in Section 8.

A final word in this introduction about notation. For a vector $x=(x_i)$ or a matrix $A=(a_{ij})$, the modulus $|\cdot|$ is taken element by element. Therefore, $|x|=(|x_i|)$ and $|A|=(|a_{ij}|)$. The maximum of the moduli is denoted by $\|\cdot\|$, i.e., $\|A\|=\max_{i,j}|a_{ij}|$ and $\|x\|=\max_i|x_i|$. The notation $\|\cdot\|$ is also used to denote the supremum of a function. For a function f, which can be vector- or matrix-valued, $\|\cdot\|_K$ signifies the supremum norm over a subset K of its domain, so that $\|f\|_K=\sup_{x\in K}\|f(x)\|$. The subset K, over which the supremum is taken, will not be specified if it is clear from the context. Standard terminologies and notations in probability and measure theory are used throughout the paper. In particular, notations for various notions of convergence such as $\to_{a.s.}$, $\to_p$ and $\to_d$ frequently appear. The notation $=_d$ signifies equality in distribution. Finally, we denote by $\mathbf{R}_+$ the set of positive real numbers.

2. THE MODEL AND PRELIMINARY RESULTS

We consider the nonlinear regression model for y_t given by

$$y_t = f(x_t, \theta_0) + u_t \tag{10}$$

where $f:\mathbf{R}\times\mathbf{R}^m\to\mathbf{R}$ is known, x_t and u_t are the regressors and regression errors, respectively, and θ_0 is an m-dimensional true parameter vector that lies in the parameter set Θ. In model (10), we let x_t be an integrated process and u_t be a martingale difference sequence, as will be specified more precisely subsequently. The model is thus critically different from the standard nonlinear regression with stationary regressors. It can be viewed as a nonlinear cointegrating regression.

This paper concentrates on the estimation of (10) by nonlinear least squares (NLS). The approach given here is applicable to other procedures, but to avoid unnecessary complications in this initial development of the theory we will keep the focus on NLS. If we let

$$Q_n(\theta) = \sum_{t=1}^{n} (y_t - f(x_t, \theta))^2,$$

then the NLS estimator $\hat{\theta}_n$ is defined as the minimizer of $Q_n(\theta)$ over $\theta \in \Theta$, i.e.,

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\arg\min}\, Q_n(\theta). \tag{11}$$

Accordingly, an error variance estimate is given by $\hat{\sigma}_n^2 = (1/n)\sum_{t=1}^n \hat{u}_t^2$, where $\hat{u}_t = y_t - f(x_t, \hat{\theta}_n)$. It is assumed that $\hat{\theta}_n$ exists and is unique for all n. Moreover, we assume throughout the paper that Θ is compact and convex, and θ_0 is an interior point of Θ. This is a standard assumption for stationary nonlinear regression.

Write x_t more specifically as

$$x_t = x_{t-1} + v_t.$$

The initial value x_0 of x_t may be any $O_p(1)$ random variable. However, we set $x_0 = 0$ in the paper to avoid unnecessary complications in exposition. In the $O_p(1)$ case, the initialization does not affect the asymptotic results anyway, as is evident from P^2. When the initialization is in the distant past and $x_0 = O_p(n^{1/2})$, the initial condition does affect the asymptotic theory (e.g., see Phillips and Park (1998)) and appropriate adjustments to some of our formulae will be required in this event, but will be fairly obvious. To focus on essentials in our development of nonlinear regression, we will retain the simplification $x_0 = 0$.

For the time series u_t and v_t, respectively, we define the stochastic processes U_n and V_n on $[0,1]$ by

$$U_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t \qquad \text{and} \qquad V_n(r) = \frac{1}{\sqrt{n}} \sum_{t=0}^{[nr]} v_{t+1}$$

where $[s]$ denotes the largest integer not exceeding s.

ASSUMPTION 2.1: (a) $(U_n, V_n) \to_d (U, V)$ *as* $n \to \infty$, *where* (U, V) *is a vector Brownian motion. Moreover, assume for each* n, *there exists a filtration* $(\mathscr{F}_{nt})$, $t = 0, \ldots, n$, *such that*:

(b) $(u_t, \mathscr{F}_{nt})$ *is a martingale difference sequence with* $\mathbf{E}(u_t^2 \,|\, \mathscr{F}_{n,t-1}) = \sigma^2$ *a.s. for all* $t = 1, \ldots, n$, *and* $\sup_{1 \le t \le n} \mathbf{E}(|u_t|^q F_{n,t-1}) < \infty$ *a.s. for some* $q > 2$, *and*

(c) x_t *is adapted to* $\mathscr{F}_{n,t-1}$, $t = 1, \ldots, n$.

Assumption 2.1 is satisfied for a wide variety of data generating processes. Condition (a) is the usual assumption routinely imposed to analyze linear models with integrated time series, as in our earlier work, Park and Phillips (1988). It is known to hold for mildly heterogeneous time series, as well as stationary processes. The martingale difference assumption for the regression errors in (b) is standard in much stationary time series regression and it is used for our development of nonstationary regression theory here, although it is not essential in some cases. As is now well known, serial correlation in the errors and cross correlation between the errors and regressors is allowed in linear cointegrating regression theory. The correlations do not affect, for instance, the consistency of the least squares estimator, but generally do affect the limit distribution theory. Since our model includes the linear cointegrating regression as a special case, it is therefore reasonable to expect that some of our subsequent results apply without condition (b), perhaps with some modification. It will be pointed out when this is the case. Moreover, as the heuristic discussion of the examples in the introduction clarify, condition (b) does play an important role in the limit distribution theory. Under condition (c), x_t becomes predetermined. The condition can simply be met by choosing the natural filtration of (u_t, x_{t+1}) for $(\mathscr{F}_{nt})$. Note that conditions (b) and (c) together imply, in particular, that $\mathbf{E}(y_t \mid \mathscr{F}_{n,t-1}) = f(x_t, \theta_0)$ a.s.

The stochastic process (U_n, V_n) takes values in $D[0,1]^2$, where $D[0,1]$ denotes the space of cadlag functions defined on the unit interval $[0,1]$. The space $D[0,1]$ is usually equipped with the Skorohod toplogy. However, it is more convenient in our context to topologize it with the uniform topology, and interpret $(U_n, V_n) \to_d (U, V)$ in Assumption 2.1(a) as weak convergence in $D[0,1]$ with the supremum norm. The reader is referred to Billingsley (1968) for detailed discussion on the subject. It then follows from the so-called Skorohod representation theorem that there is a common probability space $(\Omega, \mathscr{F}, \mathbf{P})$ supporting (U_n^0, V_n^0) and (U, V) such that

$$(U_n, V_n) =_d (U_n^0, V_n^0) \qquad \text{and} \qquad (U_n^0, V_n^0) \to_{a.s.} (U, V) \tag{12}$$

in $D[0,1]^2$ with the uniform topology. Moreover, we have the following strong approximation.

LEMMA 2.1: *Let Assumptions* 2.1(b) *and* (c) *hold. Then we may represent* U_n^0 *introduced in* (12) *by*

$$U_n^0\left(\frac{t}{n}\right) = U\left(\frac{\tau_{nt}}{n}\right)$$

with an increasing sequence of stopping times τ_{nt} *in* $(\Omega, \mathscr{F}, \mathbf{P})$ *with* $\tau_{n0} = 0$ *such that*

$$\sup_{1 \le t \le n} \left| \frac{\tau_{nt} - t}{n^{\delta}} \right| \to_{a.s.} 0 \tag{13}$$

as $n \to \infty$ *for any* $\delta > \max(1/2, 2/q)$ *where* q *is the moment exponent given in Assumption* 2.1(b).

In the paper, we establish the weak consistency and derive the asymptotic distribution of the NLS estimator $\hat{\theta}_n$ defined in (11). For our purposes, it therefore causes no loss in generality to assume $(U_n, V_n) = (U_n^0, V_n^0)$, instead of $(U_n, V_n) =_d (U_n^0, V_n^0)$ as in (12). This convention will be made throughout the paper. It allows us to avoid repetitious embedding of (U_n, V_n) in the probability space $(\Omega, \mathscr{F}, \mathbf{P})$, where (U_n^0, V_n^0) is defined. Due to the convention introduced above, all the subsequent convergence results of the sample moments with $\to_{a.s.}$ and $\to_p$, as well as those with $\to_d$, should generally be interpreted as the corresponding ones with $\to_d$. If, however, the convergence is to a nonrandom limit, then we may as well interpret it as $\to_p$, since $\to_d$ and $\to_p$ are identical in such a case.

Stronger assumptions on the data generating process for x_t will often be required to fully develop the asymptotics for the nonlinear regressions. We now introduce the following assumption.

ASSUMPTION 2.2: *Let* (a), (b), *and* (c) *be given as in Assumption* 2.1. *We let*:
(d) $v_t = \varphi(L)\varepsilon_t = \sum_{k=0}^{\infty} \varphi_k \varepsilon_{t-k}$, *with* $\varphi(1) \neq 0$ *and* $\sum_{k=0}^{\infty} k|\varphi_k| < \infty$, *and assume that* $\{\varepsilon_t\}$ *is a sequence of i.i.d. random variables with mean zero and* $\mathbf{E}|\varepsilon_t|^p < \infty$ *for some* $p > 4$, *the distribution of which is absolutely continuous with respect to Lebesgue measure and has characteristic function* $c(\lambda)$ *satisfying* $\lim_{\lambda \to \infty} \lambda^r c(\lambda) = 0$ *for some* $r > 0$.

Assumption 2.2 introduces more restrictive conditions on x_t, but still permits a wide variety of models that are used in practical applications, including all invertible Gaussian ARMA models. Under condition (d), it follows that $V_n^0 \to_d V$, as shown in Phillips and Solo (1992).

The asymptotic theory for nonlinear functions of integrated time series heavily relies on the local time of Brownian motion, or more generally that of a continuous semimartingale. The background needed is discussed in P^2 and Phillips and Park (1998) and a full discussion is contained in Revuz and Yor (1994), to which the reader is referred. In brief, for a continuous semimartingale M with quadratic variation process $[M]$, the local time of M is defined to be a two parameter stochastic process $L_M(t, s)$, which satisfies the following important lemma.

LEMMA 2.2 (The Occupation Time Formula): *Let T be a nonnegative transformation on* $\mathbf{R}$. *Then*

$$\int_0^t T(M(r))\, d[M]_r = \int_{-\infty}^{\infty} T(s) L_M(t, s)\, ds \tag{14}$$

for all $t \in \mathbb{R}$.

The local time, as a function of its spatial parameter s, has the interpretation as an occupation density. In formula (14), the occupation time is defined with respect to the measure $d[M]_r$, which may be regarded as the natural time-scale for M in terms of its variation. It is known that the local time $L_M(t, s)$ of a continuous semimartingale M is a.s. continuous in t and cadlag in s. Due, in particular, to the right continuity of $L_M(t, \cdot)$, we may apply (14) with $T(x) =$

$1\{s \le x < s+\varepsilon\}$ to get

$$(15) \qquad L_M(t,s) = \lim_{\varepsilon \to 0} \int_0^t 1\{s \le M(r) < s+\varepsilon\}\, d[M]_r,$$

a representation that explains why $L_M(\cdot, s)$ is called the local time of M at s. Roughly speaking, $L_M(t,s)$ measures the time, in the time-scale given by $[M]$, that is spent by M in the vicinity of s over the interval $[0,t]$.

The formula (14), of course, applies to Brownian motion as a special case. For Brownian motion, the result in (14) is known to hold for any locally integrable transformation T. See, for instance, Chung and Williams (1990, Corollary 7.4). It also holds for other diffusion processes such as the Ornstein-Uhlenbeck process, which has been used for the asymptotic analyses of models with near-integrated processes as in Phillips (1987). For the development of our subsequent theory, we will frequently refer to the local time L_V of the Brownian motion V. For notational simplicity, we will in fact use a scaled local time L of V defined by

$$(16) \qquad L(t,s) = (1/\sigma_v^2) L_V(t,s)$$

where σ_v^2 is the variance of V. It is often much more convenient to present our results in terms of L, instead of L_V. If we apply the formula (14) to V, then we have for any locally integrable T

$$\int_0^t T(V(r))\, dr = (1/\sigma_v^2) \int_{-\infty}^{\infty} T(s) L_V(t,s)\, ds = \int_{-\infty}^{\infty} T(s) L(t,s)\, ds$$

since $d[V]_r = \sigma_v^2\, dr$. The scaled local time $L(t,s)$ can therefore be regarded as the actual time spent by V up to time t in the neighborhood of s. It is called chronological local time in Phillips and Park (1998).

All our subsequent results are presented in terms of the Brownian motions U and V introduced in Assumption 2.1, the covariance of which will be denoted by σ_{uv}. The variances of U and V are, as already specified, written as σ^2 and σ_v^2, respectively. The scaled local time L of V defined in (16) will also be used without further reference. Finally, some of our theoretical results involve another vector Brownian motion W. The process W is independent of V, and therefore of L, and has variance $\sigma^2 I$. Of course, we may, and do, assume that W is defined in the common probability space $(\Omega, \mathscr{F}, \mathbf{P})$ containing the processes U and V. These conventions will be used throughout the paper.

3. ASYMPTOTICS FOR LINEAR FUNCTIONS OF INTEGRATED PROCESSES

3.1. *Regular Functions*

We now present some preliminary results for nonlinear transformations of integrated time series, which are used in our subsequent development of the asymptotic theory of nonlinear regression. These are related to some earlier concepts introduced in P^2. We start with the concept of a regular transformation.

DEFINITION 3.1: A transformation T on $\mathbf{R}$ is said to be *regular* if and only if

(a) it is continuous in a neighborhood of infinity, and
(b) for any compact subset K of $\mathbf{R}$ given, there exist for each $\varepsilon > 0$ continuous functions $\underline{T}_\varepsilon$, $\overline{T}_\varepsilon$, and $\delta_\varepsilon > 0$ such that $\underline{T}_\varepsilon(x) \leq T(y) \leq \overline{T}_\varepsilon(x)$ for all $|x-y| < \delta_\varepsilon$ on K, and such that $\int_K (\overline{T}_\varepsilon - \underline{T}_\varepsilon)(x)\,dx \to 0$ as $\varepsilon \to 0$.

The regularity conditions in Definition 3.1 are somewhat stronger than those in P^2. However, they are satisfied by most functions used in practical nonlinear time series analyses. The class of regular transformations is closed under the usual operations of addition, subtraction, and multiplication, as we show in Lemma A1. Continuous functions are, of course, regular. This can easily be seen by setting, for any continuous function T on $\mathbf{R}$, $\underline{T}_\varepsilon(x) = T(x) - \varepsilon$ and $\overline{T}_\varepsilon(x) = T(x) + \varepsilon$ with the usual δ_ε for the $\varepsilon - \delta$ formulation of uniform continuity, and these functions apply for any compact subset of $\mathbf{R}$. It is also quite clear that any continuous function on a compact support is regular. All piecewise continuous functions are therefore regular, due to Lemma A1 in the Appendix. Naturally, we call a vector- or matrix-valued function regular when each of its components is regular.

Logarithmic functions and reciprocals are not regular. Therefore, our subsequent theory for regular functions is not directly applicable to these functions. However, for such functions (T, say), we may consider $T_\varepsilon(x) = T(x)1\{|x| > \varepsilon\} + T(\varepsilon)1\{|x| \leq \varepsilon\}$ for some small $\varepsilon > 0$ instead of T itself. Note that T and T_ε are identical over any finite set of nonzero points, if we take $\varepsilon > 0$ to be smaller than the minimum of their moduli. Therefore, if x_t is driven by an error process whose distribution is of the continuous type, then T and T_ε are practically indistinguishable in finite samples. In this case, we have $T(x_t) = T_\varepsilon(x_t)$ for all $t = 1, \ldots, n$ if $\varepsilon > 0$ is sufficiently small. Of course, we can make this approach more rigorous by letting ε be n-dependent, say ε_n, such that $\varepsilon_n \to 0$, and considering the asymptotics of $T_n = T_{\varepsilon_n}$. This is done in P^2. We assume throughout the paper that these conventions are made for logarithmic functions and reciprocal functions.

Extending the theory in P^2, we now consider families of functions indexed by some parameter, rather than individual functions. This extension is needed for the analysis of nonlinear regressions. In the subsequent development of our theory, we are mainly concerned with a family $F: \mathbf{R} \times \Pi \to \mathbf{R}^m$ of functions from $\mathbf{R}$ to $\mathbf{R}^m$ with index set Π. Below we introduce a *regular* family of functions, which is fundamental to our analysis. We have already defined the terminology regular in Definition 3.1 for individual functions, and here it is extended to a family of functions. The asymptotics for these families then follow. In particular, we present limiting results for the sample functions $n^{-1}\sum_{t=1}^n F(x_t/\sqrt{n}, \pi)$ and $n^{-1/2}\sum_{t=1}^n F(x_t/\sqrt{n}, \pi)u_t$ for regular F. Following our terminology in the Introduction, these will be referred to subsequently as *sample mean* and *sample covariance asymptotics* for F.

DEFINITION 3.2: We say that F is *regular* on Π if

(a) $F(\cdot, \pi)$ is regular for all $\pi \in \Pi$, and
(b) for all $x \in \mathbf{R}$, $F(x, \cdot)$ is equicontinuous in a neighborhood of x.

Conditions (a) and (b) in Definition 3.2 will be called *regularity conditions*. Lemma A2 shows that regularity condition (a) is sufficient to guarantee that both sample mean and sample covariance asymptotics for $F(\cdot, \pi)$ are well defined for each $\pi \in \Pi$. Equicontinuity of $F(x, \cdot)$ in regularity condition (b) ensures, as shown in Lemma A3 in the Appendix, the existence of a neighborhood N_0 of any given $\pi_0 \in \Pi$ for which $\sup_{\pi \in N_0} F(\cdot, \pi)$ and $\inf_{\pi \in N_0} F(\cdot, \pi)$ are regular. This is required for uniform convergence in sample mean asymptotics. The condition is, of course, automatically satisfied if Π is a singleton set.

THEOREM 3.1: *Let Assumption* 2.1 *hold. If F is regular on a compact set* Π, *then as* $n \to \infty$

$$\frac{1}{n} \sum_{t=1}^{n} F\left(\frac{x_t}{\sqrt{n}}, \pi \right) \to_{a.s.} \int_0^1 F(V(r), \pi)\, dr$$

uniformly in $\pi \in \Pi$. *Moreover, if* $F(\cdot, \pi)$ *is regular, then*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} F\left(\frac{x_t}{\sqrt{n}}, \pi \right) u_t \to_d \int_0^1 F(V(r), \pi)\, dU(r)$$

as $n \to \infty$.

For regular F, Theorem 3.1 shows that sample mean asymptotics involve a random mean functional of F, specifically the time average of a nonlinear function of the Brownian motion V. The sample covariance asymptotics involve a stochastic integral of F, which will generally have a nonzero mean except for the special case where $\sigma_{uv} = 0$, i.e., U and V are independent. This stochastic integral also has a non-Gaussian distribution, although in the special case where $\sigma_{uv} = 0$ it will be a variance mixture of Gaussian distributions. For a family of homogeneous functions (like polynomials), we may also easily apply Theorem 3.1 to get asymptotics for moments of unnormalized functions of integrated time series. If specialized to the linear or quadratic functions and to a singleton parameter set Π, the results in Theorem 3.1 are already well known. Comparable results in this case have been obtained earlier by several authors under conditions weaker than Assumption 2.1. The reader is referred to Phillips and Solo (1992) and Hansen (1992), and the references cited there.

3.2. *Function Classes and Asymptotics for Unnormalized Integrated Time Series*

The limit behavior of sample moments of functions of unnormalized integrated time series critically depends on the type of function involved, as shown in P^2. P^2 consider three classes of functions—integrable functions (I), asymptotically homogeneous functions (H), and explosive functions (E). The first class includes all integrable transformations. The second class comprises functions that behave asymptotically like homogeneous functions (including homogeneous functions as a special case). This class also includes transformations such as $T(x) = \log|x|, e^x/(1+e^x)$, arctan x, and $T(x) = |x|^k$. The third class is for functions that increase at an exponential rate, and transformations like $T(x) = e^x$ or $|x|^k e^x$ belong to this class. The three classes of functions are mutually exclusive and have no function in common except for the zero function.

Here we consider the first two families of functions studied in P[2]. Introduced below are regularity conditions for the two (I- and H-) families of functions. Each family is presented with its asymptotics. Subsequently, we give asymptotic results for the sample moments $\sum_{t=1}^n F(x_t, \pi)$ and $\sum_{t=1}^n F(x_t, \pi)u_t$, appropriately normalized. As before, we refer to these as sample mean and sample covariance asymptotics for F.

3.2(a). *Integrable Functions*

DEFINITION 3.3: We say that F is *I-regular* on Π if

(a) for each $\pi_0 \in \Pi$, there exist a neighborhood N_0 of π_0 and $T: \mathbf{R} \to \mathbf{R}$ bounded integrable such that $\|F(x,\pi) - F(x,\pi_0)\| \le \|\pi - \pi_0\| T(x)$ for all $\pi \in N_0$, and

(b) for some constants $c > 0$ and $k > 6/(p-2)$ with $p > 4$ given in Assumption 2.2, $\|F(x,\pi) - F(y,\pi)\| \le c|x-y|^k$ for all $\pi \in \Pi$, on each piece S_i of their common support $S = \bigcup_{i=1}^m S_i \subset \mathbf{R}$.

We call the conditions in Definition 3.3 *I-regularity conditions*. Condition (a) requires that $F(x,\cdot)$ be continuous on Π for all $x \in \mathbf{R}$, as in standard nonlinear regression theory. The condition holds, for instance, if $\sup_{\pi \in \Pi} \partial F(\cdot,\pi)/\partial\pi$ is bounded and integrable, and implies that $\sup_{\pi \in \Pi}|F(\cdot,\pi)|$ is bounded and integrable, if Π is compact. When Π consists only of a single point π, the boundedness and integrability of $F(\cdot,\pi)$ is sufficient for the condition to hold. Condition (b) requires that all functions in the family should be sufficiently smooth piecewise on their common support, which is independent of π. The condition allows for functions that are progressively less smooth as the underlying process has higher moments.

THEOREM 3.2: *Let Assumption* 2.2 *hold. If F is I-regular on a compact set* Π, *then as* $n \to \infty$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n F(x_t,\pi) \to_p \left(\int_{-\infty}^{\infty} F(s,\pi)\, ds \right) L(1,0)$$

uniformly in $\pi \in \Pi$. *Moreover, if* $F(\cdot, \pi)$ *is I-regular,*

$$\frac{1}{\sqrt[4]{n}} \sum_{t=1}^{n} F(x_t, \pi) u_t \to_d \left(L(1,0) \int_{-\infty}^{\infty} F(s, \pi) F(s, \pi)' \, ds \right)^{1/2} W(1)$$

as $n \to \infty$.

Both in sample mean and sample covariance asymptotics, the convergence rates for functions of integrated time series are an order of magnitude slower than they are for stationary time series. Roughly speaking, this reduction in convergence rate occurs because observations from an integrated time series diverge in probability—at a rate of $\sqrt{n}$ for the sample of size n. Any observation, unless it is realized in a neighborhood of the origin, therefore loses its impact asymptotically if it is transformed by a function which vanishes at infinity as is the case with an integrable function. The asymptotics for I-regular F involve the local time L of the limit Brownian motion V. Note that both the sample mean and sample covariance asymptotics depend upon L only through its value at the spatial parameter zero. Therefore, only the time that V spends in the neighborhood of the origin matters for the asymptotics of I-regular functions. The sample covariance asymptotics yield a limit distribution that is a normal mixture with a mixing variate given by L. Note that W is independent of V, and therefore of L.

3.2(b). *Asymptotically Homogeneous Functions*

For our asymptotic analysis, the class of locally bounded transformations on **R** will play an important role and we denote this class by $\mathscr{T}_{LB}$. Any regular transformation T on **R**, defined in Definition 3.1, belongs to $\mathscr{T}_{LB}$. We often need to consider a subclass $\mathscr{T}_{LB}^0$ of $\mathscr{T}_{LB}$ consisting only of locally bounded transformations that are exponentially bounded, i.e., transformations T such that $T(x) = O(e^{c|x|})$ as $|x| \to \infty$ for some $c \in \mathbf{R}_+$. Also introduced are the class $\mathscr{T}_B$ of bounded transformations on **R**, and its subclass $\mathscr{T}_B^0$ including all transformations that are bounded and vanish at infinity, i.e., transformations T such that $T(x) \to 0$ as $|x| \to \infty$. As shown in Lemma A3 in the Appendix, these are the classes of transformations on **R** for which sample mean and sample covariance asymptotics can effectively be bounded. Clearly, $\mathscr{T}_B^0 \subset \mathscr{T}_B \subset \mathscr{T}_{LB}^0 \subset \mathscr{T}_{LB}$. As for regular functions, vector- and matrix-valued functions belong to a given class of functions when all of the individual components belong to the class.

Denote by Φ the set of parameters, and by $\mathbf{R}_+^{m^2}$ the set of $m \times m$ matrices of positive numbers. Let $\mathbf{z}: \mathbf{R}_+ \times \Phi \to \mathbf{R}_+^{m^2}$ be nonsingular, and $a, b: \mathbf{R}_+ \times \Phi \to \mathbf{R}_+^{m^2}$. We say that $a = o(z)$ and $b = O(z)$ on Φ if, as $\lambda \to \infty$,

$$\|z(\lambda, \omega)^{-1} a(\lambda, \omega)\| \to 0 \qquad \text{and} \qquad \|z(\lambda, \omega)^{-1} b(\lambda, \omega)\| < \infty$$

uniformly in $\omega \in \Phi$. Define $Z: \mathbf{R} \times \mathbf{R}_+ \times \Phi \to \mathbf{R}^m$. The following definition determines the asymptotic order of a family of functions $Z(\cdot, \lambda, \omega)$, parameterized by $\omega \in \Phi$, in terms of $z(\lambda, \omega)$ for large λ.

DEFINITION 3.4: We say that Z is *of order smaller than* z on Φ if

$$Z(x, \lambda, \omega) = a(\lambda, \omega) A(x, \omega) \qquad \text{or} \qquad b(\lambda, \omega) A(x, \omega) B(\lambda x, \omega)$$

where $a = o(z)$ and $b = O(z)$ on Φ, $\sup_{\omega \in \Phi} A(\cdot, \omega) \in \mathscr{T}_{LB}^0$ and $\sup_{\omega \in \Phi} B(\cdot, \omega) \in \mathscr{T}_B^0$.

With the notion introduced in Definition 3.4, we may now be precise about the family of asymptotically homogeneous functions that we will consider.

DEFINITION 3.5: 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 is regular on Π, and
(b) $R(x, \lambda, \pi)$ is of order smaller than $\kappa(\lambda, \pi)$ for all $\pi \in \Pi$.

We call κ the *asymptotic order* and H the *limit homogeneous function* of F. If κ does not depend upon π, then F is said to be H_0-regular.

The conditions in Definition 3.5 will be referred to as *H-regularity conditions* in our subsequent discussions. Roughly speaking, the class of H-regular functions consists of functions that are asymptotically equivalent to some regular homogeneous functions, which we call their limit homogeneous functions. Condition (b) allows us to establish this asymptotic equivalence. The limit homogeneous function H introduced in (a) is defined uniquely due to the negligibility condition for the residual term in (b). The regularity requirement for the limit homogeneous function H in the condition (a) is necessary to ensure that H has well defined asymptotics.

THEOREM 3.3: *Let Assumption* 2.1 *hold, and let F be specified as in Definition* 3.5. *If F is H-regular on a compact set* Π, *then as* $n \to \infty$

$$\frac{1}{n} \kappa(\sqrt{n}, \pi)^{-1} \sum_{t=1}^{n} F(x_t, \pi) \to_{a.s.} \int_0^1 H(V(r), \pi)\, dr$$

uniformly in $\pi \in \Pi$. *Moreover, if* $F(\cdot, \pi)$ *is H-regular, then*

$$\frac{1}{\sqrt{n}} \kappa(\sqrt{n}, \pi)^{-1} \sum_{t=1}^{n} F(x_t, \pi) u_t \to_d \int_0^1 H(V(r), \pi)\, dU(r)$$

as $n \to \infty$.

The limit theory for H-regular functions of unnormalized integrated time series are essentially identical to those of regular transformations with normalized time series, as given in Theorem 3.1. This is because

$$F(x_t, \pi) \approx \kappa(\sqrt{n}, \pi) H\left(\frac{x_t}{\sqrt{n}}, \pi\right)$$

and the residual is negligible in the limit. Notice that we may write the limit for sample mean asymptotics as $\int_{-\infty}^{\infty} H(s, \pi) L(1, s)\, ds$ using the occupation times formula. This expression is analogous in form to that of the sample mean limit of I-regular functions. For I-regular functions, only the local time at the origin matters. Here, the local time at all values of the spatial parameter influences the limit theory for the H-regular functions. Sample covariance asymptotics also differ between I- and H-regular functions. As noted earlier, sample covariances for I-regular functions have mixed normal limits. However, for H-regular functions, the limit is given in terms of a stochastic integral and is generally non-Gaussian.

Just as for regular families of functions, I- and H-regular families of functions are closed under the operations of addition, subtraction, and multiplication. It is obvious that they are closed under addition and subtraction. That they are closed under multiplication is proved in Lemma A6 in the Appendix. It is also straightforward to show that all the regularity (I- and H-regularity) conditions are preserved, if we compose a regular (I- and H-regular) family with certain types of functions. For instance, if F is regular (I- and H-regular), then so is $|F|$. This property will be used in some of our proofs. In subsequent discussion, we sometimes use the term regularity to mean any of I- and H-regularity as well as regularity in the narrow sense. This should cause no confusion.

4. CONSISTENCY

This section establishes the consistency of the NLS estimator $\hat{\theta}_n$ defined in (11). The conditions for consistency are easy to verify and, in particular, do not require differentiability of the regression function. They are satisfied for most of the commonly used nonlinear regression functions, and in these cases consistency of the NLS is readily established. However, there are some regression functions that are not covered by the conditions we impose for the consistency results in this section. They will be considered in the next section, where we derive the asymptotic distributions of the NLS estimator under stronger assumptions including differentiability of the regression function.

Define $D_n(\theta, \theta_0) = Q_n(\theta) - Q_n(\theta_0)$. To prove consistency, we show one of the following two conditions (labelled and numbered CN for reference).

CN1: *For some sequence ν_n of numbers, $\nu_n^{-1} D_n(\theta, \theta_0) \to_p D(\theta, \theta_0)$ uniformly in θ as $n \to \infty$, where $D(\cdot, \theta_0)$ is continuous and has unique minimum θ_0 a.s.*

CN2: *For any $\delta > 0$, $\liminf_{n\to\infty} \inf_{|\theta - \theta_0| \geq \delta} D_n(\theta, \theta_0) > 0$ in probability.*

Both consistency conditions CN1 and CN2 are sufficient to ensure that $\hat{\theta}_n \to_p \theta_0$, as shown in earlier work by Jennrich (1969) and Wu (1981).

For the standard nonlinear regression, Jennrich (1969) proves the consistency of the NLS estimator by establishing CN1. On the other hand, Wu (1981) derives the consistency of the NLS estimator for possibly nonstationary nonlinear regressions through CN2. Given the results in Section 3, it is not hard to show that the required conditions hold for regressions with various types of regular regression functions. The Jennrich approach is more appropriate for regression with I-regular and H_0-regular regression functions, since the regression functions converge at the same rate for all values of θ. We therefore show that CN1 is satisfied for such functions, under some identifying assumption that guarantees that $D(\cdot, \theta_0)$ has unique minimum θ_0. However, this approach is not applicable for general H-regular functions. These functions have different rates of convergence for different values of θ, and the results obtained by Wu are then more relevant. Therefore, CN2 will be shown to hold for these functions.

THEOREM 4.1: *Let Assumption* 2.2 *hold*, *and let* f *be I-regular on* Θ. *If* $\int_{-\infty}^{\infty}(f(s,\theta)-f(s,\theta_0))^2\,ds > 0$ *for all* $\theta \neq \theta_0$, *then CN*1 *holds*. *In particular*, *we have*

$$D(\theta,\theta_0) = \left(\int_{-\infty}^{\infty}(f(s,\theta)-f(s,\theta_0))^2\,ds\right)L(1,0)$$

with $\nu_n = \sqrt{n}$.

Roughly speaking, the result in Theorem 4.1 holds for all integrable functions that are bounded and piecewise smooth over their supports. We only require a mild identifying assumption for the consistency result to go through. It holds for many density type regression functions, as well as all linear-in-parameter regression functions of the type $f(x,\theta) = \theta a(x)$ with nonzero I-regular a. Theorem 4.1 is also applicable for nonlinear regression with regression function $f(x,\theta) = e^{-\theta x^2}$ with $\Theta \subset \mathbf{R}_+$, as long as $p > 8$.

It is interesting to compare two nonlinear nonstationary regressions with $f(x,\theta) = e^{-\theta x^2}$ and different specifications for x_t: the first with an integrated time series x_t, as in the paper here, and the second with $x_t = \sqrt{t}$. The latter is sometimes referred to as the first order decay model. The two regressors are comparable, since $x_t \approx O_p(\sqrt{t})$. However, their asymptotic behaviors are drastically different. The NLS estimator is consistent for the former, but for the latter it is inconsistent, as noted earlier by Malinvaud (1970). The reason is simple. In the deterministic decay model, the signal from the regressor is asymptotically negligible because $e^{-\theta t}$ tends monotonically to zero, so in the limit there is no information in the mean function about θ. On the other hand, in the stochastic trend case, while it is true that the stochastic order of x_t is $O_p(\sqrt{t})$, the process is recurrent rather than monotonic and x_t^2 keeps returning to the vicinity of the origin. In consequence, the regressor continues to carry information about the parameter θ as $t \to \infty$.

THEOREM 4.2: *Let Assumption* 2.1 *hold, and let f be H_0-regular on Θ with asymptotic order κ and limit homogeneous function h. Assume*:

(a) $\kappa(\lambda)$ *is bounded away from zero as* $\lambda \to \infty$, *and*
(b) *for all* $\theta \neq \theta_0$ *and* $\delta > 0$, $\int_{|s| \leq \delta}(h(s,\theta) - h(s,\theta_0))^2\, ds > 0$.

*Then CN*1 *holds. In particular, we have*

$$D(\theta,\theta_0) = \int_{-\infty}^{\infty} (h(s,\theta) - h(s,\theta_0))^2 L(1,s)\, ds$$

with $\nu_n = n\kappa(\sqrt{n})^2$.

Condition (a) ensures that there is sufficient variability in the regression function asymptotically to generate a signal stronger than the noise. Condition (b) is simply an identification condition for the regression with H_0-regular regression functions. Unfortunately, this condition fails to hold for some commonly used H_0-regular regression functions. These can have a limit homogeneous function $h(x)$, say, that is independent of θ, so that the true limit homogeneous function is not identified. However, we can usually achieve identification by properly reformulating the regression function in this case. Indeed, we may simply consider a regression with the transformed regression function $f_*(x,\theta) = f(x,\theta) - h(x)$ to avoid the lack of identification. Clearly, the regression (10) can then be rewritten as

$$y_t - h(x_t) = f_*(x_t,\theta_0) + u_t$$

so the regression is effectively the same as one with the regression function f_*.

EXAMPLE 4.1: (a) For the linear-in-parameter regression function $f(x,\theta) = \theta a(x)$, Theorem 4.2 is particularly easy to apply. Note that this f is H_0-regular if, in the representation $a(\lambda x) = \kappa(\lambda)b(x) + r(x,\lambda)$, b is regular and $r(x,\lambda)$ is of order smaller than $\kappa(\lambda)$. The asymptotic order and limit homogeneous function of f are given respectively by $\kappa(\lambda)$ and $h(x,\theta) = \theta b(x)$. Condition (b) of Theorem 4.2 is satisfied whenever $\int_{|s| \leq \delta}|b(s)|\, ds > 0$ for all $\delta > 0$. One may now easily see that Theorem 4.2 is applicable for regression functions such as $f(x,\theta) = \theta e^x/(1+e^x)$, $\theta \log|x|$, and $\theta|x|^k$, among many others. The asymptotic orders of these functions are given respectively by $\kappa(\lambda) = 1$, $\log \lambda$, and λ^k. The corresponding limit homogeneous functions are $h(x,\theta) = \theta 1\{x \geq 0\}$, θ, and $\theta|x|^k$.

(b) Consider the regression function given by $f(x,\theta) = x(1+\theta x)^{-1}1\{x \geq 0\}$ with $\Theta \subset \mathbf{R}_+$. This is a reparameterized and restricted version of the Michaelis-Menten model used in Bates and Watts (1988) to fit the relationship between the velocity of an enzymatic reaction and the substrate concentration. One may easily see that it is H_0-regular with asymptotic order $\kappa(\lambda) = 1$ and limit homogeneous function $h(x,\theta) = \theta^{-1}1\{x \geq 0\}$. Clearly, it satisfies all the conditions of Theorem 4.2.

(c) The result in Theorem 4.2 is not directly applicable to the regression function $f(x,\theta)=(x+\theta)^2$, which was considered in Wu (1981). Clearly, the function is H_0-regular with asymptotic order $\kappa(\lambda)=\lambda^2$ for which condition (a) holds. However, this function has limit homogeneous function $h(x)=x^2$ for all values of θ, which fails to satisfy condition (b). Nevertheless, as indicated above, we may reformulate the regression with the regression function $f_*(x,\theta)=f(x,\theta)-h(x)=(x+\theta)^2-x^2=2\theta x+\theta^2$ to apply Theorem 4.2. Obviously, f_* is H_0-regular with asymptotic order $\kappa_*(\lambda)=\lambda$ and limit homogeneous function $h_*(x,\theta)=2\theta x$, and satisfies the conditions of Theorem 4.2.

EXAMPLE 4.2: The logistic regression function $f(x,\theta)=e^{\theta x}/(1+e^{\theta x})$ has the same lack of identification problem as the model in Example 4.1(c). The function is H_0-regular with the asymptotic order $\kappa(\lambda)=1$, and therefore satisfies condition (a). However, the limit homogeneous function is given by $h(x)=1\{x\geq 0\}$ and the identification condition (b) fails. To analyze such a regression model, we need to reformulate the model in terms of the regression function

$$f_*(x,\theta)=f(x,\theta)-h(x)=\frac{e^{\theta x}}{1+e^{\theta x}}1\{x<0\}-\frac{1}{1+e^{\theta x}}1\{x\geq 0\}.$$

The reformulated regression function f_*, however, is no longer H_0-regular. However, it is I-regular and satisfies the conditions of Theorem 4.1. So our asymptotic theory is applicable in this case also.

THEOREM 4.3: *Let Assumption* 2.1 *hold, and let f be H-regular on Θ with asymptotic order κ and limit homogeneous function h. Then CN*2 *holds if*:

(a) *for any* $\bar\theta\neq\theta_0$ *and* $\bar p,\bar q>0$, *there exist* $\varepsilon>0$ *and a neighborhood N of* $\bar\theta$ *such that as* $\lambda\to\infty$

$$\inf_{\substack{|p-\bar p|<\varepsilon\\ |q-\bar q|<\varepsilon}}\ \inf_{\theta\in N}|p\kappa(\lambda,\theta)-q\kappa(\lambda,\theta_0)|\to\infty;$$

(b) *for all* $\theta\in\Theta$ *and* $\delta>0$, $\int_{|s|\leq\delta}h(s,\theta)^2\,ds>0$.

EXAMPLE 4.3: Consider the Box-Cox transformation $f(x,\theta)=(|x|^{\theta}-1)/\theta$ with $\theta\in\Theta\subset\mathbf{R}_+$. It is straightforward to see that f is H-regular with asymptotic order κ and limit homogeneous function h given by $\kappa(\lambda,\theta)=\lambda^{\theta}$ and $h(x,\theta)=|x|^{\theta}/\theta$, respectively. We may easily show that such an f satisfies the conditions of Theorem 4.3. It is obvious that condition (b) holds. To see that condition (a) is satisfied, set $0<\varepsilon<\min(\bar p,\bar q)$ and, for any given $\bar\theta\neq\theta_0$, let N be any neighborhood of $\bar\theta$ such that $\theta_0\notin N$.

COROLLARY 4.4: *Suppose that the assumptions in Theorem* 4.1 *or Theorem* 4.2 *with* $\kappa<\infty$ *hold. Then* $\hat\sigma_n^2\to_p\sigma^2$, *as* $n\to\infty$.

By this Corollary, the error variance estimator $\hat{\theta}_n^2$ is consistent in nonlinear regressions with regression functions that are I-regular or H_0-regular with nonexplosive asymptotic order. Consistency for nonlinear regressions with more general H-regular regression functions will be shown in the next section under stronger assumptions.

It is interesting to note that $\hat{\sigma}_n^2 = (1/n)\sum_{t=1}^n y_t^2 \to_p \sigma^2$ for a nonlinear regression with an I-regular regression function. This follows immediately from Theorem 3.2. We may therefore estimate the error variance directly from y_t, rather than residuals. However, the convergence rate of the resulting estimator $\tilde{\sigma}_n^2$ is slower. For, if we define $\sigma_n^2 = (1/n)\sum_{t=1}^n u_t^2$, then $\tilde{\sigma}_n^2 = \sigma_n^2 + O_p(n^{-1/2})$ whereas $\hat{\sigma}_n^2 = \sigma_n^2 + o_p(n^{-1/2})$. This is shown in the proof of Corollary 4.4.

5. LIMIT DISTRIBUTIONS

This section of the paper derives the asymptotic distribution of the NLS estimator $\hat{\theta}_n$ defined in (11). As in standard nonlinear regression theory, we require conditions on the regression function that ensure it is sufficiently smooth as a function of the unknown parameter θ. Assuming differentiability of the regression function also allows us to establish the consistency of the NLS in models where the results in the previous section are not applicable. For such models, the results in this section will give consistency as well as the asymptotic distribution of the NLS estimator.

Define

$$\dot{f} = \left(\frac{\partial f}{\partial \theta_i}\right), \qquad \ddot{f} = \left(\frac{\partial^2 f}{\partial \theta_i\, \partial \theta_j}\right), \qquad \dddot{f} = \left(\frac{\partial^3 f}{\partial \theta_i\, \partial \theta_j\, \partial \theta_k}\right)$$

to be all vectors, arranged by the lexicographic ordering of their indices. It is sometimes more convenient to define the second derivatives of f in matrix form as in $\ddot{F} = \partial^2 f/\partial\theta\,\partial\theta'$. Clearly, we may obtain $\ddot{f}$ from $\ddot{F}$ by stacking its rows into a column vector. In what follows, we denote by $\dot{h}$ the limit homogeneous function of H-regular $\dot{f}$. Moreover, the asymptotic orders of $\dot{f}$, $\ddot{f}$, and $\dddot{f}$ of H-regular functions will be written as $\dot{\kappa}$, $\ddot{\kappa}$, and $\dddot{\kappa}$. Whenever $\dot{f}$, $\ddot{f}$, and $\dddot{f}$ are introduced, we assume that they exist.

Now let $\dot{Q}_n$ and $\ddot{Q}_n$ be the first and second derivatives of Q_n with respect to θ defined in the usual way, i.e., $\dot{Q}_n = \partial Q_n/\partial\theta$ and $\ddot{Q}_n = \partial^2 Q_n/\partial\theta\,\partial\theta'$. We have

$$\dot{Q}_n(\theta) = -\sum_{t=1}^n \dot{f}(x_t, \theta)(y_t - f(x_t, \theta)),$$

$$\ddot{Q}_n(\theta) = \sum_{t=1}^n \dot{f}(x_t, \theta)\dot{f}(x_t, \theta)' - \sum_{t=1}^n \ddot{F}(x_t, \theta)(y_t - f(x_t, \theta)),$$

ignoring a constant, which is unimportant. As in standard nonlinear regression, the asymptotic distribution of $\hat{\theta}_n$ in our model can be obtained from the first order Taylor expansion of $\dot{Q}_n$, which is written as

$$\dot{Q}_n\left(\hat{\theta}_n\right) = \dot{Q}_n(\theta_0) + \ddot{Q}_n(\theta_n)\left(\hat{\theta}_n - \theta_0\right), \tag{17}$$

where θ_n lies in the line segment connecting $\hat{\theta}_n$ and θ_0. We have $\dot{Q}_n(\hat{\theta}_n) = 0$ if $\hat{\theta}_n$ is an interior solution to the minimization problem (11).

Let $\dot{f}$ be one of the regular functions introduced in Section 3. For an appropriately chosen normalizing sequence ν_n, it follows immediately from the sample covariance asymptotics in Section 3 that $\nu_n^{-1}\dot{Q}_n(\theta_0) \to_d \dot{Q}(\theta_0)$ for some random vector $\dot{Q}(\theta_0)$. Also, if we let

$$\ddot{Q}_n^0(\theta_0) = \sum_{t=1}^{n} \dot{f}(x_t, \theta_0)\dot{f}(x_t, \theta_0)',$$

then $\nu_n^{-1}\ddot{Q}_n^0(\theta_0)\nu_n^{-1\prime} \to_p \ddot{Q}(\theta_0)$ for some random matrix $\ddot{Q}(\theta_0)$, due to Lemma A6 and sample mean asymptotics in Section 3. Therefore, under suitable conditions that ensure $\nu_n^{-1}\ddot{Q}_n(\theta_n)\nu_n^{-1\prime} = \nu_n^{-1}\ddot{Q}_n^0(\theta_0)\nu_n^{-1\prime} + o_p(1)$ and $\ddot{Q}(\theta_0) > 0$ a.s., we may expect from (17) that

$$\begin{aligned} \nu_n'\left(\hat{\theta}_n - \theta_0\right) &= -\left(\nu_n^{-1}\ddot{Q}_n(\theta_n)\nu_n^{-1\prime}\right)^{-1}\nu_n^{-1}\dot{Q}_n(\theta_0) \\ &= -\left(\nu_n^{-1}\ddot{Q}_n^0(\theta_0)\nu_n^{-1\prime}\right)^{-1}\nu_n^{-1}\dot{Q}_n(\theta_0) + o_p(1) \\ &\to_d -\ddot{Q}(\theta_0)^{-1}\dot{Q}(\theta_0) \end{aligned} \tag{18}$$

as $n \to \infty$.

For easy reference, we list a set of sufficient conditions (labelled and numbered AD for reference) that lead to (18), using the notation introduced above.

AD1: $\nu_n^{-1}\dot{Q}_n(\theta_0) \to_d \dot{Q}(\theta_0)$ *as* $n \to \infty$.

AD2: $\nu_n^{-1}\ddot{Q}_n(\theta_0)\nu_n^{-1\prime} = \nu_n^{-1}\ddot{Q}_n^0(\theta_0)\nu_n^{-1\prime} + o_p(1)$ *for large* n.

AD3: $\nu_n^{-1}\ddot{Q}_n(\theta_0)\nu_n^{-1\prime} \to_p \ddot{Q}(\theta_0)$ *as* $n \to \infty$.

AD4: $\ddot{Q}(\theta_0) > 0$ *a.s.*

AD5: $\dot{Q}_n(\hat{\theta}_n) = 0$ *with probability approaching to one as* $n \to \infty$.

AD6: $\nu_n^{-1}(\ddot{Q}_n(\theta_n) - \ddot{Q}_n(\theta_0))\nu_n^{-1\prime} \to_p 0$ *as* $n \to \infty$.

The asymptotic distribution conditions AD1–AD6 are standard in nonlinear regression analysis. Given AD1–AD6, (18) follows immediately from (17).

It is generally simple to check AD1–AD4 for a given nonlinear regression. For all types of regular regression functions, AD1–AD3 directly follow from the results in Section 3, if we properly choose the normalizing sequence ν_n. Moreover, given AD2, AD4 can readily be deduced under an identifying assumption to avoid asymptotic multicollinearity in $\dot{f}$. For regressions with I- or H_0-regular $\dot{f}$ and $\ddot{f}$, it is also not difficult to show that AD5 and AD6 hold if we presume the consistency of $\hat{\theta}_n$, as established in Theorems 4.1 and 4.2. Clearly, AD5 is an immediate consequence of the assumption that θ_0 is an interior point of Θ. Moreover, AD6 can also be easily deduced for this type of regression because, for a normalizing sequence ν_n independent of θ, $\nu_n^{-1}\ddot{Q}_n^0(\theta)\nu_n^{-1'}$ converges uniformly to a continuous function $\ddot{Q}(\theta)$, say, of θ.

THEOREM 5.1: *Let Assumption* 2.2 *hold. Assume*

(a) *f satisfies conditions in Theorem* 4.1,
(b) *$\dot{f}$ and $\ddot{f}$ are I-regular on Θ, and*
(c) $\int_{-\infty}^{\infty} \dot{f}(s,\theta_0)\dot{f}(s,\theta_0)'\,ds > 0$.

Then we have

$$\sqrt[4]{n}\left(\hat{\theta}_n - \theta_0\right) \to_d \left(L(1,0)\int_{-\infty}^{\infty}\dot{f}(s,\theta_0)\dot{f}(s,\theta_0)'\,ds\right)^{-1/2} W(1)$$

as $n \to \infty$.

The conditions in Theorem 5.1 hold for a wide range of integrable regression functions that are used in practical applications. They are satisfied, for instance, for the regression function $f(x,\alpha,\beta) = \alpha e^{-\beta x^2}$, where $\theta = (\alpha,\beta) \in \Theta \subset \mathbf{R} \times \mathbf{R}_+$. Clearly, the results in Theorem 5.1 are also applicable for all nonzero I-regular linear-in-parameter regression functions. For regressions with I-regular regression functions, the NLS estimator converges at the rate of $\sqrt[4]{n}$, and has a mixed Gaussian limiting distribution. The asymptotic theory is not likely to provide a good approximation in small samples for regressions with I-regular regression functions, due to the slower than usual rate of convergence. However, this matter needs to be investigated in simulations.

As one may well expect from our earlier results, the asymptotic behavior of the NLS estimator can be quite different for regressions with other types of regression functions. For regressions with H_0-regular regression functions, the convergence rate is given by $\sqrt{n}\,\dot{\kappa}(\sqrt{n})$. It therefore converges faster than the standard $\sqrt{n}$ rate, when $\dot{\kappa}(\sqrt{n})$ diverges, as is usually the case. The limiting distribution theory, however, is not Gaussian, except for the special case where $\sigma_{uv} = 0$. We have the following result in this case.

THEOREM 5.2: *Let Assumption* 2.1 *hold. Assume*

(a) *f satisfies conditions in Theorem* 4.2,
(b) *$\dot{f}$ and $\ddot{f}$ are H_0-regular on Θ,*
(c) $\|(\dot{\kappa} \otimes \dot{\kappa})^{-1}\kappa\ddot{\kappa}\| < \infty$, *and*
(d) $\int_{|s| \le \delta} \dot{h}(s,\theta_0)\dot{h}(s,\theta_0')\,ds > 0$ *for all* $\delta > 0$.

Then

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

as $n \to \infty$.

Theorem 5.2 is applicable for all H_0-regular regression functions considered in Example 4.1. It is easy to see that the conditions in Theorem 5.2 hold for all the linear-in-parameter regression functions in Example 4.1(a). The zero function can of course be regarded as H_0-regular with any asymptotic order, since it has zero limit homogeneous function. For the regression function $f(x,\theta)=x(1+\theta x)^{-1}1\{x\geq 0\}$ in Example 4.1(b), both $\dot{f}(x,\theta)=-x^2(1+\theta x)^{-2}1\{x\geq 0\}$ and $\ddot{f}(x,\theta)=2x^3(1+\theta x)^{-3}1\{x\geq 0\}$ are H_0-regular with $\dot{\kappa}(\lambda)=\ddot{\kappa}(\lambda)=1(\equiv\kappa(\lambda))$ and $\dot{h}(x,\theta)=-\theta^{-2}1\{x\geq 0\}$. One may easily check that all the conditions of Theorem 5.2 are met. The reformulated regression function $f(x,\theta)=2\theta x+\theta^2$ in Example 4.1(c) also satisfies conditions in Theorem 5.2. Both $\dot{f}$ and $\ddot{f}$ are H_0-regular in this case, respectively with $\dot{\kappa}(\lambda)=\lambda$ and $\dot{h}(x,\theta)=2x$, and $\ddot{\kappa}(\lambda)=1$ and $\ddot{h}(x,\theta)=2$. Another example would be the regression function $f(x,\alpha,\beta)=(\alpha+\beta e^x)/(1+e^x)$ with $\theta=(\alpha,\beta)\in\Theta\subset\mathbf{R}^2$, which represents smooth transition from level α to level β. It is easily seen that this function satisfies CN2 and Theorem 5.2 holds.

It is more difficult to establish AD5 and AD6 for general H-regular functions. If we assume consistency, AD5 would follow immediately. However, we still cannot invoke the uniform convergence of $\ddot{Q}_n(\theta)$ to prove AD6, since the convergence rate is dependent upon θ. Moreover, the conditions of Theorem 4.3, where the consistency of $\hat{\theta}_n$ is established for general H-regular functions, are somewhat restrictive and do not allow for some commonly used regression functions. Here we do not presume consistency to derive the asymptotic distributions. For our approach, we need to introduce the following condition:

AD7: *There is a sequence* μ_n *such that* $\mu_n\nu_n^{-1}\to_{a.s.}0$, *and such that*

$$\sup_{\theta\in N_n}\left\|\mu_n^{-1}\left(\ddot{Q}_n(\theta)-\ddot{Q}_n(\theta_0)\right)\mu_n^{-1\prime}\right\|\to_p 0$$

where $N_n=\{\theta:\|\mu_n'(\theta-\theta_0)\|\leq 1\}$.

As shown in Wooldridge (1994), AD7 implies both AD5 and AD6, given AD1–AD4. Therefore, we may check AD1–AD4 and AD7, instead of AD1–AD6, to deduce (18).

We now present the asymptotic distribution of $\hat{\theta}_n$ for regressions with H-regular $\dot{f}$. For notational brevity, we write $\dot{\kappa}_0(\cdot)=\dot{\kappa}(\cdot,\theta_0)$. Moreover, to properly formulate a sufficient set of conditions for AD7, define a neighborhood of θ_0 by

$$N(\varepsilon,\lambda)=\{\theta:\|\dot{\kappa}_0(\lambda)'(\theta-\theta_0)\|\leq\lambda^{-1+\varepsilon}\}$$

for $\varepsilon>0$ given. We have the following theorem.

THEOREM 5.3: *Let Assumption* 2.1 *hold. Assume*

(a) $\dot{f}$ *is H-regular on* Θ,
(b) *for any* $\bar{s}>0$ *given, there exists* $\varepsilon>0$ *such that as* $\lambda\to\infty$

(19) $$\left\| (\dot{\kappa}_0 \otimes \dot{\kappa}_0)(\lambda)^{-1} \left(\sup_{|s| \le \bar{s}} |\ddot{f}(\lambda s, \theta_0)| \right) \right\| \to 0,$$

(20) $$\lambda^{-1+\varepsilon} \left\| (\dot{\kappa}_0 \otimes \dot{\kappa}_0)(\lambda)^{-1} \left(\sup_{|s| \le \bar{s}} \sup_{\theta \in N(\varepsilon, \lambda)} |\ddot{f}(\lambda s, \theta)| \right) \right\| \to 0,$$

(21) $$\lambda^{-1+\varepsilon} \left\| \left(\dot{\kappa}_0 \otimes \dot{\kappa}_0 \otimes \dot{\kappa}_0 \right)(\lambda)^{-1} \left(\sup_{|s| \le \bar{s}} \sup_{\theta \in N(\varepsilon, \lambda)} |\dddot{f}(\lambda s, \theta)| \right) \right\| \to 0;$$

(c) $\int_{|s| \le \delta} \dot{h}(s, \theta_0) \dot{h}(s, \theta_0)' \, ds > 0$ *for all* $\delta > 0$.

Then

$$\sqrt{n}\, \dot{\kappa}_0(\sqrt{n})' \left(\hat{\theta}_n - \theta_0 \right) \to_d \left(\int_0^1 \dot{h}(V, \theta_0) \dot{h}(V, \theta_0)' \right)^{-1} \int_0^1 \dot{h}(V, \theta_0) \, dU$$

as $n \to \infty$.

REMARKS: For the regression function f with H-regular $\dot{f}$ specified as in (a), the identification condition is given by (c). The conditions in (a) and (c) are usually easy to check. The conditions in (b), however, are awkward and cumbersome. We may replace them with a stronger, yet easier to verify, condition as discussed below.

(a) For many H-regular functions, there exists $\varepsilon > 0$ such that

(22) $$\lambda^{\varepsilon} \left\| (\dot{\kappa}_0 \otimes \dot{\kappa}_0)(\lambda)^{-1} \left(\sup_{|s| \le \bar{s}} \sup_{\theta \in N(\varepsilon, \lambda)} |\ddot{f}(\lambda s, \theta)| \right) \right\| \to 0$$

for any $\bar{s} > 0$. Clearly, (19) and (20) are satisfied if (22) holds. Moreover, we show in the proof of Theorem 5.3 that thrice differentiability of f and (21) are unnecessary if (22) holds true. Conditions in (19)–(21) can thus be replaced by (22), which is much simpler to check.

(b) Define $N(\delta) = \{\theta : \|\theta - \theta_0\| < \delta\}$ for $\delta > 0$. If there exists $\varepsilon > 0$ such that

(23) $$\lambda^{-1+\varepsilon} \left\| \dot{\kappa}_0(\lambda)^{-1} \right\| \to 0$$

as $\lambda \to \infty$, then we have for any $\delta > 0$ $N(\varepsilon, \lambda) \subset N(\delta)$ when λ is sufficiently large. Therefore, it suffices to show that there exist $\varepsilon > 0$ satisfying (23) and

(24) $$\lambda^{\varepsilon} \left\| (\dot{\kappa}_0 \otimes \dot{\kappa}_0)(\lambda)^{-1} \left(\sup_{|s| \le \bar{s}} \sup_{\theta \in N(\delta)} |\ddot{f}(\lambda s, \theta)| \right) \right\| \to 0$$

for some $\delta > 0$, instead of (22). It is indeed quite easy and straightforward to show that (23) and (24) hold for many H-regular functions that are used in nonlinear analyses.

EXAMPLE 5.1: Let $\Theta \subset \mathbf{R}_{+}^{2}$ and write $\theta = (\alpha, \beta)$. Consider $f(x, \alpha, \beta) = \alpha x^{\beta}$. It is straightforward to see that f is H-regular with asymptotic order κ and limit homogeneous function h given respectively by

$$\kappa(\lambda, \alpha, \beta) = \alpha\lambda^{\beta} \qquad \text{and} \qquad h(x, \alpha, \beta) = x^{\beta}.$$

Moreover, we have

$$\dot{\kappa}(\lambda, \alpha, \beta) = \begin{pmatrix} \lambda^{\beta} & 0 \\ \alpha\lambda^{\beta} \log \lambda & \alpha\lambda^{\beta} \end{pmatrix} \quad \text{and} \quad \dot{h}(x, \alpha, \beta) = \begin{pmatrix} x^{\beta} \\ x^{\beta} \log x \end{pmatrix}.$$

It is obvious that conditions (23) and (24) are satisfied for f. To show this, we simply let ε and δ in (23) and (24) be any numbers such that $0 < \varepsilon, \delta < \beta_0$ for any β_0.

The asymptotic results for regressions with general H-regular regression functions are essentially identical to those for regressions with H_0-regular regression functions given in Theorem 5.2. The only difference is that the convergence rate is now given by $\dot{\kappa}_0(\sqrt{n})$, which is dependent upon the true value θ_0 of θ.

COROLLARY 5.4: *Suppose that the assumptions in Theorem* 5.3 *or Theorem* 5.4 *hold. Then* $\hat{\sigma}_n^2 \to_p \sigma^2$ *as* $n \to \infty$.

Corollaries 4.4 and 5.4 establish the consistency of the error variance estimator $\hat{\sigma}_n^2$ for regressions with I- and H-regular regression functions. The estimator can therefore be used for consistent estimation of the error variance in a wide class of nonlinear regressions. In consequence, hypotheses about θ can be tested using standard procedures like the Wald, Lagrange multiplier, and likelihood ratio tests. The statistical limit theories for these tests are straightforward given the results in this section. For regressions with I-regular regression functions, test statistics of the usual form all have limiting chi-square distributions, respectively, under the assumptions of Theorem 5.1. However, for regressions with H-regular regression functions, they are generally dependent upon a nuisance parameter generated by σ_{uv}. These become chi-square under the assumptions of Theorems 5.2 and 5.3, only when $\sigma_{uv} = 0$.

6. CONCLUSION

This paper develops some new technology that makes possible the analysis of nonlinear regressions with unit root nonstationary time series. The techniques rely on the spatial properties of Brownian motion and these are used to assist in representing the limiting forms of sample moments and sample covariance functions of integrated time series. Under fairly general conditions and for an extensive family of nonlinear regression functions, the paper proves the consistency of nonlinear regressions, finds rates of convergence, and obtains forms for the limit distributions. The convergence rates can be both slower ($\sqrt[4]{n}$) and faster (powers of $\sqrt{n}$) than that of traditional nonlinear regression, depending on whether the signal is attenuated or strengthened by the presence of integrated regressors. When the regression function involves exponentials, it is shown that the convergence rates are path dependent. For the regressions with integrable regression functions, the limit distributions of the nonlinear regression estimators are mixed normal as long as the equation errors are martingale differences. In such cases, nonlinear inference procedures apply in the usual manner, so that although the estimators may have non-Gaussian limit distributions, inference is unaffected.

Some remarks on the limitations of this present work are in order. Broadly stated, our purpose has been to initiate nonlinear econometric analysis for stochastically nonstationary time series and provide new tools that are equal to the task of providing an asymptotic theory. We have not attempted a complete theory that encompasses all, or even most, of the interesting cases that can be expected to arise in applied work. As we have seen, one of the distinguishing characteristics of this new field is that the spatial features of a time series can play a significant role in the asymptotics. In some cases, even the rate of convergence of a nonlinear estimator can be influenced by the sample path of the regressors, making a significant departure from traditional nonlinear asymptotic theory. The models we have studied cover the case of parametric nonlinear cointegration and should prove useful in some empirical studies of nonlinear cointegrating links between economic time series where martingale difference errors can be expected, as in formulations arising from rational expectations models. A general treatment that applies to practical cases in which short run dynamics are unspecified is still to be provided. Moreover, a combination of the ideas presented here and those in our other paper, Phillips and Park (1998), is needed to form the basis of a nonparametric analysis of cointegration. The results of this paper are also limited to nonlinear regressions with a single integrated regressor. The theory for general regressions with multiple regressors can be expected to differ, often in substantial ways, from the theory presented here. Indeed, such differences are to be anticipated, because the asymptotics for integrated time series are naturally approximated by functionals of Brownian motion, and functionals of vector Brownian motion can behave quite differently from those of scalar Brownian motion. For instance, it is well known (e.g., Revuz and Yor (1994, Ch. V)) that while vector Brownian motion is recurrent in $\mathbf{R}^d$ for $d \leq 2$, it is transient when $d \geq 3$, so that the spatial behavior of multivariate Brownian motion involves major changes as the dimension rises above the univariate case. In a certain sense, the commonly cited curse of dimensionality is worse in the nonstationary case because the number of visits to a spatial location are more drastically reduced as the dimension increases than they are in a stationary environment. Nevertheless, some statistical theory for nonlinear regressions with more than one integrated regressor is possible, especially in econometric models that involve a single index, and this theory is currently under development by the authors. Readers are referred to Park and Phillips (2000) for some work along these lines with binary choice models.

School of Economics, Seoul National University, Seoul 151-742, Korea; jpark@plaza.snu.ac.kr; http://econ.snu.ac.kr/faculty/jpark/index-e.html

and

Cowles Foundation for Research in Economics, Yale University, Box 208281, New Haven, CT 06520-8281, U.S.A.; University of Auckland and University of York; peterphillips@yale.edu; http://korora.econ.yale.edu

Manuscript received July, 1998; final revision received November, 1999.

APPENDIX A: TECHNICAL RESULTS FOR I(1) FUNCTIONALS

Useful Lemmas

We give several lemmas that will be used repeatedly in the proofs of the main theorems and corollaries. The proofs of these lemmas are given in the next section.

LEMMA A1: *Let T_1 and T_2 be transformations on* $\mathbf{R}$. *If T_1 and T_2 are regular, then so are $T_1 \pm T_2$ and $T_1 T_2$.*

LEMMA A2: *Let Assumption* 2.1 *hold. If T is regular, then*

$$\frac{1}{n}\sum_{t=1}^{n} T\left(\frac{x_t}{\sqrt{n}}\right) \to_{a.s.} \int_0^1 T(V(r))\,dr,$$

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} T\left(\frac{x_t}{\sqrt{n}}\right) u_t \to_d \int_0^1 T(V(r))\,dU(r),$$

as $n \to \infty$.

LEMMA A3: (a) *If $F(\cdot,\pi)$ is a regular family on Π and $\pi_0 \in \Pi$, then there is a neighborhood N_0 of π_0 such that $\sup_{\pi\in N} F(\cdot,\pi)$ and $\inf_{\pi\in N} F(\cdot,\pi)$ are regular for all $N \subset N_0$.*

(b) *If F is regular on Π and Π is compact, then $\sup_{\pi\in\Pi}|F(\cdot,\pi)|$ is locally bounded.*

LEMMA A4: *Let Assumption* 2.1 *hold. Then as $n\to\infty$:*

(a) $(1/n)\sum_{t=1}^{n} T(x_t/\sqrt{n}) = O_{a.s.}(1)$ *for* $T = \mathscr{T}_{LB}$.
(b) $(1/n)\sum_{t=1}^{n} T_1(x_t/\sqrt{n})T_2(x_t) = o_{a.s.}(1)$ *for* $T_1 \in \mathscr{T}_{LB}$, $T_2 \in T_B^0$.
(c) $(1/\sqrt{n})\sum_{t=1}^{n} T(x_t/\sqrt{n})u_t = O_p(1)$ *for* $T \in \mathscr{T}_{LB}^0$.
(d) $(1/\sqrt{n})\sum_{t=1}^{n} T_1(x_t/\sqrt{n})T_2(x_t)u_t = o_p(1)$ *for* $T_1 \in \mathscr{T}_{LB}^0$, $T_2 \in \mathscr{T}_B^0$.

LEMMA A5: (a) *Let (Z, z, Φ) and (Z_i, z_i, Φ_i), $i = 1,2$, be defined as in Definition* 3.4, *and let $W: \mathbf{R}\times\Phi \to \mathbf{R}^m$ be such that $\sup_{w\in\Phi} W(\cdot,\omega) \in \mathscr{T}_{LB}^0$. If Z is of order smaller than z on Φ, then $W\otimes Z$ is of order smaller than $I_m \otimes z$. Moreover, if Z_i is of order smaller than z_i on Φ_i for $i=1,2$, then $Z_1\otimes Z_2$ is of order smaller than $z_1\otimes z_2$ on $\Phi_1\times\Phi_2$.*

(b) *Suppose that (Z, z, Φ) and (Z_i, z_i, Φ_i), $i=1,2$, be defined as in Definition* 3.4 *and that Z is of order smaller than z on Φ. Also, let Assumption* 2.1 *hold, and write $Z_{nt}(\omega) = Z(x_t/\sqrt{n}, \sqrt{n}, \omega)$ and $z_n(\omega) = z(\sqrt{n}, \omega)$ for short. Then we have $n^{-1}z_n(\omega)^{-1}\sum_{t=1}^{n} Z_{nt}(\omega) \to_{a.s.} 0$ and $n^{-1}z_n(\omega)^{-1}\sum_{t=1}^{n} Z_{nt}(\omega)u_t \to_p 0$ uniformly in $\omega\in\Phi$. Moreover, for each $\omega\in\Phi$ we have $n^{-1/2}z_n(\omega)^{-1}\sum_{t=1}^{n} Z_{nt}(\omega)u_t \to_p 0$.*

LEMMA A6: *Let* $F_i : R \times \Pi_i \to R$ *for* $i = 1,2$, *and let* $\Pi = \Pi_1 \times \Pi_2$. *Define* $F : R \times \Pi \to R$ *by* $F(\cdot, \pi) = F_1(\cdot, \pi_1) \otimes F_2(\cdot, \pi_2)$, *where* $\pi = (\pi_1, \pi_2)$.

(a) *If* F_i *is regular on* Π_i *for* $i = 1,2$, *then so is* F *on* Π.
(b) *If* F_i *is I-regular on compact* Π_i *for* $i = 1,2$, *then so is* F *on* Π.
(c) *If* F_i *is H-regular on compact* Π_i *with asymptotic order* $\kappa(\cdot, \pi) = \kappa_1(\cdot, \pi_1) \otimes \kappa_2(\cdot, \pi_2)$ *and limit homogeneous function* H_i *for* $i = 1,2$, *then so is* F *on* Π *with asymptotic order* $\kappa(\cdot, \pi) = \kappa_1(\cdot, \pi_1) \otimes \kappa_2(\cdot, \pi_2)$ *and limit homogeneous function* $H(\cdot, \pi) = H_1(\cdot, \pi_1) \otimes H_2(\cdot, \pi_2)$.

LEMMA A7: (a) *Let Assumption* 2.1 *hold. If* F *is regular on a compact set* Π, *then for large* n, $n^{-1}\sum_{t=1}^n F(x_t/\sqrt{n}, \pi)u_t = o_p(1)$ *uniformly in* $\pi \in \Pi$.
(b) *Let Assumption* 2.2 *hold. If* F *is I-regular on a compact set* Π, *then for large* n $n^{-1/2}\sum_{t=1}^n F(x_t, \pi)u_t = o_p(1)$ *uniformly in* $\pi \in \Pi$.
(c) *Let Assumption* 2.1 *hold. If* F *is H-regular on a compact set* Π, *then for large* n $n^{-1}\kappa(\sqrt{n}, \pi)\sum_{t=1}^n F(x_t, \pi)u_t = o_p(1)$ *uniformly in* $\pi \in \Pi$.

LEMMA A8: (a) *If* F *is regular on a compact set* Π, *then* $\int_0^1 F(V(r), \cdot)\, dr$ *is continuous a.s. on* Π.
(b) *If* F *is I-regular on a compact set* Π, $\int_{-\infty}^{\infty} F(s, \cdot)\, ds$ *is continuous on* Π.

Proofs

PROOF OF LEMMA A1: It is obvious that $T_1 \pm T_2$ and T_1T_2 satisfy regularity condition (a), if T_1 and T_2 do. To show that they also satisfy regularity condition (b), let $K \subset \mathbf{R}$ be compact, and for each $\varepsilon > 0$, let $\overline{T}_{i\varepsilon}$, $\underline{T}_{i\varepsilon}$, and $\delta_{i\varepsilon} > 0$ be given accordingly by regularity condition (b) for T_i, $i = 1,2$. For each of $T = T_1 + T_2$ and $T_1 - T_2$, we set

$$\underline{T}_\varepsilon = \underline{T}_{1\varepsilon} + \underline{T}_{2\varepsilon}, \underline{T}_{1\varepsilon} - \overline{T}_{2\varepsilon},$$

$$\overline{T}_\varepsilon = \overline{T}_{1\varepsilon} + \overline{T}_{2\varepsilon}, \overline{T}_{1\varepsilon} - \underline{T}_{2\varepsilon},$$

and $\delta_\varepsilon = \min(\delta_{1\varepsilon}, \delta_{2\varepsilon})$. It is obvious that $\underline{T}_\varepsilon$ and $\overline{T}_\varepsilon$ are continuous, $\underline{T}_\varepsilon(x) \le T(y) \le \overline{T}_\varepsilon(x)$ for all $|x - y| < \delta_\varepsilon$ on K, and $\int_K (\overline{T}_\varepsilon - \underline{T}_\varepsilon)(x)\, dx \to 0$, as $\varepsilon \to 0$, as required for the regularity of $T = T_1 \pm T_2$.

For $T = T_1T_2$, it suffices to look at the case where $T_1, T_2 \ge 0$. Given the above result, we write $T_i = T_i^+ - T_i^-$, where T_i^+ and T_i^- are the positive and negative parts of T_i, $i = 1,2$, respectively, and then consider each product term separately. To show that regularity condition (b) holds for T, let

$$\underline{T}_\varepsilon = \underline{T}_{1\varepsilon}\underline{T}_{2\varepsilon} \quad \text{and} \quad \overline{T}_\varepsilon = \overline{T}_{1\varepsilon}\overline{T}_{2\varepsilon}$$

and $\delta_\varepsilon = \min(\delta_{1\varepsilon}, \delta_{2\varepsilon})$ for each $\varepsilon > 0$. Clearly, $\underline{T}_\varepsilon$ and $\overline{T}_\varepsilon$ are continuous, and $\underline{T}_\varepsilon(x) \le T(y) \le \overline{T}_\varepsilon(x)$ for all $|x - y| < \delta_\varepsilon$ on K. Moreover, since $\underline{T}_{i\varepsilon}$ and $\overline{T}_{i\varepsilon}$, $i = 1,2$, are bounded on K, $\int_K (\overline{T}_\varepsilon - \underline{T}_\varepsilon)(x)\, dx \to 0$, as $\varepsilon \to 0$. This completes the proof. *Q.E.D.*

PROOF OF LEMMA A2: The first part is due to Park and Phillips (1999). To prove the second part, first let $s_{\max} = \max_{0 \le r \le 1} V(r)$ and $s_{\min} = \min_{0 \le r \le 1} V(r)$, and define $s_m = \max(s_{\max}, -s_{\min}) + 1$. Since $s_m < \infty$ a.s., we have $\mathbf{P}\{s_m > c\} \to 0$ as $c \to \infty$. Therefore, we may take $c > 0$ large so that $\mathbf{P}\{s_m > c\}$ is arbitrarily small. Fix $c > 0$ large, and define $K = [-c, c]$. We then denote by $\underline{T}_\varepsilon$ and $\overline{T}_\varepsilon$ the functions given for each $\varepsilon > 0$ by regularity condition (b) on the compact set K. In view of regularity condition (a), we may find $\underline{T}_\varepsilon$ and $\overline{T}_\varepsilon$ such that they are continuous on $\mathbf{R}$, and $\overline{T}_\varepsilon - \underline{T}_\varepsilon$ is bounded.

Let $T_\varepsilon = \underline{T}_\varepsilon$ or $\overline{T}_\varepsilon$. Since T_ε is continuous, $T_\varepsilon(V_n) \to_{a.s.} T_\varepsilon(V)$. Therefore, by Kurtz and Protter (1991),

$$\int_0^1 T_\varepsilon(V_n)\, dU_n \to_d \int_0^1 T_\varepsilon(V)\, dU \tag{25}$$

as $n \to \infty$. It therefore suffices to show that as $\varepsilon \to 0$

$$\left| \int_0^1 T(V_n)\, dU_n - \int_0^1 T_\varepsilon(V_n)\, dU_n \right| \to_p 0, \tag{26}$$

uniformly for all large n including $n = \infty$, in which case by convention V_n and U_n reduce to V and U respectively.

Define

$$A_{n\varepsilon} = \int_0^1 (\overline{T}_\varepsilon(V_n) - \underline{T}_\varepsilon(V_n))^2 1\{|V_n| \le c\},$$

$$B_{n\varepsilon} = \int_0^1 (\overline{T}_\varepsilon(V_n) - \underline{T}_\varepsilon(V_n))^2 1\{|V_n| > c\}.$$

Then we have

$$\mathbf{E}\left(\int_0^1 (T(V_n) - T_\varepsilon(V_n))\, dU_n \right)^2 \le \sigma^2 \mathbf{E}\left(\int_0^1 (\overline{T}_\varepsilon(V_n) - \underline{T}_\varepsilon(V_n))^2 \right)$$
$$= \sigma^2 \mathbf{E}(A_{n\varepsilon} + B_{n\varepsilon}).$$

The result in (26) will therefore follow if we show that $\mathbf{E}A_{n\varepsilon}$ and $\mathbf{E}B_{n\varepsilon}$ can be made arbitrarily small for all large n by choosing $\varepsilon > 0$ sufficiently small.

Let $D_\varepsilon(x) = (\overline{T}_\varepsilon(x) - \underline{T}_\varepsilon(x))^2 1\{|x| \le c\}$. Since D_ε is regular, we have by the result in the first part of the lemma that

$$A_{n\varepsilon} = \int_0^1 D_\varepsilon(V_n) \to_{a.s.} \int_0^1 D_\varepsilon(V) := A_\varepsilon.$$

Moreover,

$$A_\varepsilon = \int_K (\overline{T}_\varepsilon(s) - \underline{T}_\varepsilon(s))^2 L(1,s)\, ds$$
$$\le \|\overline{T}_\varepsilon - \underline{T}_\varepsilon\| \left(\sup_{s \in K} L(1,s) \right) \int_K (\overline{T}_\varepsilon(s) - \underline{T}_\varepsilon(s))\, ds \to_{a.s.} 0,$$

as $\varepsilon \to 0$. It now follows that $\mathbf{E}A_{n\varepsilon} \to \mathbf{E}A_\varepsilon$ as $n \to \infty$, and $\mathbf{E}A_\varepsilon \to 0$ as $\varepsilon \to 0$, since $A_{n\varepsilon}$ and A_ε are all bounded. Consequently, $\mathbf{E}A_{n\varepsilon}$ can be made small for all large n by choosing $\varepsilon > 0$ appropriately.

Finally, notice that

$$B_{n\varepsilon} \le \|\overline{T}_\varepsilon - \underline{T}_\varepsilon\|^2 1\{s_m > c\}$$

for all large n including $n = \infty$, and therefore,

$$\mathbf{E}B_{n\varepsilon} \le \|\overline{T}_\varepsilon - \underline{T}_\varepsilon\|^2 \Pr\{s_m > c\},$$

which, as we noted earlier, can be made arbitrarily small by taking c large. We now have (26), which along with (25), completes the proof. *Q.E.D.*

Proof of Lemma A3: For part (a), let $\pi_0 \in \Pi$ and a compact set $K \subset \mathbf{R}$ be given. Since $F(\cdot, \pi_0)$ is regular, there exist for each $\varepsilon > 0$ continuous $\underline{T}_\varepsilon^0$, $\overline{T}_\varepsilon^0$, and $\delta_\varepsilon > 0$ satisfying

$$\underline{T}_\varepsilon^0(x) \le F(y, \pi_0) \le \overline{T}_\varepsilon^0(x)$$

for all $|x - y| < \delta_\varepsilon$ on K, and $\int_K (\overline{T}_\varepsilon^0 - \underline{T}_\varepsilon^0)(x)\, dx \to 0$ as $\varepsilon \to 0$. However, due to the equicontinuity of $F(x, \cdot)$, there is a neighborhood N_0 of π_0 such that

$$\underline{T}_\varepsilon^0(x) - \varepsilon \le F(y, \pi) \le \overline{T}_\varepsilon^0(x) + \varepsilon$$

for all $\pi \in N_0$ and all $|x - y| < \delta_\varepsilon$ on K.

We now let

$$\underline{T}_\varepsilon(x) = \underline{T}_\varepsilon^0(x) - \varepsilon \qquad \text{and} \qquad \overline{T}_\varepsilon(x) = \overline{T}_\varepsilon^0(x) + \varepsilon.$$

It is easy to see that $\underline{T}_\varepsilon$ and $\overline{T}_\varepsilon$ are continuous, and for all $N \subset N_0$,

$$\underline{T}_\varepsilon(x) \le \sup_{\pi \in N} F(y, \pi), \qquad \inf_{\pi \in N} F(y, \pi) \le \overline{T}_\varepsilon(x)$$

for all $|x - y| < \delta_\varepsilon$ on K, and finally, $\int_K (\overline{T}_\varepsilon - \underline{T}_\varepsilon)(x)\,dx \to 0$ as $\varepsilon \to 0$ since K is bounded.

To prove part (b), use the result in part (a) to deduce that for every $\pi_0 \in \Pi$ there exists a neighborhood N_0 such that $\sup_{\pi \in N_0} F(\cdot, \pi)$ and $\inf_{\pi \in N_0} F(\cdot, \pi)$ are regular, and therefore locally bounded. The local boundedness of $\sup_{\pi \in \Pi} |F(\cdot, \pi)|$ now follows directly from the compactness of Π. *Q.E.D.*

Proof of Lemma A4: Let $K = [s_{\min} - 1, s_{\max} + 1]$. Part (a) is trivial, because

$$\left| \frac{1}{n} \sum_{t=1}^n T\left(\frac{x_t}{\sqrt{n}} \right) \right| \le \|T\|_K < \infty \qquad \text{a.s.}$$

for large n. Part (c) is also immediate since

$$\mathbf{E}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^n T\left(\frac{x_t}{\sqrt{n}} \right) u_t \right)^2 = \sigma^2 \mathbf{E}\left(\frac{1}{n} \sum_{t=1}^n T\left(\frac{x_t}{\sqrt{n}} \right)^2 \right) \le \sigma^2 \mathbf{E}\|T^2\|_K,$$

which is finite, due to that $T \in \mathscr{T}_{LB}^0$.

For the proofs of parts (b) and (d), recall that for all $T \in \mathscr{T}_{LB}^0$

$$\frac{1}{n} \sum_{t=1}^n T(x_t) \to_{a.s.} 0,$$

as $n \to \infty$. This is shown in Park and Phillips (1999). To show part (b), we note that

$$\left| \frac{1}{n} \sum_{t=1}^n T_1\left(\frac{x_t}{\sqrt{n}} \right) T_2(x_t) \right| \le \|T_1\|_K \frac{1}{n} \sum_{t=1}^n |T_2(x_t)| \to_{a.s.} 0$$

since $T_1 \in \mathscr{T}_{LB}$ and $T_2 \in \mathscr{T}_{LB}^0$. Finally, we observe for the proof of part (d) that

$$\left| \frac{1}{n} \sum_{t=1}^n T_1\left(\frac{x_t}{\sqrt{n}} \right)^2 T_2(x_t)^2 \right| \le \|T_1^2\|_K \frac{1}{n} \sum_{t=1}^n T_2(x_t)^2 \to_{a.s.} 0$$

since $T_1 \in \mathscr{T}_{LB}$ and $T_2 \in \mathscr{T}_{LB}^0$. Moreover, we note that

$$\left| \frac{1}{n} \sum_{t=1}^n T_1\left(\frac{x_t}{\sqrt{n}} \right)^2 T_2(x_t)^2 \right| \le \|T_1^2\|_K \|T_2^2\|$$

and, since $T_1 \in \mathscr{T}_{LB}^0$, $\mathbf{E}\|T_1^2\|_K < \infty$. It therefore follows by dominated convergence that

$$\mathbf{E}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^n T_1\left(\frac{x_t}{\sqrt{n}} \right) T_2(x_t) u_t \right)^2 = \sigma^2 \mathbf{E}\left(\frac{1}{n} \sum_{t=1}^n T_1 \right) \left(\frac{x_t}{\sqrt{n}} \right)^2 T_2(x_t)^2 \to 0,$$

which completes the proof. *Q.E.D.*

Proof of Lemma A5: Part (a) follows directly from Definition 3.4. Note that both $\mathscr{T}_B^0$ and $\mathscr{T}_{LB}^0$ are closed under multiplication. To prove the first two results in part (b), let

$$T(x) = \left\| \sup_{\omega \in \Phi} A(x, \omega) \right\| \qquad \text{and} \qquad S(x) = \left\| \sup_{\omega \in \Phi} B(x, \omega) \right\|,$$

and write $a_n(\omega) = a(\sqrt{n}, \omega)$ and $b_n(\omega) = b(\sqrt{n}, \omega)$. We have

$$\left\| \frac{1}{n} z_n(\omega)^{-1} \sum_{t=1}^{n} Z_{nt}(\omega) \right\| \leq \|z_n(\omega)^{-1} a_n(\omega)\| \left(\frac{1}{n} \sum_{t=1}^{n} T\left(\frac{x_t}{\sqrt{n}} \right) \right)$$

$$\text{or } \|z_n(\omega)^{-1} b_n(\omega)\| \left(\frac{1}{n} \sum_{t=1}^{n} T\left(\frac{x_t}{\sqrt{n}} \right) S(x_t) \right),$$

and then the first result follows immediately from Lemma A4. To deduce the second result, note that

$$\left\| \frac{1}{n} z_n(\omega)^{-1} \sum_{t=1}^{n} Z_{nt}(\omega) u_t \right\| \leq \|z_n(\omega)^{-1} a_n(\omega)\| \left(\frac{1}{n} \sum_{t=1}^{n} T\left(\frac{x_t}{\sqrt{n}} \right) |u_t| \right)$$

$$\text{or } \|z_n(\omega)^{-1} b_n(\omega)\| \left(\frac{1}{n} \sum_{t=1}^{n} T\left(\frac{x_t}{\sqrt{n}} \right) S(x_t) |u_t| \right),$$

and subsequently observe that

$$\frac{1}{n} \sum_{t=1}^{n} T\left(\frac{x_t}{\sqrt{n}} \right) |u_t| \leq \left(\frac{1}{n} \sum_{t=1}^{n} T\left(\frac{x_t}{\sqrt{n}} \right)^2 \right)^{1/2} \left(\frac{1}{n} \sum_{t=1}^{n} u_t^2 \right)^{1/2} = O_p(1),$$

$$\frac{1}{n} \sum_{t=1}^{n} T\left(\frac{x_t}{\sqrt{n}} \right) S(x_t) |u_t| \leq \left(\frac{1}{n} \sum_{t=1}^{n} T\left(\frac{x_t}{\sqrt{n}} \right)^2 S(x_t)^2 \right)^{1/2} \left(\frac{1}{n} \sum_{t=1}^{n} u_t^2 \right)^{1/2} = o_p(1),$$

due to Lemma A4, since $T^2 \in \mathscr{T}_{LB}^0$ and $S^2 \in \mathscr{T}_B^0$. The third result in part (b) is an immediate consequence of Lemma A4. *Q.E.D.*

PROOF OF LEMMA A6: For the proof of part (a), assume that F_1 and F_2 are regular. It follows immediately from Lemma A2 that regularity condition (a) holds for F, since it is the product of two regular functions F_1 and F_2. To show that F satisfies regularity condition (b), we fix x_0 and $\pi_0 = (\pi_1^0, \pi_2^0)$ arbitrarily, and let $\varepsilon > 0$ be given. Due to the regularity of F_i, there exists $\delta > 0$ such that $\|F_i(x, \pi_i) - F_i(x, \pi_i^0)\| < \varepsilon$ for all $|x - x_0| < \delta$ and $\|\pi_i - \pi_0\| < \delta$. We therefore have

$$\begin{aligned} \|F(x, \pi) - F(x, \pi_0)\| &\leq \|F_1(x, \pi_1) - F_1(x, \pi_1^0)\| \, \|F_2(x, \pi_2)\| \\ &\quad + \|F_1(x, \pi_1^0)\| \, \|F_2(x, \pi_2) - F_2(x, \pi_2^0)\| \\ &\leq \varepsilon \max(\|F_1(x, \pi_1^0)\|, \|F_2(x, \pi_2^0)\| + \varepsilon) \end{aligned}$$

for all $|x - x_0| < \delta$ and $\|\pi - \pi_0\| < \delta$. This establishes regularity condition (b) for F.

To prove part (b), let F_1 and F_2 be I-regular. To show that I-regularity condition (a) is satisfied for F, we choose arbitrary π_1^0 and π_2^0. Since the F_i are I-regular, there exist neighborhoods N_i^0 of π_i^0 and bounded and integrable T_i such that

$$\|F_i(x, \pi_i) - F_i(x, \pi_i^0)\| \leq \|\pi_i - \pi_i^0\| T_i(x)$$

for all $\pi_i \in N_i^0$. Therefore, if we let $\pi_0 = (\pi_1^0, \pi_2^0)$, then it follows for all $\pi \in N_0 = N_1^0 \times N_2^0$ that

$$\begin{aligned} &\|F(x, \pi) - F(x, \pi_0)\| \\ &\quad \leq \|F_1(x, \pi_1) - F_1(x, \pi_1^0)\| S_2(x) + S_1(x) \|F_2(x, \pi_2) - F_2(x, \pi_2^0)\| \\ &\quad \leq \|\pi_1 - \pi_1^0\| T_1(x) S_2(x) + \|\pi_2 - \pi_2^0\| S_1(x) T_2(x) \\ &\quad \leq \|\pi - \pi_0\| T(x), \end{aligned}$$

where we set $S_i(x) = \sup_{\pi_i \in \Pi} \|F_i(x, \pi_i)\|$ and $T = \max(T_1, T_2, S_1, S_2)$. Note that S_i are bounded and integrable, since Π_i are assumed to be compact. Therefore, T is bounded and integrable. Finally, we let

$$\|F_i(x, \pi_i) - F_i(y, \pi_i)\| \leq c_i |x - y|^k,$$

and $a = \max(\|F_1\|, \|F_2\|)$ and $b = \max(c_1, c_2)$. Then it follows immediately that

$$\begin{aligned}
&\|F(x, \pi) - F(y, \pi)\| \\
&\quad \leq \|F_1(x, \pi_1) \otimes F_2(x, \pi_2) - F_1(y, \pi_1) \otimes F_2(x, \pi_2)\| \\
&\qquad + \|F_1(y, \pi_1) \otimes F_2(x, \pi_2) - F_1(y, \pi_1) \otimes F_2(y, \pi_2)\| \\
&\quad \leq a(\|F_1(x, \pi_1) - F_1(y, \pi_1)\| + \|F_2(x, \pi_2) - F_2(y, \pi_2)\|) \\
&\quad \leq ab|x - y|^k,
\end{aligned}$$

which proves that I-regularity condition (b) also holds for F.

For part (c), let

$$F_i(\lambda x, \pi_i) = \kappa_i(\lambda, \pi_i) H_i(x, \pi_i) + R_i(x, \lambda, \pi_i),$$

for $i = 1, 2$, and define

$$\kappa(\lambda, \pi) = \kappa_1(\lambda, \pi_1) \otimes \kappa_2(\lambda, \pi_2) \qquad \text{and} \qquad H(x, \pi) = H_1(x, \pi_1) \otimes H_2(x, \pi_2).$$

As shown in part (a), H is regular, and the H-regularity condition (a) is satisfied. Moreover, if we write

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

then the residual function R becomes

$$\begin{aligned}
R(x, \lambda, \pi) = {} & R_1(x, \lambda, \pi_1) \otimes R_2(x, \lambda, \pi_2) \\
& + \kappa_1(\lambda, \pi_1) H_1(x, \pi_1) \otimes R_2(x, \lambda, \pi_2) \\
& + \kappa_2(\lambda, \pi_2) H_2(x, \pi_2) \otimes R_1(x, \lambda, \pi_1).
\end{aligned}$$

It therefore follows immediately from Lemma A5(a) that $R(x, \lambda, \pi)$ is of order smaller than $\kappa(\lambda, \pi)$ on Π, and so H-regularity condition (b) is also met. This completes the proof for part (c). *Q.E.D.*

PROOF OF LEMMA A7: In what follows, we assume w.l.o.g. that F is real-valued by taking each component separately. For part (a), let $\pi_0 \in \Pi$ be chosen arbitrarily. We show that

$$\sup_{\pi \in N_0} \left| \frac{1}{n} \sum_{t=1}^{n} F\left(\frac{x_t}{\sqrt{n}}, \pi \right) u_t \right| = o_p(1), \tag{27}$$

for some neighborhood N_0 of π_0, from which the stated result follows immediately because of the compactness of Π. From Theorem 3.1

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} F\left(\frac{x_t}{\sqrt{n}}, \pi_0 \right) u_t = O_p(1), \tag{28}$$

so it suffices to show that

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

can be made arbitrarily small a.s. uniformly in $\pi \in N_0$, which we now set out to do.

Using Cauchy–Schwarz we have

$$\left|\frac{1}{n}\sum_{t=1}^{n}\left(F\left(\frac{x_t}{\sqrt{n}},\pi\right)-F\left(\frac{x_t}{\sqrt{n}},\pi_0\right)\right)u_t\right| \le \left(\frac{1}{n}\sum_{t=1}^{n}\left(F\left(\frac{x_t}{\sqrt{n}},\pi\right)-F\left(\frac{x_t}{\sqrt{n}},\pi_0\right)\right)^2\right)^{1/2}\left(\frac{1}{n}\sum_{t=1}^{n}u_t^2\right)^{1/2}. \tag{29}$$

However, it follows from Lemma A6(a) and Theorem 3.1 that

$$\frac{1}{n}\sum_{t=1}^{n}\left(F\left(\frac{x_t}{\sqrt{n}},\pi\right)-F\left(\frac{x_t}{\sqrt{n}},\pi_0\right)\right)^2 \to_{a.s.} \int_0^1 (F(V(r),\pi)-F(V(r),\pi_0))^2\,dr,$$

uniformly in $\pi \in \Pi$. Let N_δ be the δ-neighborhood of π_0. Then, for any $x \in \mathbf{R}$

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

as $\delta \to 0$, due to the continuity of $F(x,\cdot)$. Since $\sup_{\pi\in\Pi}|F(\cdot,\pi)|$ is locally integrable as shown in Lemma A3(b), we may invoke dominated convergence to get

$$\int_0^1 (F(V(r),\pi)-F(V(r),\pi_0))^2\,dr = \int_{-\infty}^{\infty}(F(s,\pi)-F(s,\pi_0))^2 L(1,s)\,ds \to_{a.s.} 0$$

uniformly on N_δ, as $\delta \to 0$. It therefore follows from (34) that there exists a neighborhood N_0 of π_0 such that

$$\sup_{\pi\in N_0}\left|\frac{1}{n}\sum_{t=1}^{n}\left(F\left(\frac{x_t}{\sqrt{n}},\pi\right)-F\left(\frac{x_t}{\sqrt{n}},\pi_0\right)\right)u_t\right| < \varepsilon \qquad \text{a.s.} \tag{30}$$

for any $\varepsilon > 0$ given. We may now easily deduce (27) from the results in (28) and (30). The proof for part (a) is therefore complete.

We now prove part (b). As in the proof of part (a), we fix an arbitrary $\pi_0 \in \Pi$. Due to the compactness of Π, it suffices to show that there exists a neighborhood N_0 of π_0 for which

$$\sup_{\pi\in N_0}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}F(x_t,\pi)u_t\right| = o_p(1). \tag{31}$$

Since it follows from Theorem 3.2 that

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

it suffices to show that

$$\sup_{\pi\in N_0}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(F(x_t,\pi)-F(x_t,\pi_0))u_t\right| = o_p(1) \tag{32}$$

to deduce (31).

However, we have by I-regularity condition (a) that

$$(33)\qquad \sum_{t=1}^{n} |F(x_t,\pi)-F(x_t,\pi_0)|\,|u_t| \le \|\pi-\pi_0\| \left(\sigma \sum_{t=1}^{n} |T(x_t)| + \sum_{t=1}^{n} |T(x_t)| w_t \right),$$

where $w_t = |u_t| - \mathbf{E}(|u_t| \| \mathscr{F}_{t-1})$. Note that $\mathbf{E}(|u_t| \| \mathscr{F}_{t-1})^2 \le \sigma^2$ by Jensen's inequality. Since T is bounded and integrable, we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} |T(x_t)| = O_p(1),$$

and

$$\mathbf{E}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} |T(x_t)| w_t \right)^2 \le \sigma^2 \mathbf{E}\left(\frac{1}{n} \sum_{t=1}^{n} T(x_t)^2 \right) \to 0.$$

It is therefore clear from (33) that we may choose N_0 such that (32) holds, which completes the proof.

For the proof of part (c), note that

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

due to Theorem 3.3. Moreover, by Cauchy–Schwarz,

$$\left| \frac{1}{n\kappa(\sqrt{n})} \sum_{t=1}^{n} (F(x_t,\pi)-F(x_t,\pi_0)) u_t \right| \le \left(\frac{1}{n\kappa(\sqrt{n})^2} \sum_{t=1}^{n} (F(x_t,\pi)-F(x_t,\pi_0))^2 \right)^{1/2} \left(\frac{1}{n} \sum_{t=1}^{n} u_t^2 \right)^{1/2},$$

and we have from Lemma A6(c) and Theorem 3.3 that

$$\frac{1}{n\kappa(\sqrt{n})^2} \sum_{t=1}^{n} (F(x_t,\pi)-F(x_t,\pi_0))^2 \to_{a.s.} \int_0^1 (H(V(r),\pi)-H(V(r),\pi_0))^2\, dr$$

uniformly in $\pi \in \Pi$. We may now use the same argument as that in the proof of part (a) to get the stated result. *Q.E.D.*

PROOF OF LEMMA A8: For the proof of part (a), it suffices to show that

$$(34)\qquad \int_{-\infty}^{\infty} F(s,\cdot) L(1,s)\, ds$$

is continuous a.s., due to the occupation time formula. The continuity of (34), however, is an immediate consequence of dominated convergence, and follows immediately from the a.s. integrability of $\sup_{\pi\in\Pi} |F(\cdot,\pi)| L(1,\cdot)$. Note that $\sup_{\pi\in\Pi} |F(\cdot,\pi)|$ is locally bounded, as shown in Lemma A3(b), and hence locally integrable, and $L(1,\cdot)$ has compact support a.s. We may also easily deduce part (b) from dominated convergence, due to the continuity of $F(x,\cdot)$ for all $x \in \mathbf{R}$, and the integrability of $\sup_{\pi\in\Pi} |F(\cdot,\pi)|$. *Q.E.D.*

APPENDIX B: PROOFS OF THE MAIN RESULTS

PROOF OF LEMMA 2.1: Let $(U_n, V_n) \to_d (U, V)$, as given by condition (a) of Assumption 2.1, and denote by $(\Omega, \mathscr{F}, \Pr)$ the probability space where (U,V) is defined. For each n, we construct a sequence of stopping times $(\tau_{nt})_{t=0}^{n}$ and random variables $(V_{nt})_{t=0}^{n}$ on $(\Omega, \mathscr{F}, \Pr)$, from which the desired (U_n^0, V_n^0) is then defined. In the subsequent construction, we let

$$\mathscr{F}_{nt}^0 = \sigma\left(\left(U(r), r \le \frac{\tau_{nt}}{n}\right), (V_{ni})_{i=0}^t\right), \qquad \mathscr{F}_{nt} = \sigma((U_n(n_i), V_n(n_i))_{i=0}^t),$$

$$\mathscr{G}_{nt}^0 = \sigma\left(\left(U(r), r \le \frac{\tau_{nt}}{n}\right), (V_{ni})_{i=0}^{t-1}\right), \qquad \mathscr{G}_{nt} = \sigma((U_n(n_i))_{i=0}^t (V_n(n_i))_{i=0}^{t-1}),$$

where $n_i = i/n$ for $0 \le i \le n$. Also, the symbolism '$\cdot|\mathscr{F}$' is used to signify 'distribution conditional on the σ-field $\mathscr{F}$'.

Let n be given and fixed. First, we choose any random variable V_{n0} on $(\Omega, \mathscr{F}, \mathbf{P})$, which has the same distribution as $V_n(0)$, i.e., $V_{n0} =_d V_n(0)$. Second, let π_{n1} be a stopping time defined on $(\Omega, \mathscr{F}, \mathbf{P})$, for which $U(\tau_{n1}/n)|\mathscr{F}_{n0}^0 =_d U_n(1/n)|\mathscr{F}_{n0}$. Such a stopping time exists, as shown in Hall and Heyde (1980, Theorem A1). We then define a random variable on $(\Omega, \mathscr{F}, \mathbf{P})$, denoted by V_{n1}, such that $V_{n1}|\mathscr{G}_{n1}^0 =_d V_n(1/n)|\mathscr{G}_{n1}$, and so on. It is obvious that we may proceed to find $(\tau_{nt})_{t=0}^n$ and $(V_{nt})_{t=0}^n$ on $(\Omega, \mathscr{F}, \mathbf{P})$ successively so that

$$U\left(\frac{\tau_{nt}}{n}\right)\bigg|\mathscr{F}_{n,t-1}^0 =_d U_n\left(\frac{t}{n}\right)\bigg|\mathscr{F}_{n,t-1} \qquad \text{and} \qquad V_{nt}\bigg|\mathscr{G}_{nt}^0 =_d V_n\left(\frac{t}{n}\right)\bigg|\mathscr{G}_{nt}$$

in a zig-zag fashion. If we let $(\tau_{nt})_{t=0}^n$ and $(V_{nt})_{t=0}^n$ be constructed in this way, and define

$$U_n^0(r) = U\left(\frac{\tau_{n[nr]}}{n}\right) \qquad \text{and} \qquad V_n^0(r) = V_{n[nr]},$$

it follows immediately that $(U_n, V_n) =_d (U_n^0, V_n^0)$. Such processes (U_n^0, V_n^0) can, of course, be found for all n.

It is shown by Park and Phillips (1999) that we may choose the stopping times τ_{nt} so that they satisfy condition (13). In particular, it follows from the Hölder continuity of the sample path of U that

$$|U_n^0(r) - U(r)| \le c\left|\frac{\tau_{[nr]}}{n} - r\right|^{1/2-\varepsilon}$$

a.s., for some constant c and any $\varepsilon > 0$. Now we may easily deduce from (13) that

$$\sup_{r\in[0,1]} |U_n^0(r) - U(r)| = o(n^{(-1+\delta)/2+\varepsilon}) \qquad \text{a.s.}$$

for $\max(1/2, 2/q) < \delta < 1$ and any $\varepsilon > 0$. In particular, $U_n^0 \to_{a.s.} U$ uniformly on $[0,1]$. Moreover, since $(U_n^0, V_n^0) \to_d (U, V)$, we may redefine V_n^0, if necessary, so that the distribution of (U_n^0, V_n^0) is unchanged and $V_n^0 \to_{a.s.} V$ uniformly on $[0,1]$. This is possible due to the representation theorem of a weakly convergent sequence of probability measures by an almost sure convergent sequence—e.g. see Pollard (1984, pp. 71–72). *Q.E.D.*

Proof of Lemma 2.2: See Corollary 1.6, p. 215, of Revuz and Yor (1994). *Q.E.D.*

Proof of Theorem 3.1: For sample mean asymptotics, we write

$$\frac{1}{n}\sum_{t=1}^n F\left(\frac{x_t}{\sqrt{n}}, \pi\right) = \int_0^1 F(V_n(r), \pi)\, dr,$$

and show that

$$\int_0^1 F(V_n(r), \pi)\, dr \to_{a.s.} \int_0^1 F(V(r), \pi)\, dr, \tag{35}$$

uniformly in $\pi \in \Pi$. Fix an arbitrary $\pi_0 \in \Pi$. Due to Lemma A3(a), there exists a neighborhood N_0

of π_0 such that $\sup_{\pi \in N} F(\cdot,\pi)$ and $\inf_{\pi \in N} F(\cdot,\pi)$ are regular for any neighborhood $N \subset N_0$ of π_0. Therefore, it follows from Lemma A2 that

$$\int_0^1 \sup_{\pi\in N} F(V_n(r),\pi)\,dr \to_{a.s.} \int_0^1 \sup_{\pi\in N} F(V(r),\pi)\,dr, \tag{36}$$

$$\int_0^1 \inf_{\pi\in N} F(V_n(r),\pi)\,dr \to_{a.s.} \int_0^1 \inf_{\pi\in N} F(V(r),\pi)\,dr, \tag{37}$$

as $n\to\infty$.

Let $N_\delta \subset N_0$ be the δ-neighborhood of π_0. Then we have

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

as $\delta\to 0$, due to the continuity of $F(x,\cdot)$. Moreover, as shown in Lemma A3(b), $\sup_{\pi\in\Pi}|F(\cdot,\pi)|$ is locally bounded. It therefore follows from the occupation time formula and dominated convergence that

$$\begin{aligned}\int_0^1\left(\sup_{\pi\in N} F(V(r),\pi) - \inf_{\pi\in N} F(V(r),\pi)\right)dr \\ = \int_{-\infty}^{\infty}\left(\sup_{\pi\in N_\delta} F(s,\pi) - \inf_{\pi\in N_\delta} H(s,\pi)\right)L(1,s)\,ds \to_{a.s.} 0,\end{aligned} \tag{38}$$

as $\delta\to 0$. We may now easily deduce from (36)–(38) that there exists a neighborhood of π_0 where (35) holds uniformly in π. Since π_0 was chosen arbitrary and Π is compact, (35) holds uniformly on Π, as was to be shown. The sample covariance asymptotics are given in Lemma A2. *Q.E.D.*

Proof of Theorem 3.2: Fix $\pi_0\in\Pi$. For any neighborhood N of π_0, we have by I-regularity condition (b)

$$\begin{aligned}&\left|\inf_{\pi\in N} F(x,\pi) - \inf_{\pi\in N} F(y,\pi)\right|, \left|\sup_{\pi\in N} F(x,\pi) - \sup_{\pi\in N} F(y,\pi)\right| \\ &\quad\le \sup_{\pi\in N}|F(x,\pi)-F(y,\pi)| \\ &\quad\le c|x-y|^k.\end{aligned}$$

It follows that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n \sup_{\pi\in N} F(x_t,\pi) \to_p \left(\int_{-\infty}^{\infty}\sup_{\pi\in N} F(s,\pi)\,ds\right)L(1,0), \tag{39}$$

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n \inf_{\pi\in N} F(x_t,\pi) \to_p \left(\int_{-\infty}^{\infty}\inf_{\pi\in N} F(s,\pi)\,ds\right)L(1,0), \tag{40}$$

due to Theorem 5.1 of Park and Phillips (1999).

Let N_δ be the δ-neighborhood of π_0. By I-regularity condition (a), we have

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

as $\delta\to 0$, and, in view of I-regularity condition (a) and dominated convergence,

$$\int_{-\infty}^{\infty}\sup_{\pi\in N_\delta} F(s,\pi)\,ds - \int_{-\infty}^{\infty}\inf_{\pi\in N_\delta} F(s,\pi)\,ds \to 0, \tag{41}$$

as $\delta \to 0$. We may now easily deduce from (39)–(41) that there exists a neighborhood of π_0 such that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} F(x_t, \pi) \to_p \left(\int_{-\infty}^{\infty} F(s, \pi)\, ds \right) L(1,0),$$

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

We now prove the result on sample covariance asymptotics. For notational simplicity, we assume that F is real-valued. The proof for a vector F follows by considering an arbitrary linear combination of the components of F. Define

$$(42) \qquad M_n(r) = \sqrt[4]{n} \sum_{t=1}^{k-1} F\left(\sqrt{n}\, V_n\left(\frac{t-1}{n} \right), \pi \right) \left(U\left(\frac{\tau_{nt}}{n} \right) - U\left(\frac{\tau_{n,t-1}}{n} \right) \right)$$
$$+ \sqrt[4]{n}\, F\left(\sqrt{n}\, V_n\left(\frac{k-1}{n} \right), \pi \right) \left(U(r) - U\left(\frac{\tau_{n,k-1}}{n} \right) \right),$$

for $\tau_{n,k-1}/n < r \le \tau_{nk}/n$, where τ_{nk}, $k = 1, \dots, n$, are the stopping times introduced in Lemma 2.1. One may easily see that M_n is a continuous martingale such that

$$(43) \qquad \frac{1}{\sqrt[4]{n}} \sum_{t=1}^{n} F(x_t, \pi) u_t = M_n\left(\frac{\tau_{nn}}{n} \right),$$

and that

$$(44) \qquad \sup_{1 \le t \le n} \left| \left(\frac{\tau_{nt}}{n} - \frac{\tau_{n,t-1}}{n} \right) - \frac{1}{n} \right| = o_{a.s.}(1),$$

by Lemma 2.1.

The quadratic variation process $[M_n]$ of M_n is given by

$$[M_n]_r = \sqrt{n} \sum_{t=1}^{k-1} F\left(\sqrt{n}\, V_n\left(\frac{t-1}{n} \right), \pi \right)^2 \left(\frac{\tau_{nt}}{n} - \frac{\tau_{n,t-1}}{n} \right)$$
$$+ \sqrt{n}\, F\left(\sqrt{n}\, V_n\left(\frac{k-1}{n} \right), \pi \right)^2 \left(r - \frac{\tau_{n,k-1}}{n} \right)$$
$$= \sqrt{n} \int_0^r F(\sqrt{n}\, V_n(s), \pi)^2\, ds (1 + o_{a.s.}(1)),$$

due to (44), and therefore,

$$(45) \qquad [M_n]_r \to_p \left(\int_{-\infty}^{\infty} F(s, \pi)^2\, ds \right) L(r,0),$$

uniformly in $r \in [0,1]$, from the result obtained in the first part of this theorem. Moreover, if we denote by $[M_n, V]$ the covariation process of M_n and V, then

$$[M_n, V]_r = \sqrt[4]{n} \sum_{t=1}^{k-1} F\left(\sqrt{n}\, V_n\left(\frac{t-1}{n} \right), \pi \right) \left(\frac{\tau_{nt}}{n} - \frac{\tau_{n,t-1}}{n} \right) \sigma_{uv}$$
$$+ \sqrt[4]{n}\, F\left(\sqrt{n}\, V_n\left(\frac{k-1}{n} \right), \pi \right) \left(r - \frac{\tau_{n,k-1}}{n} \right) \sigma_{uv}$$
$$= \sigma_{uv} \sqrt[4]{n} \int_0^r F(\sqrt{n}\, V_n(s), \pi)\, ds (1 + o_{a.s.}(1))$$

uniformly in $r \in [0,1]$, due to (44). However, for all $r \in [0,1]$,

$$\left| \sqrt[4]{n} \int_0^r F(\sqrt{n}\, V_n(s), \pi)\, ds \right| \leq \frac{1}{\sqrt[4]{n}} \left(\sqrt{n} \int_0^1 |F(\sqrt{n}\, V_n(s), \pi)|\, ds \right) \to_p 0,$$

as $n \to \infty$. It follows that

$$[M_n, V]_{\rho_n(r)} \to_p 0, \tag{46}$$

where $\rho_n(r) = \inf\{s \in [0,1] : [M_n]_s > r\}$ is a sequence of time changes.

The asymptotic distribution of the continuous martingale M_n in (42) is completely determined by (45) and (46), as shown in Revuz and Yor (1994, Theorem 2.3, page 496). Now define

$$W_n(r) = M_n(\rho_n(r)).$$

The process W_n is called the DDS (or Dambis, Dubins-Schwarz) Brownian motion of the continuous martingale M_n (see, for example, Revuz and Yor (1994), Theorem 1.6, page 173). It now follows that (V, W_n) converges jointly in distribution to two independent Brownian motions (V, W). Therefore,

$$\begin{aligned} M_n\left(\frac{\tau_{nn}}{n}\right) &= M_n(1) + o_p(1) \\ &\to_d W\left(L(1,0) \int_{-\infty}^{\infty} F(s,\pi)^2\, ds \right) \end{aligned}$$

which, in view of (43), completes the proof for the second part. *Q.E.D.*

PROOF OF THEOREM 3.3: Due to Lemma A5(b),

$$\frac{1}{n} \kappa(\sqrt{n}, \pi)^{-1} \sum_{t=1}^n R\left(\frac{x_t}{\sqrt{n}}, \sqrt{n}, \pi \right) \to_{a.s.} 0,$$

uniformly in $\pi \in \Pi$, and

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

for each $\pi \in \Pi$. We therefore have

$$\frac{1}{n} \kappa(\sqrt{n}, \pi)^{-1} \sum_{t=1}^n F(x_t, \pi) = \frac{1}{n} \sum_{t=1}^n H\left(\frac{x_t}{\sqrt{n}}, \pi \right) + o_{a.s.}(1),$$

uniformly in $\pi \in \Pi$, and

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

for each $\pi \in \Pi$. The stated results follow directly from the application of Theorem 3.1 to the limit homogeneous function H. *Q.E.D.*

PROOF OF THEOREM 4.1: It follows readily from Lemma A7(b) that

$$\frac{1}{\sqrt{n}} D_n(\theta, \theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^n (f(x_t, \theta) - f(x_t, \theta_0))^2 + o_p(1),$$

uniformly in $\theta \in \Theta$. The stated result now follows immediately from Lemma A6(b) and Theorem 3.2. Notice that $L(1,0) > 0$ a.s., and, therefore, $D(\cdot, \theta_0)$ has a unique minimum at θ_0 a.s. when and only when the given identification condition holds. The continuity of $D(\cdot, \theta_0)$ follows from Lemma A8(b). *Q.E.D.*

PROOF OF THEOREM 4.2: Due to Lemma A7(c), we have

$$\frac{1}{n\kappa(\sqrt{n})}\sum_{t=1}^{n}(f(x_t,\theta)-f(x_t,\theta_0))u_t=o_p(1),$$

uniformly in $\theta\in\Theta$. Furthermore, since κ is bounded away from zero by condition (a), we have

$$\frac{1}{n\kappa(\sqrt{n})^2}D_n(\theta,\theta_0)=\frac{1}{n\kappa(\sqrt{n})^2}\sum_{t=1}^{n}(f(x_t,\theta)-f(x_t,\theta_0))^2+o_p(1),$$

uniformly in $\theta\in\Theta$. The stated result follows directly from Lemma A6(c) and Theorem 3.3. Since $L(1,0)>0$ and $L(1,\cdot)$ is continuous a.s., there exists a neighborhood of zero on which $L(1,\cdot)>0$ a.s. Therefore, $D(\cdot,\theta_0)$ has unique minimum θ_0 a.s. when and only when the identification condition (b) holds. The continuity of $D(\cdot,\theta_0)$ is due to Lemma A8(a). *Q.E.D.*

PROOF OF THEOREM 4.3: Let

$$m(\sqrt{n},\theta)^2=\frac{1}{n\kappa(\sqrt{n},\theta)^2}\sum_{t=1}^{n}f(x_t,\theta)^2,$$

$$m(\theta)^2=\int_{-\infty}^{\infty}h(s,\theta)^2L(1,s)\,ds.$$

We have $m(\sqrt{n},\theta)\to_{a.s.}m(\theta)$ uniformly in $\theta\in\Theta$, by Lemma A6(c) and Theorem 3.3. It follows from Lemma A8(a) that m is continuous a.s. Also, due to condition (b), $m>0$ a.s.

Let $\delta>0$ be given, and define $\Theta_0=\{\|\theta-\theta_0\|\geq\delta\}\subset\Theta$. Fix an arbitrary $\bar{\theta}\in\Theta_0$, and let N be the neighborhood of $\bar{\theta}$ given by the condition (a). Also, set $\bar{p}=m(\bar{\theta})$ and $\bar{q}=m(\theta_0)$. For large n, we have

$$\sup_{\theta\in N}|m(\sqrt{n},\theta)-\bar{p}|<\varepsilon,$$

since $m(\sqrt{n},\theta)\to_{a.s.}m(\theta)$ uniformly in $\theta\in\Theta$ and m is continuous. Moreover, $|m(\sqrt{n},\theta_0)-\bar{q}|<\varepsilon$ for sufficiently large n. Notice that $\bar{p},\bar{q}>0$ since $m>0$. Therefore,

$$\inf_{\substack{|p-\bar{p}|<\varepsilon\\|q-\bar{q}|<\varepsilon}}\inf_{\theta\in N}|p\kappa(\lambda,\theta)-q\kappa(\lambda,\theta_0)|\leq|\kappa(\sqrt{n},\theta)m(\sqrt{n},\theta)-\kappa(\sqrt{n},\theta_0)m(\sqrt{n},\theta_0)| \tag{47}$$

for large n.

Define

$$A_n(\theta,\theta_0)=\frac{1}{n}\sum_{t=1}^{n}(f(x_t,\theta)-f(x_t,\theta_0))^2,$$

$$B_n(\theta,\theta_0)=\frac{1}{n}\sum_{t=1}^{n}(f(x_t,\theta)-f(x_t,\theta_0))u_t.$$

Since $\sum_{t=1}^{n}u_t^2/n=\sigma^2+o_p(1)$, we have by the Cauchy–Schwarz inequality

$$|B_n(\theta,\theta_0)|\leq A_n(\theta,\theta_0)^{1/2}(\sigma^2+o_p(1)),$$

and, therefore,

$$(A_n^{-1}|B_n|)(\theta,\theta_0)\leq A_n(\theta,\theta_0)^{-1/2}(\sigma^2+o_p(1)), \tag{48}$$

uniformly in $\theta \in \Theta$. However, using (47) and the inequality

$$\sum_{t=1}^{n}(a_t - b_t)^2 \geq \left(\left(\sum_{t=1}^{n} a_t^2\right)^{1/2} - \left(\sum_{t=1}^{n} b_t^2\right)^{1/2}\right)^2,$$

which holds for any real-valued sequences (a_t) and (b_t), we may deduce that

$$A_n(\theta,\theta_0) \geq \left(\kappa(\sqrt{n},\theta)m(\sqrt{n},\theta) - \kappa(\sqrt{n},\theta_0)m(\sqrt{n},\theta_0)\right)^2 \to_{a.s.} \infty, \tag{49}$$

uniformly in $\theta \in N$.

Now it follows from (48) and (49) that

$$\begin{aligned} n^{-1}D_n(\theta,\theta_0) &= A_n(\theta,\theta_0)(1 - 2(A_n^{-1}|B_n|)(\theta,\theta_0)) \\ &= A_n(\theta,\theta_0)(1 + o_p(1)) \to_p \infty, \end{aligned}$$

uniformly in $\theta \in N$. Since Θ_0 is compact and $\bar{\theta}$ was chosen arbitrarily, we may easily deduce that

$$n^{-1} \inf_{\theta \in \Theta_0} D_n(\theta,\theta_0) \to_p \infty,$$

from which the stated result follows immediately. *Q.E.D.*

PROOF OF COROLLARY 4.4: Let

$$\sigma_n^2 = \frac{1}{n}\sum_{t=1}^{n} u_t^2. \tag{50}$$

It follows from Assumption 2.1(b) that $\sigma_n^2 \to_p \sigma^2$. Assume first that f satisfies the assumptions in Theorem 4.1. Then, as shown in the proof of Theorem 4.1, $n^{-1/2}D_n(\theta,\theta_0) \to_p D(\theta,\theta_0)$ uniformly in $\theta \in \Theta$, with $D(\theta,\theta_0)$ given in Theorem 4.1. We have, in particular, that $D(\cdot,\theta_0)$ is continuous a.s., and therefore $D(\cdot,\theta_0) = o_{a.s.}(1)$ near θ_0. It follows from the consistency of $\hat{\theta}_n$ that $n^{-1/2}D_n(\hat{\theta}_n,\theta_0) = o_p(1)$, which implies $\hat{\sigma}_n^2 = \sigma_n^2 + o_p(n^{-1/2})$, and hence, the stated result follows. For f satisfying the assumptions in Theorem 4.2 with $\kappa < \infty$, we note that $n^{-1}\kappa_n^{-2}D_n(\theta,\theta_0) \to_p D(\theta,\theta_0)$ uniformly in $\theta \in \Theta$, where $D(\cdot,\theta_0)$ is an a.s. continuous function given in Theorem 4.2. Therefore, in the same way as above, we have $\hat{\sigma}_n^2 = \sigma_n^2 + o_p(\kappa_n^2)$. The stated result follows immediately, since $\kappa_n^2 = O_p(1)$. *Q.E.D.*

PROOF OF THEOREM 5.1: Given the I-regularity of $\dot{f}$ in condition (b), AD1–AD3 follow directly from Lemma A6(b) and Theorem 3.2 with $\nu_n = \sqrt[4]{n}$. Since we have in particular

$$\ddot{Q}(\theta_0) = L(1,0)\int_{-\infty}^{\infty} \dot{f}(s,\theta_0)\dot{f}(s,\theta_0)'\,ds,$$

we may easily deduce AD4 from condition (c). Moreover, AD5 holds trivially, since $\hat{\theta}_n \to_p \theta_0$, due to condition (a) and Theorem 4.1, and we assume that θ_0 is an interior point of Θ. It therefore remains to show AD6. For AD6, we prove that $\ddot{Q}_n(\theta) \to_p \ddot{Q}_0(\theta)$ uniformly on a neighborhood of θ_0. In view of the consistency of $\hat{\theta}_n$, this establishes AD6.

We now write

$$\ddot{Q}_n(\theta) = \sum_{t=1}^{n} \dot{f}(x_t,\theta)\dot{f}(x_t,\theta)' + \sum_{t=1}^{n} \ddot{F}(x_t,\theta)(f(x_t,\theta) - f(x_t,\theta_0)) - \sum_{t=1}^{n} \ddot{F}(x_t,\theta)u_t. \tag{51}$$

By Lemma A6(b) and Theorem 3.2, we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\dot f(x_t,\theta)\dot f(x_t,\theta)' \to_p L(1,0)\int_{-\infty}^{\infty}\dot f(s,\theta)\dot f(s,\theta)'\,ds, \tag{52}$$

uniformly in $\theta\in\Theta$. Also, it follows from Lemma A7(b) that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\ddot f(x_t,\theta)u_t \to_p 0, \tag{53}$$

uniformly in $\theta\in\Theta$. Finally, since $|f(\cdot,\theta)-f(\cdot,\theta_0)|$ is I-regular on Θ and $\ddot f$ is bounded, we have from Theorem 3.2

$$\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\ddot f(x_t,\theta)(f(x_t,\theta)-f(x_t,\theta_0))\right\| \le \|\ddot f\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}|f(x_t,\theta)-f(x_t,\theta_0)|$$
$$\to_p \|\ddot f\|L(1,0)\int_{-\infty}^{\infty}|f(s,\theta)-f(s,\theta_0)|\,ds,$$

uniformly in $\theta\in\Theta$. Furthermore, the limit function is continuous in θ by Lemma A8(b), and we can make it arbitrarily small in a neighborhood of θ_0. This, together with (52) and (53), completes the proof. *Q.E.D.*

PROOF OF THEOREM 5.2: We have AD1–AD3 directly from Lemma A6(c) and Theorem 3.3 with $\nu_n=\sqrt{n}\,\dot\kappa(\sqrt{n})$, due to the H-regularity of $\dot h$ in condition (b). Moreover,

$$\ddot Q(\theta_0)=\int_0^1\dot h(V(r),\theta_0)\dot h(V(r),\theta_0)'\,dr=\int_{-\infty}^{\infty}\dot h(s,\theta_0)\dot h(s,\theta_0)'L(1,s)\,ds, \tag{54}$$

and AD4 follows immediately from condition (d). As in the proof of Theorem 5.1, AD5 holds trivially because of the consistency of $\hat\theta_n$ implied by condition (a). To finish the proof, it therefore suffices to show AD6.

Write $\kappa_n=\kappa(\sqrt{n})$, $\dot\kappa_n=\dot\kappa(\sqrt{n})$, and $\ddot\kappa_n=\ddot\kappa(\sqrt{n})$ for notational simplicity. It follows from Lemma A6(c) and Theorem 3.3 that

$$\nu_n^{-1}\sum_{t=1}^{n}\dot f(x_t,\theta)\dot f(x_t,\theta)'\nu_n^{-1\prime} \to_{a.s.} \int_0^1 h(V(r),\theta)\dot h(V(r),\theta)'\,dr, \tag{55}$$

and from Lemma A7(c) that

$$(\nu_n\otimes\nu_n)^{-1}\sum_{t=1}^{n}\ddot f(x_t,\theta)u_t=(\dot\kappa_n\otimes\dot\kappa_n)^{-1}\ddot\kappa_n\left(\frac{\ddot\kappa_n^{-1}}{n}\sum_{t=1}^{n}\ddot f(x_t,\theta)u_t\right)\to_p 0, \tag{56}$$

uniformly in $\theta\in\Theta$. Therefore, for AD6, we only need to show

$$(\nu_n\otimes\nu_n)^{-1}\sum_{t=1}^{n}\ddot f(x_t,\theta_n)(f(x_t,\theta_n)-f(x_t,\theta_0)) \tag{57}$$
$$=((\dot\kappa_n\otimes\dot\kappa_n)^{-1}\kappa_n\ddot\kappa_n)\frac{1}{n}\sum_{t=1}^{n}(\ddot\kappa_n^{-1}\ddot f(x_t,\theta_n))(\kappa_n^{-1}(f(x_t,\theta_n)-f(x_t,\theta_0)))=o_p(1).$$

It is easily seen from (51) that (55)–(57) imply AD6.

To prove (57), apply Theorem 3.3 to $|f(\cdot,\theta)-f(\cdot,\theta_0)|$ and use the local boundedness of $\ddot{h}$ established in Lemma A3(b) to deduce that

$$\left\| \frac{1}{n}\sum_{t=1}^{n}(\ddot{\kappa}_n^{-1}\ddot{f}(x_t,\theta))(\kappa_n^{-1}(f(x_t,\theta)-f(x_t,\theta_0)))\right\|$$

$$\leq \|\ddot{h}\|_K \frac{1}{n\kappa_n}\sum_{t=1}^{n}|f(x_t,\theta)-f(x_t,\theta_0)|$$

$$\to_{a.s.} \|\ddot{h}\|_K \int_0^1 |h(V(r),\theta)-h(V(r),\theta_0)|\,dr,$$

uniformly in $\theta\in\Theta$, where $K=[s_{\min}-1, s_{\max}+1]\times\Theta$. To deduce (57), simply note that the limit function is continuous in θ, due to Lemma A8(a). *Q.E.D.*

Proof of Theorem 5.3: We show that AD1–AD4 and AD7 hold to establish (18). Write $\dot{\kappa}_n = \dot{\kappa}(\sqrt{n})$ to simplify notation. It follows directly from condition (a) and Theorem 3.3 that AD1 holds. Also, AD2 is immediate, since we have from (19)

$$\left\|(\nu_n\otimes\nu_n)^{-1}\sum_{t=1}^{n}\ddot{f}(x_t,\theta_0)u_t\right\| \leq \left\|(\dot{\kappa}_n\otimes\dot{\kappa}_n)^{-1}\left(\sup_{|s|\leq\bar{s}}|f(\sqrt{n}\,s,\theta_0)|\right)\right\|\frac{1}{n}\sum_{t=1}^{n}|u_t| \to_p 0.$$

Under AD2, we may easily get AD3 by applying Lemma A6(b) and Theorem 3.3. In particular, we have $\ddot{Q}(\theta_0)$ given by (54), and therefore, AD4 follows straightforwardly from condition (c).

To show AD7, fix δ such that $0<\delta<\varepsilon/3$, and define $\mu_n = n^{1/2-\delta}\dot{\kappa}_n$ and $\nu_n = n^{1/2}\dot{\kappa}_n$ so that $\mu_n\nu_n^{-1}\to 0$ as required. Let N_n be defined as in AD7. We first write

$$\ddot{Q}_n(\theta)-\ddot{Q}_n(\theta_0) = (\ddot{D}_{1n}(\theta)+\ddot{D}_{1n}(\theta)') + \ddot{D}_{2n}(\theta)+\ddot{D}_{3n}(\theta)+\ddot{D}_{4n}(\theta),$$

where

$$\ddot{D}_{1n}(\theta) = \sum_{t=1}^{n}\dot{f}(x_t,\theta_0)(\dot{f}(x_t,\theta)-\dot{f}(x_t,\theta_0))',$$

$$\ddot{D}_{2n}(\theta) = \sum_{t=1}^{n}(\dot{f}(x_t,\theta)-\dot{f}(x_t,\theta_0))(\dot{f}(x_t,\theta)-\dot{f}(x_t,\theta_0))',$$

$$\ddot{D}_{3n}(\theta) = \sum_{t=1}^{n}\ddot{F}(x_t,\theta)(f(x_t,\theta)-f(x_t,\theta_0)),$$

$$\ddot{D}_{4n}(\theta) = -\sum_{t=1}^{n}(\ddot{F}(x_t,\theta)-\ddot{F}(x_t,\theta_0))u_t,$$

and define

$$\overline{\omega}_{in}^2(\theta) = \|\mu_n^{-1}\ddot{D}_{in}(\theta)\mu_n^{-1\prime}\|,$$

for $i=1,\ldots,4$. For all $\theta\in N_n$, we have

$$\overline{\omega}_{1n}^2(\theta) \leq \sum_{t=1}^{n}\|\mu_n^{-1}\dot{f}(x_t,\theta_0)\|\,\|(\mu_n\otimes\mu_n)^{-1}\ddot{f}(x_t,\bar{\theta})\|, \tag{58}$$

$$\overline{\omega}_{2n}^2(\theta) \leq \sum_{t=1}^{n}\|(\mu_n\otimes\mu_n)^{-1}\ddot{f}(x_t,\bar{\theta})\|^2, \tag{59}$$

$$\overline{\omega}_{3n}^2(\theta) \leq \sum_{t=1}^{n} \| \mu_n^{-1} \dot{f}(x_t, \theta_0) \| \, \| (\mu_n \otimes \mu_n)^{-1} \ddot{f}(x_t, \theta) \| + \frac{1}{2} \sum_{t=1}^{n} \| (\mu_n \otimes \mu_n)^{-1} \ddot{f}(x_t, \overline{\theta}) \| \, \| (\mu_n \otimes \mu_n)^{-1} \ddot{f}(x_t, \theta) \|, \tag{60}$$

$$\overline{\omega}_{4n}^2(\theta) \leq \sum_{t=1}^{n} \| (\mu_n \otimes \mu_n \otimes \mu_n)^{-1} \dddot{f}(x_t, \overline{\theta}) \| \, |u_t|, \tag{61}$$

where $\overline{\theta}$ lies in the line segment connecting θ and θ_0.

Let $\bar{s} = \max(s_{\max}, -s_{\min}) + 1$. Then we have for large n

$$\sup_{\theta \in N_n} | \ddot{f}(x_t, \theta) | \leq \sup_{|s| \leq \bar{s}} \sup_{\theta \in N_n} | \ddot{f}(\sqrt{n}\, s, \theta) |,$$

for all $t = 1, \ldots, n$. It now follows from (58)–(61) that

$$\overline{\omega}_{1n}^2(\theta) \leq \frac{n^{3\delta}}{\sqrt{n}} \left\| (\dot{\kappa}_n \otimes \dot{\kappa}_n)^{-1} \left(\sup_{|s| \leq \bar{s}} \sup_{\theta \in N_n} | \ddot{f}(\sqrt{n}\, s, \theta) | \right) \right\| \frac{1}{n} \sum_{t=1}^{n} \| \dot{\kappa}_n^{-1} \dot{f}(x_t, \theta_0) \|, \tag{62}$$

$$\overline{\omega}_{2n}^2(\theta) \leq \frac{n^{4\delta}}{n} \left\| (\dot{\kappa}_n \otimes \dot{\kappa}_n)^{-1} \left(\sup_{|s| \leq \bar{s}} \sup_{\theta \in N_n} | \ddot{f}(\sqrt{n}\, s, \theta) | \right) \right\|^2, \tag{63}$$

$$\overline{\omega}_{3n}^2(\theta) \leq \frac{n^{3\delta}}{\sqrt{n}} \left\| (\dot{\kappa}_n \otimes \dot{\kappa}_n)^{-1} \left(\sup_{|s| \leq \bar{s}} \sup_{\theta \in N_n} | \ddot{f}(\sqrt{n}\, s, \theta) | \right) \right\| \frac{1}{n} \sum_{t=1}^{n} \| \dot{\kappa}_n^{-1} \dot{f}(x_t, \theta_0) \| + \frac{n^{4\delta}}{2n} \left\| (\dot{\kappa}_n \otimes \dot{\kappa}_n)^{-1} \left(\sup_{|s| \leq \bar{s}} \sup_{\theta \in N_n} | \ddot{f}(\sqrt{n}\, s, \theta) | \right) \right\|^2, \tag{64}$$

$$\overline{\omega}_{4n}^2(\theta) \leq \frac{n^{3\delta}}{\sqrt{n}} \left\| (\dot{\kappa}_n \otimes \dot{\kappa}_n \otimes \dot{\kappa}_n)^{-1} \left(\sup_{|s| \leq \bar{s}} \sup_{\theta \in N_n} | \dddot{f}(\lambda s, \theta) | \right) \right\| \frac{1}{n} \sum_{t=1}^{n} |u_t|, \tag{65}$$

from which we may easily deduce that $\overline{\omega}_{in}^2(\theta) = o_{a.s.}(1)$, $i = 1, \ldots, 4$, uniformly in $\theta \in N_n$, due to (20) and (21). Now AD7 follows immediately from (62)–(65). This completes the proof. *Q.E.D.*

PROOF OF COROLLARY 5.4: Let σ_n^2 be given as (50). Due to Assumption 2.1(b), it suffices to show that $\hat{\sigma}_n^2 = \sigma_n^2 + o_p(1)$. To show this, we define

$$A_n = \sum_{t=1}^{n} (f(x_t, \hat{\theta}_n) - f(x_t, \theta_0))^2,$$

$$B_n = \sum_{t=1}^{n} (f(x_t, \hat{\theta}_n) - f(x_t, \theta_0)) u_t,$$

so that $D_n(\hat{\theta}_n, \theta_0) = A_n - 2B_n$.

Let $\nu_n = n^{1/2} \dot{\kappa}_n$, where $\dot{\kappa}_n = \dot{\kappa}_0(\sqrt{n})$, as defined in the proof of Theorem 5.3. Then we have

$$A_n \leq \| \nu_n' (\hat{\theta}_n - \theta_0) \|^2 \frac{1}{n} \sum_{t=1}^{n} \| \dot{\kappa}_n^{-1} \dot{f}(x_t, \theta_n) \|^2 = O_p(1),$$

$$\frac{|B_n|}{\sqrt{n}} \leq \| \nu_n' (\hat{\theta}_n - \theta_0) \| \left\| \frac{\dot{\kappa}_n^{-1}}{n} \sum_{t=1}^{n} \dot{f}(x_t, \theta_n) u_t \right\| = o_p(1),$$

by Theorem 3.4 and Lemma A7. It therefore follows that $n^{-1/2} D_n(\hat{\theta}_n, \theta_0) = o_p(1)$, from which we have $\hat{\sigma}_n^2 = \sigma_n^2 + o_p(n^{-1/2})$, as required. *Q.E.D.*

REFERENCES

ANDREWS, D. W. K., AND C. J. MCDERMOTT (1995): "Nonlinear Econometric Models with Deterministically Trending Variables," *Review of Economic Studies*, 62, 343–360.

BATES, D. M., AND D. G. WATTS (1988): *Nonlinear Regression Analysis and Its Applications*. New York: Wiley.

BILLINGSLEY, P. (1968): *Convergence of Probability Measures*. New York: Wiley.

CHUNG, K. L., AND R. J. WILLIAMS (1990): *Introduction to Stochastic Integration*, 2nd ed. Boston: Birkhäuser.

GRANGER, C. W. J. (1995): "Nonlinear Relationships between Extended-Memory Variables," *Econometrica*, 63, 265–280.

HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and Its Application*. New York: Academic Press.

HANSEN, B. E. (1992): "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes," *Econometric Theory*, 8, 489–500.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

JENNRICH, R. I. (1969): "Asymptotic Properties of Non-linear Least Squares Estimation," *Annals of Mathematical Statistics*, 40, 633–643.

KURTZ, T. G., AND P. PROTTER (1991): "Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations," *Annals of Probability*, 19, 1035–1070.

MALINVAUD, E. (1970): "The Consistency of Nonlinear Regressions," *Annals of Mathematical Statistics*, 41, 956–969.

PARK, J. Y., AND P. C. B. PHILLIPS (1988): "Statistical Inference in Regressions with Integrated Processes. Part 1," *Econometric Theory*, 4, 468–497.

——— (1999): "Asymptotics for Nonlinear Transformations of Integrated Time Series," *Econometric Theory*, 15, 269–298.

PARK, J. Y., AND P. C. B. PHILLIPS (2000): "Nonstationary Binary Choice," *Econometrica*, 68, 1249–1280.

PHILLIPS, P. C. B. (1986): "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, 33, 311–340.

——— (1987): "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–301.

——— (1991): "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283–306.

PHILLIPS, P. C. B., AND S. N. DURLAUF (1986): "Multiple Time Series with Integrated Variables," *Review of Economic Studies*, 53, 473–496.

PHILLIPS, P. C. B., AND J. Y. PARK (1998): "Nonstationary Density Estimation and Kernel Autoregression," Yale University, Mimeographed.

PHILLIPS, P. C. B., AND V. SOLO (1992): "Asymptotics for Linear Processes," *Annals of Statistics*, 20, 971–1001.

POLLARD, D. (1984): *Convergence of Stochastic Processes*. New York: Springer-Verlag.

REVUZ, D., AND M. YOR (1994): *Continuous Martingale and Brownian Motion*, 2nd ed. New York: Springer-Verlag.

SAIKKONEN, P. (1995): "Problems with the Asymptotic Theory of Maximum Likelihood Estimation in Integrated and Cointegrated Systems," *Econometric Theory*, 11, 888–911.

SHORACK, G. R., AND J. A. WELLNER (1986): *Empirical Processes with Applications to Statistics*. New York: Wiley.

WOOLDRIDGE, J. M. (1994): "Estimation and Inference for Dependent Processes," in *Handbook of Econometrics*, Vol. IV, ed. by R. F. Engle and D. L. McFadden. Amsterdam: Elsevier, pp. 2639–2738.

WU, C. F. (1981): "Asymptotic Theory of Nonlinear Least Squares Estimation," *Annals of Statistics*, 9, 501–513.