COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1190

NONLINEAR REGRESSIONS WITH INTEGRATED TIME SERIES

Joon Y. Park and Peter C. B. Phillips

August 1998

Nonlinear Regressions with Integrated Time Series[1]

Joon Y. Park
Division of Economics
Seoul National University

and

Peter C. B. Phillips
Cowles Foundation for Research in Economics
Yale University

Abstract

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, asymptotically homogeneous and explosive functions. Sufficient conditions for weak consistency are given and a limit distribution theory is provided. In general, the limit theory is mixed normal with mixing variates that depend on the sojourn time of the limiting Brownian motion of the integrated process. 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, to be generally polynomial in $n^{1/2}$ for homogeneous functions, and to be path dependent in the case of explosive functions.

First draft: January 1998

Completed version: July 1998

JEL Classification: C22

Key words and phrases: Functionals of Brownian motion, Brownian motion, integrated process, local time, mixed normal limit theory, nonlinear transformations, nonparametric density estimation, occupation time, nonlinear regression.

[1]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. Phillips thanks the NSF for support under Grant No. SBR 94–22922. The paper was typed by the authors in SW2.5.

1. Introduction

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 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) and 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 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 greatly inhibits the use of the traditional machinery of asymptotic analysis. As Saikonnen put it 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."*

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 analysing even simple nonlinear functionals, as the following example makes clear.

Let x_t be a standard Gaussian random walk with zero initialization. Then, $n^{-\frac{1}{2}}x_{[n\cdot]} = X_n(\cdot) \to_d W(\cdot)$. The nonlinear function $f(x) = 1/(1+x^2)$ is everywhere continuous and well behaved at the limit of the domain of definition of x_t. What is the limit behavior of the sample mean function $\sum_{t=1}^{n} f(x_t)$? The standard approach outlined in the previous paragraph suggests the following

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

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, 1987) 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

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

for $\varepsilon > 0$. Thus, $\int_0^1 \frac{dr}{W(r)^2} = \infty \ a.s.$ and all we have shown in (1) is that $\sum_{t=1}^{n} \frac{1}{1+x_t^2}$ diverges as $n \to \infty$.

How then do we analyse 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. $\frac{1}{2\delta}\times\text{time}\,(x_t \in [s-\delta, s+\delta];\ t = 1, ..., n)$. 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

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \frac{1}{1+x_t^2} \sim \frac{1}{\sqrt{n}} \sum_{\min_{t\leq n}(x_t)}^{\max_{t\leq n}(x_t)} \frac{1}{1+s^2} \times \frac{1}{2\delta} \times \text{time}\,(x_t \in [s-\delta, s+\delta];\ t = 1, ..., n)$$

$$= \sum_{\min_{t\leq n}(x_t)}^{\max_{t\leq n}(x_t)} \frac{1}{1+s^2} \times \frac{1}{n}\frac{1}{2\varepsilon} \times \text{time}\left(\frac{x_t}{\sqrt{n}} \in \left[\frac{s}{\sqrt{n}} - \varepsilon, \frac{s}{\sqrt{n}} + \varepsilon\right];\ t = 1, ..., n\right).$$

Now as $n \to \infty$ we note that $\max_{t\leq n}(x_t) \to \infty$, $\min_{t\leq n}(x_t) \to -\infty$, so that for large n we have

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

Also for all finite s we have

$$\frac{1}{n}\frac{1}{2\varepsilon} \times \text{time}\left(\frac{x_t}{\sqrt{n}} \in \left[\frac{s}{\sqrt{n}} - \varepsilon, \frac{s}{\sqrt{n}} + \varepsilon\right];\ t = 1, ..., n\right)$$

$$\sim \frac{1}{n}\frac{1}{2\varepsilon} \times \text{time}\left(\frac{x_t}{\sqrt{n}} \in [-\varepsilon, \varepsilon]\,;\ t = 1, ..., n\right)$$
$$\sim \frac{1}{2\varepsilon}\int_0^1 1(|X_n(r)| \le \varepsilon)dr$$
$$\sim \frac{1}{2\varepsilon}\int_0^1 1(|W(r)| \le \varepsilon)dr.$$

From these heuristics we get the approximate expression

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

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 (2). Noting that ε was arbitrary, we can let $\varepsilon \to 0$ in (2). 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

$$L_W(1, 0) = \lim_{\varepsilon \to 0+}\frac{1}{2\varepsilon}\int_0^1 1(|W(r)| \le \varepsilon)dr \tag{3}$$

is well defined and is known as the local time of standard Brownian motion at the origin. 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 (2) and (3), our heuristics lead us to the following asymptotic behaviour as $n \to \infty$

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{1}{1+x_t^2} \to_d \left(\int_{-\infty}^{\infty}\frac{ds}{1+s^2}\right)L_W(1, 0). \tag{4}$$

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 (4) could have been used to obtain the limit behavior of the sample mean of an arbitrary integrable function f, specifically

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

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

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 (1997) and makes rigorous the ideas described above. Consistency in nonlinear regression is proved in Section 4 and the limit distribution theory is developed in Section 5. Some extensions to additive nonlinear regression models, including partial linear regressions, are given in Section 6. Section 7 concludes. A technical Appendix is provided and is divided into two parts. Some useful technical lemmas are given in Section 8 and proofs of the theorems in the paper are given in Section 9.

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 $\mathbf{R}_+$ ($\mathbf{R}_-$) the set of positive (negative) numbers.

2. The Model and Preliminary Results

We consider the nonlinear regression model for y_t given by

$$y_t = f(x_t, \theta_0) + u_t \tag{6}$$

where $f : \mathbf{R} \to \mathbf{R}$ is known, x_t and u_t are the regressors and regression errors, respectively, and θ_0 is the true parameter value which lies in the parameter set Θ. In model (6), 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.

The nonlinear regression (6) can be estimated by nonlinear least squares (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.,

$$\hat{\theta}_n = \underset{\theta\in\Theta}{\arg\max}\ Q_n(\theta) \tag{7}$$

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 Park and Phillips (1997). When the initialization is in the distant past and $x_0 = O_p\left(n^{\frac{1}{2}}\right)$, 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

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

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* $(\mathcal{F}_{nt})$, $t = 0, \ldots, n$, *such that*

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

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

Assumption 2.1 is quite weak, and 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 — e.g., 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 stationary time series regression. However, it is not essential for regressions with integrated time series. As is well known, serial correlation in the errors and cross correlation between the errors and regressors can be allowed in linear cointegrating regressions. They do not affect, for instance, the consistency of the least squares estimator. 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. 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 $(\mathcal{F}_{nt})$. Note that conditions (b) and (c) together imply, in particular, that $\mathbf{E}(y_t|\mathcal{F}_{n,t-1}) = f(x_t, \theta_0)$ 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 topology. 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 $(\Omega, \mathcal{F}, \mathbf{P})$ supporting (U_n°, V_n°) and (U, V) such that

$$(U_n, V_n) =_d (U_n^\circ, V_n^\circ) \quad \text{and} \quad (U_n^\circ, V_n^\circ) \to_{a.s.} (U, V) \tag{8}$$

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° *introduced in* (8) *by*

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

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

$$\sup_{1 \le t \le n} \left| \frac{\tau_{nt} - t}{n^\delta} \right| \to_{a.s.} 0 \tag{9}$$

as $n \to \infty$ *for any* $\delta > 2/q$.

In the paper, we establish the weak consistency and derive the asymptotic distribution of the NLS estimator $\hat{\theta}_n$ defined in (7). For our purposes, it therefore causes no loss in generality to assume $(U_n, V_n) = (U_n^\circ, V_n^\circ)$, instead of $(U_n, V_n) =_d (U_n^\circ, V_n^\circ)$ as in (8). This convention will be made throughout the paper. It allows us to avoid repetitious embedding of (U_n, V_n) in the probabilty space $(\Omega, \mathcal{F}, \mathbf{P})$, where (U_n°, V_n°) 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 correponding 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

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 $V_n^\circ \to_d V$, as shown in Phillips and Solo (1992).

Assumption 2.3 *Let*(a) *and*(b) *be given as in*Assumption 2.1, *and let* (d) *be given as in*Assumption 2.2. *We assume*
(c) x_t^* *is adapted to* $\mathcal{F}_{n,t-1}$, $t = 1, \ldots, n$,
where $x_t^* = x_t - \max_{1\leq t\leq n} x_t$.

Condition (c) of Assumption 2.3 can be truly restrictive. Though it is certainly satisfied when x_t and u_t are independent and the filtration $\mathcal{F}_{n,t}$ is suitably defined, it may not hold for many interesting econometric models. Roughly speaking, it requires that knowledge of deviations from the maximum taken over future and past values of the regressor should not help in predicting future values of the regression error. This may be unlikely and unrealistic, though it is not totally unacceptable. As will become clear, we need the condition only to fully analyze the asymptotics for the case of nonlinear regression with explosive regression functions.

Under condition (c) of Assumption 2.3, it is often more convenient to work with

$$V_n^*(r) = V_n(r) - \sup_{0\leq s\leq 1} V_n(s) \quad \text{and} \quad V^*(r) = V(r) - \sup_{0\leq s\leq 1} V(s)$$

in place of V_n and V. Given condition (a), we have $(U_n, V_n^*) \to_d (U, V^*)$ by the continuous mapping theorem. Therefore, we may embed (U_n, V_n^*) into the probability space supporting $(U_n^\circ, V_n^{\circ *})$ and (U, V^*) so that $(U_n, V_n^*) =_d (U_n^\circ, V_n^{\circ *})$ and $(U_n^\circ, V_n^{\circ *}) \to_{a.s.} (U, V^*)$ uniformly on $[0, 1]$. This is precisely analogous to (8). Moreover, we may also represent U_n° by U with appropriate time changes, as in Lemma 2.1, under conditions (b) and (c) of Assumption 2.3. This can be seen from the proof of Lemma 2.1. For the subsequent development of our theory, we also use the convention $(U_n, V_n^*) = (U_n^\circ, V_n^{\circ *})$, correponding to our earlier convention $(U_n, V_n) = (U_n^\circ, V_n^\circ)$.
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. For a continuous semimartingale M with quadratic variation $[M]$, 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* **R**. *Then*

$$\int_0^t T(M(r))\,d[M]_r = \int_{-\infty}^{\infty} T(s)L_M(t,s)\,ds \tag{10}$$

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 (10), the occupation time is defined with respect to $d[M]_r$, 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 (10) with $T(x) = 1\{s \le x < s+\varepsilon\}$ to get

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

a representation which 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 (10), of course, applies to Brownian motion as a special case. For Brownian motion, the result in (10) 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, 1988). 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

$$L(t,s) = (1/\sigma_v^2)\,L_V(t,s) \tag{11}$$

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 (10) 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 (11) will also be used without further reference.

We often need to evaluate L at the maximum or the minimum of the sample path of V, which we define by

$$s_{\max} = \max_{0 \leq r \leq 1} V(r) \quad \text{and} \quad s_{\min} = \min_{0 \leq r \leq 1} V(r)$$

Therefore, expressions like $L(\cdot, s_{\max})$ or $L(\cdot, s_{\min})$ will often appear in our formulae. 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 $(\Omega, \mathcal{F}, \mathbf{P})$ containing the processes U and V. These conventions will be used throughout the paper.

3. Asymptotics for Nonlinear 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 Park and Phillips (1997), which we denote hereafter by 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 component is regular.

Logarithmic functions and reciprocals are not regular. Therefore, our subsequent theory for regular functions are 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. For any fixed n, 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. 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. 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 non zero 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.

We consider three different families of functions, corresponding to each of the three types of functions studied in P^2. Introduced below are regularity conditions for the three (I-, H- and E-) 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 *or* 2.3, *$\|F(x,\pi) - F(y,\pi)\| \le c\,|x-y|^k$ 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 $\mathbf{R}$ will play an important role and we denote this class by $\mathcal{T}_{LB}$. Any regular transformation T on $\mathbf{R}$, defined in Definition 3.1, belongs to $\mathcal{T}_{LB}$. We often need to consider a sub-class $\mathcal{T}_{LB}^0$ of $\mathcal{T}_{LB}$ consisting only of locally bounded transformations which 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 $\mathcal{T}_B$ of bounded transformations on

$\mathbf{R}$, and its subclass $\mathcal{T}_B^0$ 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 $\mathbf{R}$ for which sample mean and sample covariance asymptotics can effectively be bounded. Clearly, $\mathcal{T}_B^0 \subset \mathcal{T}_B \subset \mathcal{T}_{LB}^0 \subset \mathcal{T}_{LB}$. 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 $z : \mathbf{R}_+ \times \Phi \to \mathbf{R}_+^{m^2}$ 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 $\lambda \to \infty$,

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

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 $Z(\cdot,\lambda,\omega)$, parametrized 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*

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

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

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 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 neglibile 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 analagous 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.

3.2(c) Exponential Functions

We start with the following definition.

Definition 3.6 *Let*

$$F(\lambda x,\pi) = \kappa(\lambda s,\pi)E\left(\lambda(x-s),\pi\right) + R(x-s,\lambda,s,\pi)$$

where κ is a nonsingular matrix, and define $E_-(x,\pi) = E(x,\pi)1\{x \le 0\}$ *and* $R_-(x,\lambda,s,\pi) = R(x,\lambda,s,\pi)1\{x \le 0\}$. *We say that F is E-regular on* Π *if*
(a) E_- is I*-regular, and*
(b) $R_-(\cdot - s,\lambda,s,\pi)$ *is of order smaller than* $\kappa(\lambda s,\pi)/\lambda$ *on* $\mathbf{R}_+ \times \Pi$.
We call κ the asymptotic order, *and* E *the* limit exponential function.

These conditions will be called E-*regularity conditions.* The functions in the E-regular class are essentially exponentials, as implied by the form of F and the E-regularity condition (a). Note that ordinary exponential functions, when restricted to $\mathbf{R}_-$, are I-regular, and therefore satisfy the condition. Similarly as in the second H-regularity condition, E-regularity condition (b) ensures that the residual component is negligible. Note that the parameter set over which we define asymptotic order is $\mathbf{R}_+ \times \Pi$, not Π. We should therefore compare R_- with κ for each $(s,\pi) \in \mathbf{R}_+ \times \Pi$, not just for $\pi \in \Pi$. Naturally, the class of E-regular functions accomodates the exponential function e^x. However, unlike the corresponding condition in Park and Phillips (1997), the function xe^x is not allowed here. We can nevertheless analyze such functions by looking at the vector function $T(x) = (e^x, xe^x)'$. Note that we may decompose $T(x)$ as

$$\begin{pmatrix} e^{\lambda x} \\ \lambda x e^{\lambda x} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \lambda s & 1 \end{pmatrix} \begin{pmatrix} e^{\lambda x} \\ \lambda(x-s)e^{\lambda x} \end{pmatrix} = \begin{pmatrix} e^{\lambda s} & 0 \\ \lambda s e^{\lambda s} & e^{\lambda s} \end{pmatrix} \begin{pmatrix} e^{\lambda(x-s)} \\ \lambda(x-s)e^{\lambda(x-s)} \end{pmatrix}$$

from which form it is apparent that T is in the E-regular family.

Theorem 3.4 *Let F be specified as in* Definition 3.6. *If F is* E*-regular on a compact set Π, then under* Assumption 2.2 *or* 2.3

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

as $n \to \infty$, uniformly in $\pi \in \Pi$. Moreover, if $F(\cdot,\pi)$ is E*-regular, then under* Assumption 2.3,

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

as $n \to \infty$.

Asymptotics for E-regular functions have a special feature that does not typically arise elsewhere, viz. path-dependent rates of convergence. As we will show, the convergence rate of the asymptotic sample moments of E-regular functions depends not only on the size, but also on the actual path of the sample. Aside from this important feature, the asymptotics of E-regular functions is analogous to that of I-regular functions. To find the asymptotic behaviour of an E-regular function, we simply restrict the support of the function to $\mathbf{R}_-$ and evaluate the local time L at the maximum of the sample path of the limit Brownian motion V. The asymptotics for the function F such that $F(-\cdot,\pi)$ is E-regular can also be easily obtained. For such functions, we may well expect that the convergence rate is dependent upon $\min x_t$ instead of $\max x_t$. Also, the asymptotics in this case will obviously be given in terms of $L(\cdot, s_{\min})$, not $L(\cdot, s_{\max})$.

Just as for regular families of functions, I-, H- and E-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-, H- and E-regularity) conditions are preserved, if we compose a regular (I-, H- and E-regular) family with certain types of functions. For instance, if F is regular (I-, H-, and E-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-, H- and E-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 (7). 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.

Defining $D_n(\theta, \theta_0) = Q_n(\theta) - Q_n(\theta_0)$. To prove consistency, we show one of the following two conditions.

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 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 guarantee that $D(\cdot, \theta_0)$ has unique minimum θ_0. However, this approach is not applicable for general H-regular and E-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 CN1 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 which 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 CN1 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 (6) 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 analyse 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* CN2 *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 $\theta_0 \notin N$.

Theorem 4.4 *Let* Assumption 2.2 *or* 2.3 *hold, and let f be* E*-regular on Θ with asymptotic power order κ and limit exponential function e. Then* CN2 *holds if*

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

$$\lambda^{-1} \inf_{\substack{|p-\bar{p}|<\varepsilon \\ |q-\bar{q}|<\varepsilon}} \inf_{|s-\bar{s}|<\varepsilon} \inf_{\theta \in N} |pe(\lambda s,\theta) - qe(\lambda s,\theta_0)| \to \infty$$

(b) *for all $\theta \in \Theta$, $\int_{-\infty}^0 e(s,\theta)^2 ds > 0$.*

It is straightforward to show that $f(x,\theta) = e^{\theta x}$ with $\Theta \subset \mathbf{R}_+$ satisfies the conditions in Theorem 4.4. Notice that f is E-regular, is its own limit exponential function, and therefore, $e(\lambda s,\theta) = e^{\theta\lambda s}$. To show that condition (a) holds, simply set $0 < \varepsilon < \min(\bar{s},\bar{p},\bar{q})$ for any given $\bar{s},\bar{p},\bar{q} > 0$ and define N to be any neighborhood of $\bar{\theta} \neq \theta_0$ such that $\theta_0 \notin N$, as in Example 4.3. It is trivial to show that condition (b) also holds.

Corollary 4.5 *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 $\hat{\sigma}_n^2$ 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 and E-regular regression functions will be shown in the next section under stronger assumptions.

It is interesting to note that $\tilde{\sigma}_n^2 = (1/n)\sum_{t=1}^n y_t^2 \to_p \sigma^2$ 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 4.5.

It is also possible to establish consistency of the NLS estimator using a triangular array asymptotic framework. Suppose that the actual size of the sample is fixed at, say, n_0. Then we can consider a regression model, which is given for each n by

$$y_{nt} = f(x_{nt}, \theta_0) + u_t \quad \text{with} \quad x_{nt} = \left(\frac{n_0}{n}\right)^{1/2} x_t \tag{12}$$

$t = 1, \ldots, n$, and analyze the asymptotic behavior of the NLS estimator of θ_0 as $n \to \infty$. When $n = n_0$, we have $x_{nt} = x_t$ and $y_{nt} = y_t$, and (12) reduces to the original nonlinear regression (6). If n_0 is large, the large n asymptotics for the reformulated regression model in (12) should therefore provide a reasonable approximation for the original regression model (6). We denote by $\bar{\theta}_n$ the NLS estimator of θ_0, and by $\bar{\sigma}_n^2$ the corresponding estimator of σ^2, obtained from the regression (12). We get the following asymptotics.

Theorem 4.6 *Let* Assumption 2.1 *hold, and let* f *be regular on* Θ *such that, for all* $\theta \neq \theta_0$ *and* $\delta > 0$, $\int_{|s|\leq\delta}(f(s,\theta) - f(s,\theta_0))^2 ds < \infty$. *Then* $\bar{\theta}_n \to_p \theta_0$ *and* $\bar{\sigma}_n^2 \to_p \sigma^2$ *as* $n \to \infty$.

The advantage of triangular array asymptotics is that we can derive the consistency of the NLS estimator quite easily and under relatively simple conditions, which are applicable to all the types of regression function that are considered in this paper. However, they also have some obvious drawbacks. For instance, models like (12) are really only approximations to (6) and it is not clear how to interpret the accuracy of the approximations they provide, in contrast to standard sequential asymptotics which generate approximations with a degree of precision that is given explicitly in terms of the sample size and for a model that is unchanging.[2] Clearly, this approach

[2]For more discussion on the simplifying characteristics of triangular array asymptotics, the reader may refer to Andrews and McDermott (1995) and the references therein, where the approach is used to analyse deterministic trends in nonlinear regression.

to asymptotic theory for trending nonlinear regression avoids the difficulty of dealing directly with nonlinear functions of integrated process by substituting approximations in terms of normalized processes. In doing so, some of the character of the nonlinear dependence is lost. It will be of interest to conduct simulations with both approaches to furnish information about the adequacy of this alternative approach to an asymptotic theory.

5. Limit Distributions

This section of the paper derives the asymptotic distribution of the NLS estimator $\hat{\theta}_n$ defined in (7). 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

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

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 respectively by $\dot{h}$ and $\dot{e}$ the limit homogeneous function and limit exponential function of H- and E-regular $\dot{f}$. Moreover, the asymptotic orders of $\dot{f}, \ddot{f}$ and $\dddot{f}$ of H- or E-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

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

ignoring 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

$$\dot{Q}_n(\hat{\theta}_n) = \dot{Q}_n(\theta_0) + \ddot{Q}_n(\theta_n)(\hat{\theta}_n - \theta_0), \tag{13}$$

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 (7).

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

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

then $\nu_n^{-1}\ddot{Q}_n^\circ(\theta_0)\nu_n^{-1\prime} \to_p \ddot{Q}(\theta_0)$ 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 $\nu_n^{-1}\ddot{Q}_n(\theta_n)\nu_n^{-1\prime} = \nu_n^{-1}\ddot{Q}_n^\circ(\theta_0)\nu_n^{-1\prime} + o_p(1)$ and $\ddot{Q}(\theta_0) > 0$ a.s., we may expect from (13) that

$$\begin{aligned} \nu_n'(\hat{\theta}_n - \theta_0) &= -\left(\nu_n^{-1}\ddot{Q}_n(\theta_n)\nu_n^{-1\prime}\right)^{-1} \nu_n^{-1}\dot{Q}_n(\theta_0) \\ &= -\left(\nu_n^{-1}\ddot{Q}_n^\circ(\theta_0)\nu_n^{-1\prime}\right)^{-1} \nu_n^{-1}\dot{Q}_n(\theta_0) + o_p(1) \\ &\to_d -\ddot{Q}(\theta_0)^{-1}\dot{Q}(\theta_0) \end{aligned} \tag{14}$$

as $n \to \infty$.

For easy reference, we list a set of sufficient conditions that lead to (14), 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: $\nu_n^{-1}\ddot{Q}_n(\theta_0)\nu_n^{-1\prime} = \nu_n^{-1}\ddot{Q}_n^\circ(\theta_0)\nu_n^{-1\prime} + o_p(1)$ *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* $n \to \infty$.

Conditions AD1–AD6 are standard in nonlinear regression analysis. Given AD1–AD6, (14) follows immediately from (13).

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, $\nu_n^{-1}\ddot{Q}_n^\circ(\theta)\nu_n^{-1\prime}$ 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

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

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

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

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. Finally, 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 $\ddot{h}(x,\theta) = 2$.

It is more difficult to establish AD5 and AD6 for general H- or E-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 Theorems 4.3 and 4.4, where the consistency of $\hat{\theta}_n$ is established for general H and E-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

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

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

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 (14).

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

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

for $\varepsilon > 0$ given. We have:

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* $\lambda \to \infty$

$$\left\| (\dot{\kappa}_0 \otimes \dot{\kappa}_0)(\lambda)^{-1} \left(\sup_{|s| \leq \bar{s}} |\ddot{f}(\lambda s, \theta_0)| \right) \right\| \to 0 \tag{15}$$

$$\lambda^{-1+\varepsilon} \left\| (\dot{\kappa}_0 \otimes \dot{\kappa}_0)(\lambda)^{-1} \left(\sup_{|s| \leq \bar{s}} \sup_{\theta \in N(\varepsilon,\lambda)} |\ddot{f}(\lambda s, \theta)| \right) \right\| \to 0 \tag{16}$$

$$\lambda^{-1+\varepsilon} \left\| (\dot{\kappa}_0 \otimes \dot{\kappa}_0 \otimes \dot{\kappa}_0)(\lambda)^{-1} \left(\sup_{|s| \leq \bar{s}} \sup_{\theta \in N(\varepsilon,\lambda)} |\dddot{f}(\lambda s, \theta)| \right) \right\| \to 0 \tag{17}$$

(c) $\int_{|s| \leq \delta} \dot{h}(s, \theta_0) \dot{h}(s, \theta_0)' ds > 0$ *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-regrular functions, there exist $\varepsilon>0$ such that

$$\lambda^\varepsilon\left\|(\dot\kappa_0\otimes\dot\kappa_0)(\lambda)^{-1}\left(\sup_{|s|\le\bar s}\sup_{\theta\in N(\varepsilon,\lambda)}|\ddot f(\lambda s,\theta)|\right)\right\|\to 0 \tag{18}$$

for any $\bar s>0$. Clearly, (15) and (16) are satisfied if (18) holds. Moreover, we show in the proof of Theorem 5.3 that thrice differentiability of f and (17) are unnecessary if (18) holds true. Conditions in (15)–(17) can thus be replaced by (18), 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

$$\lambda^{-1+\varepsilon}\left\|\dot\kappa_0(\lambda)^{-1}\right\|\to 0 \tag{19}$$

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 (19) and

$$\lambda^\varepsilon\left\|(\dot\kappa_0\otimes\dot\kappa_0)(\lambda)^{-1}\left(\sup_{|s|\le\bar s}\sup_{\theta\in N(\delta)}|\ddot f(\lambda s,\theta)|\right)\right\|\to 0 \tag{20}$$

for some $\delta>0$, instead of (18). It is indeed quite easy and straightforward to show that (19) and (20) 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

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

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 (19) and (20) are satisfied for f. To show this, we simply let ε and δ in (19) and (20) 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. Likewise, the distribution theory for the regressions with E-regular regression functions in Theorem 5.4 below is somewhat similar to those with I-regular regression functions in Theorem 5.1. The limiting distribution involved is mixed Gaussian in both cases. However, the convergence rate for the NLS estimator in regressions with E-regular regression functions is path-dependent. It depends not only on the size of the sample, but also on the actual values of the path by way of their maximum. We may expect that with large probability (and eventually with probability one) the convergence rate is much faster than other types of regressions, since x_t has a stochastic trend and $\dot{\kappa}_0$ is like an exponential function.

The asymptotic distribution of $\hat{\theta}_n$ in the regression with E-regular $\dot{f}$ can be obtained in a similar way. We let $\dot{\kappa}(\cdot) = \dot{\kappa}(\cdot, \theta_0)$ as earlier, and define a neighborhood of θ_0 for E-regular $\dot{f}$ as

$$N(\bar{s}, \varepsilon, \lambda) = \left\{ \theta : \left\| \left(\inf_{|s-\bar{s}|\leq\varepsilon} \dot{\kappa}_0(\lambda s) \right)' (\theta - \theta_0) \right\| \leq \lambda^{-1/2+\varepsilon} \right\}$$

for given $\bar{s}$ and $\varepsilon > 0$.

Theorem 5.4 *Let* Assumption 2.3 *hold. Assume:*
(a) $\dot{f}$ *is E-regular on* Θ;
(b) *for any* $\bar{s} > 0$ *given, there exists* $\varepsilon > 0$ *such that as* $\lambda \to \infty$

$$\lambda \left\| \left(\inf_{|s-\bar{s}|\leq\varepsilon} (\dot{\kappa}_0 \otimes \dot{\kappa}_0)(\lambda s) \right)^{-1} \sup_{s \leq \bar{s}+\varepsilon} |\ddot{f}(\lambda s, \theta_0)| \right\| \to 0, \quad (21)$$

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

$$\lambda^{1/2+\varepsilon} \left\| \left(\inf_{|s-\bar{s}|\leq\varepsilon} (\dot{\kappa}_0 \otimes \dot{\kappa}_0 \otimes \dot{\kappa}_0)(\lambda s) \right)^{-1} \sup_{s \leq \bar{s}+\varepsilon} \sup_{\theta \in N(\bar{s},\varepsilon,\lambda)} |\dddot{f}(\lambda s, \theta)| \right\| \to 0; \quad (23)$$

(c) $\int_{-\infty}^{0} \dot{e}(s, \theta_0)\dot{e}(s, \theta_0)' ds > 0$.
Then

$$\sqrt[4]{n} \dot{\kappa}_0 \left(\max_{1\leq t\leq n} x_t \right)' \left(\hat{\theta}_n - \theta_0 \right) \to_d \left(L(1, s_{\max}) \int_{-\infty}^{0} \dot{e}(s, \theta_0)\dot{e}(s, \theta_0)' ds \right)^{-1/2} W(1)$$

as $n \to \infty$.

Remark If the regression function f has E-regular $\dot{f}$ as assumed in (a), it suffices to have condition (c) as the identification condition. The conditions in (a) and (c) are easy to verify. Similarly, in the case of H-regular functions, we may replace the conditions in (b) with a simpler one. Indeed, it suffices to have $\varepsilon, \delta > 0$ such that

$$\lambda^{-1/2+\varepsilon} \left\| \left(\inf_{|s-\bar{s}|\leq\varepsilon} \dot{\kappa}_0(\lambda s) \right)^{-1} \right\| \to 0 \tag{24}$$

and

$$\lambda^{1+\varepsilon} \left\| \left(\inf_{|s-\bar{s}|\leq\varepsilon} (\dot{\kappa}_0 \otimes \dot{\kappa}_0)(\lambda s) \right)^{-1} \sup_{s\leq\bar{s}+\varepsilon} \sup_{\theta\in N(\delta)} |\ddot{f}(\lambda s, \theta)| \right\| \to 0 \tag{25}$$

for any $\bar{s} > 0$. It seems that conditions (24) and (25) are satisfied for many E-regular functions that are used in nonlinear analyses.

Example 5.2 Let $\Theta \subset \mathbf{R}_+ \times \mathbf{R}_+$ and write $\theta = (\alpha, \beta)$. Consider $f(x, \alpha, \beta) = \alpha e^{\beta x}$, for which we have

$$\dot{\kappa}(w, \alpha, \beta) = \begin{pmatrix} e^{\beta w} & 0 \\ \alpha w e^{\beta w} & e^{\beta w} \end{pmatrix} \quad \text{and} \quad \dot{e}(x, \alpha, \beta) = \begin{pmatrix} e^{\beta x} \\ \alpha x e^{\beta x} \end{pmatrix}$$

and

$$\int_{-\infty}^{0} \dot{e}(s, \alpha, \beta)\dot{e}(s, \alpha, \beta)' ds = \begin{pmatrix} \frac{1}{2\beta} & -\frac{\alpha}{4\beta^2} \\ -\frac{\alpha}{4\beta^2} & \frac{\alpha^2}{4\beta^3} \end{pmatrix} > 0$$

whenever $\alpha \neq 0$ and $\beta > 0$. Moreover, it is easy to see that conditions in (24) and (25) are satisfied if we choose ε and $\delta > 0$ so that $-2\beta_0(\bar{s} - \varepsilon) + (\beta_0 + \delta)(\bar{s} + \varepsilon) < 0$, which holds whenever $0 < \varepsilon < \bar{s}/3$ and $0 < \delta < \beta_0(\bar{s} - 3\varepsilon)/(\bar{s} + \varepsilon)$ as can be readily checked.

Corollary 5.5 *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 4.5 and 5.5 establish the consistency of the error variance estimator $\hat{\sigma}_n^2$ for regressions with all types of 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- and E-regular regression functions, test statistics of the usual form all have limiting chi-square distibutions, respectively, under the assumptions of Theorems 5.1 and 5.4. 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 $\sigma_{uv} = 0$.

We conclude this section by considering the asymptotic distribution of the NLS estimator $\bar{\theta}_n$ defined in Section 4 for the case of simplifying triangular array asymptoics. As shown below, the convergence rate is $\sqrt{n}$, regardless of the type of regression. The distribution theory is completely analogous to that of the conventional NLS estimator $\hat{\theta}_n$ from a regression with H-regular regression functions. It is non Gaussian/mixed Gaussian unless $\sigma_{uv} = 0$.

Theorem 5.6 *Let* Assumption 2.1 *and the conditions of* Theorem 4.6 *hold. Assume that $\dot{f}$ and $\ddot{f}$ are regular, and $\int_{|s|\leq\delta} \dot{f}(s,\theta_0)\dot{f}(s,\theta_0)'ds > 0$ for all $\delta > 0$. Then*

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

as $n \to \infty$, where $\dot{f}_0(x,\theta) = \dot{f}(\sqrt{n}_0 x, \theta)$.

Tests can be constructed from $\bar{\theta}_n$ in the usual way with results analogous to those discussed above. As indicated above, the limit theory for these tests will be chi square only when $\sigma_{uv} = 0$.

6. Asymptotics for Additive Models

In this section, we consider a nonlinear additive regression in which the regression function has additive components of the form

$$f(x,\theta) = \sum_{i=1}^{m} f_i(x,\theta_i) \tag{26}$$

where the f_i's are the I-, H- or E-regular functions introduced earlier in the paper, and the θ_i's are the unknown parameters. Such a formulation allows us to consider a wider class of nonlinear regression models including, in particular, regressions that are partially nonlinear. Unless explicitly stated otherwise, we will assume that there are no restriction across the θ_i's, and $\theta = (\theta_1', \ldots, \theta_m')'$. Define $\theta_0 = (\theta_{10}', \ldots, \theta_{m0}')'$.

It is useful to compare regression (6) with the regression function (26) and the regressions

$$y_t = f_i(x_t, \theta_i) + u_t \tag{27}$$

for $i = 1, \ldots, m$. Denote by $\tilde{\theta}_{in}$ the NLS estimator for θ_i based on the regression (27) for each $i = 1, \ldots, m$. Clearly, all our previous results apply to these regressions. In particular, for some sequence ν_{in}, $\nu_{in}'\left(\tilde{\theta}_{in} - \theta_{i0}\right)$ has a well-defined limiting distribution under appropriate regularity conditions for f_i for $i = 1, \ldots, m$.

The asymptotics for the additive model (26) can also be derived as in (14) by showing that AD1–AD4 and AD7 hold for some normalizing sequence ν_n. To derive the asymptotic theory for the additive model along these lines, it is first necessary to reformulate the conditions of Theorems 5.1 and 5.2 conformably with those in Theorems 5.3 and 5.4.

Theorem 6.1 *Let* Assumption 2.2 *hold. Assume that* $\dot{f}$, $\ddot{f}$ *and* $\dddot{f}$ *are I-regular, and* $\int_{-\infty}^{\infty} \dot{f}(s,\theta_0)\dot{f}(s,\theta_0)'ds > 0$. *Then the result of* Theorem 5.1 *follows.*

Theorem 6.2 *Let*Assumption 2.1 hold. *Assume that* $\dot{f}$, $\ddot{f}$ *and* $\dddot{f}$ *are* H_0*-regular,* $\|(\dot{\kappa}\otimes\dot{\kappa})^{-1}\ddot{\kappa}\|, \|(\dot{\kappa}\otimes\dot{\kappa}\otimes\dot{\kappa})^{-1}\dddot{\kappa}\| < \infty$, *and* $\int_{|s|\le\delta} \dot{h}(s,\theta_0)\dot{h}(s,\theta_0)'ds > 0$ *for all* $\delta > 0$. *Then the result of* Theorem 5.2 *follows.*

Let ν_{in} be the normalizing sequence for the NLS estimator $\tilde{\theta}_{in}$ from (27), $i = 1,\ldots,m$, as introduced above. We define a normalizing sequence ν_n by

$$\nu_n = \mathrm{diag}\,(\nu_{1n},\ldots,\nu_{mn}) \tag{28}$$

for the NLS estimator $\hat{\theta}_n$ in the regression (6) with the regression function (26).

Lemma 6.3 *Let* f *be given by* (26). *If the set of assumptions in any of* Theorems 6.1, 6.2, 5.3 *and* 5.4 *are satisfied for each of* f_i, $i = 1,\ldots,m$, *then* AD1, AD2 *and* AD7 *hold with the normalizing sequence* ν_n *in* (28).

Therefore, we only need to establish AD3 and AD4 to deduce (14). That is, it is sufficient to show that $\nu_n^{-1}\ddot{Q}_n^{\circ}(\theta_0)\nu_n^{-1\prime}$ converges in probability to some $\ddot{Q}(\theta_0)$ and that $\ddot{Q}(\theta_0) > 0$ a.s. The former follows directly from our results in Section 3, and the latter will be fulfilled under appropriate identifying assumptions. Often the limit matrix $\ddot{Q}(\theta_0)$ becomes block diagonal, in which case AD4 is automatically satisfied by the assumptions on each component function. Moreover, we have asymptotic equivalence between $\hat{\theta}_{in}$ and $\tilde{\theta}_{in}$ in such cases. In subsequent discussion on the asymptotics for the additive model (26), we will simply call the f_i's *separable* if $\hat{\theta}_{in}$ has the same limiting distribution as $\tilde{\theta}_{in}$ for $i = 1,\ldots,m$.

Theorem 6.4 *Let* f *be given as in* (26) *with* $m = 2$. *If* f_1 *and* f_2 *satisfy the conditions respectively in* Theorem 5.3 *(or* 6.2*) and* Theorem 6.1*, and if* Assumption 2.2 *holds, then* f_1 *and* f_2 *are separable.*

Theorem 6.5 *Let* f *be given as in* (26) *with* $m = 2$. *If* f_1 *and* f_2 *satisfy the conditions respectively in* Theorem 5.3 *(or* 6.2*) and* Theorem 5.4*, and if Assumption* 2.3 *holds, then* f_1 *and* f_2 *are separable.*

Example 6.1 For the partially linear model in which

$$f(x,\alpha,\beta) = \alpha x + g(x,\beta)$$

with I-regular g satisfying the conditions of Theorem 6.1, the result in Theorem 6.2 is applicable. Therefore, under Assumptions 2.2, we have

$$\sqrt{n}(\hat{\alpha} - \alpha_0) \to_d \quad \left(\int_0^1 V^2\right)^{-1} \int_0^1 V\,dU,$$

$$\sqrt[4]{n}(\hat{\beta}-\beta_0) \to_d \qquad \left(L(1,0)\int_{-\infty}^{\infty}\dot{g}(s,\beta_0)\dot{g}(s,\beta_0)'ds\right)^{-1/2} W(1),$$

which can be easily obtained from the results in Theorems 5.1 and 5.2, respectively for the regressions on αx and $g(x,\beta)$ alone.

Thus far, we have assumed that there are no restrictions on θ_i's. But, models with restrictions on θ_i's can also be considered within our framework. Here we just look at one such model which seems to appear most frequently in nonlinear analyses. Let $m=2$ in (26), and write the parameters θ_1 and θ_2 more explicitly as

$$\theta_1 = \alpha \quad \text{and} \quad \theta_2 = (\alpha', \beta')'$$

Also, let $\theta = (\alpha', \beta')'$ and $\theta_0 = (\alpha_0', \beta_0')'$. The most important case seems to be the model where f_1 is H$_0$-regular and f_2 is I-regular. Some examples will be given below. Define

$$f_2^*(x,\beta) = f_2(x,\alpha_0,\beta)$$

and $f_* = f_1 + f_2^*$. Then, we have

Theorem 6.6 *Let* Assumption 2.2 *hold. Assume*
(a) f_1 *satisfies the conditions in* Theorem 6.2,
(b) f_2 *and* f_2^* *satisfy, respectively,* (a) *and* (b) *of* Theorem 6.1*, and*
(c) $\lambda^{1/2}\dot{\kappa}_1(\lambda) \to \infty$.
Then regression on f is asymptotically equivalent to that on f_*, for which f_1 and f_2^* are separable.

Example 6.2 For the logistic regression function $f(x,\alpha,\beta) = \alpha e^{\beta x}/(1+e^{\beta x})$ with two parameters α and β, we may write

$$f(x,\alpha,\beta) = f_1(x,\alpha) + f_2(x,\alpha,\beta)$$

where

$$\begin{aligned} f_1(x,\alpha) &= \alpha 1\{x \geq 0\} \\ f_2(x,\alpha,\beta) &= \frac{\alpha e^{\beta x}}{1+e^{\beta x}} 1\{x<0\} - \frac{\alpha}{1+e^{\beta x}} 1\{x \geq 0\} \end{aligned}$$

to apply Theorem 6.6. It is straightforward to check that the conditions of Theorem 6.6 are satisfied for such an f. We therefore have

$$\begin{aligned} \sqrt{n}(\hat{\alpha}-\alpha_0) &\to_d \quad \left(\int_0^1 1\{V\geq 0\}\right)^{-1}\int_0^1 1\{V \geq 0\}\, dU \\ \sqrt[4]{n}(\hat{\beta}-\beta_0) &\to_d \quad \left(\frac{\alpha_0^2(\pi^2-6)}{18\beta_0^3}L(1,0)\right)^{-1/2} W(1) \end{aligned}$$

as $n \to \infty$.

Other types of regressions with additive regression functions as in (26) can be analyzed in a similar way. Here we have just presented results for models which seem most frequently to appear in practical applications. It is pretty straightforward to show that the usual error variance estimator $\hat{\sigma}_n^2$ in the additive models is consistent, following the proof of Corollary 5.5. Also, the limit theory for hypothesis testing in additive models follows as before. Indeed, Lemma 6.3 establishes the asymptotic equivalence of the nonlinear regression (6) with the additive regression function (26) and the *linear* regression with the regressors $(\dot{f}_1(x_t, \theta_{10}), \ldots, \dot{f}_m(x_t, \theta_{m0}))$ and regression coefficients $(\theta_1, \ldots, \theta_m)$. The statistical limit theories for the estimators and test statistics in such linear regressions follow directly from our results in Section 3.

7. 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 $\left(n^{\frac{1}{4}}\right)$ 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. In most cases and for most regression functions, the limit distributions of the nonlinear regression estimators are mixed normal and are always so in the cases we consider when 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.

Our purpose in this paper has been to initiate nonlinear econometric analysis for stochastically nonstationary time series. 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 empirical studies of nonlinear cointegrating links between economic time series. A combination of the ideas presented here and those in our other paper, Phillips and Park (1998), will form the basis of a nonparametric analysis of cointegration and the authors plan to report on this further extension at a later date.

8. Appendix A: Technical Results for I(1) Functionals

8.1 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 Section 8.2 below.

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) I*f $F(\cdot,\pi)$ 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) I*f 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) $\frac{1}{n}\sum_{t=1}^{n} T\left(\frac{x_t}{\sqrt{n}}\right) = O_{a.s.}(1)$ for $T \in \mathcal{T}_{LB}$.
(b) $\frac{1}{n}\sum_{t=1}^{n} T_1\left(\frac{x_t}{\sqrt{n}}\right)T_2(x_t) = o_{a.s.}(1)$ for $T_1 \in \mathcal{T}_{LB},\ T_2 \in \mathcal{T}_B^0$.
(c) $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} T\left(\frac{x_t}{\sqrt{n}}\right) u_t = O_p(1)$ for $T \in \mathcal{T}_{LB}^0$.
(d) $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} T_1\left(\frac{x_t}{\sqrt{n}}\right) T_2(x_t)u_t = o_p(1)$ for $T_1 \in \mathcal{T}_{LB}^0$, $T_2 \in \mathcal{T}_B^0$.

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* $\sup_{\omega\in\Phi} W(\cdot,\omega) \in \mathcal{T}_{LB}^0$. *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) I*f 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 $H(\cdot, \pi) = H_1(\cdot, \pi_1) \otimes H_2(\cdot, \pi_2)$.*

(d) *If F_i is* E*-regular on compact Π_i with asymptotic order κ_i and limit exponential function E_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 exponential function $E(\cdot, \pi) = E_1(\cdot, \pi_1) \otimes E_2(\cdot, \pi_2)$.*

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 $\pi \in \Pi$.*
(d) *Let* Assumption 2.3 *hold. If F is E-regular on a compact set Π, then for large n $n^{-1/2}\kappa(\max x_t, \pi)^{-1}F(x_t, \pi)\, u_t = o_p(1)$ 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.*

8.2 Proofs

Proof of Lemma A1 It is obvious that $T_1 \pm T_2$ and $T_1 T_2$ 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

$$\begin{aligned} \underline{T}_\varepsilon &= \underline{T}_{1\varepsilon} + \underline{T}_{2\varepsilon},\ \underline{T}_{1\varepsilon} - \overline{T}_{2\varepsilon} \\ \overline{T}_\varepsilon &= \overline{T}_{1\varepsilon} + \overline{T}_{2\varepsilon},\ \overline{T}_{1\varepsilon} - \underline{T}_{2\varepsilon} \end{aligned}$$

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 $T = T_1 T_2$, 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. ■

Proof of Lemma A2 The first part is due to Park and Phillips (1997). To prove the second part, first let $s_m = \max(s_{\max}, -s_{\min}) + 1$. 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),

$$\int_0^1 T_\varepsilon(V_n)\, dU_n \to_d \int_0^1 T_\varepsilon(V)\, dU \tag{29}$$

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

$$\left| \int_0^1 T(V_n)\, dU_n - \int_0^1 T_\varepsilon(V_n)\, dU_n \right| \to_p 0, \tag{30}$$

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

$$\begin{aligned} A_{n\varepsilon} &= \int_0^1 \left(\overline{T}_\varepsilon(V_n) - \underline{T}_\varepsilon(V_n)\right)^2 1\{|V_n| \le c\}, \\ B_{n\varepsilon} &= \int_0^1 \left(\overline{T}_\varepsilon(V_n) - \underline{T}_\varepsilon(V_n)\right)^2 1\{|V_n| > c\}. \end{aligned}$$

Then we have

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

The result in (30) 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,

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

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

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

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

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

which, as we noted earlier, can be made arbitrarily small by taking c large. We now have (30), which along with (29), completes the proof. ■

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 $\underline{T}^0_\varepsilon, \overline{T}^0_\varepsilon$ and $\delta_\varepsilon > 0$ satisfying

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

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

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

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

We now let

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

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

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

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. ■

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

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

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

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

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

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

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

as $n \to \infty$. This is shown in Park and Phillips (1997). To show part (b), we note that

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

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

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

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

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

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

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

which completes the proof. ■

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

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

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

$$\begin{aligned}
\left\| \frac{1}{n} z_n(\omega)^{-1} \sum_{t=1}^n Z_{nt}(\omega) \right\| &\leq \|z_n(\omega)^{-1} a_n(\omega)\| \left(\frac{1}{n} \sum_{t=1}^n T\left(\frac{x_t}{\sqrt{n}} \right) \right) \\
&\quad \text{or } \|z_n(\omega)^{-1} b_n(\omega)\| \left(\frac{1}{n} \sum_{t=1}^n T\left(\frac{x_t}{\sqrt{n}} \right) S(x_t) \right),
\end{aligned}$$

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

$$\begin{aligned}
\left\| \frac{1}{n} z_n(\omega)^{-1} \sum_{t=1}^n Z_{nt}(\omega) u_t \right\| &\leq \|z_n(\omega)^{-1} a_n(\omega)\| \left(\frac{1}{n} \sum_{t=1}^n T\left(\frac{x_t}{\sqrt{n}} \right) |u_t| \right) \\
&\quad \text{or } \|z_n(\omega)^{-1} b_n(\omega)\| \left(\frac{1}{n} \sum_{t=1}^n T\left(\frac{x_t}{\sqrt{n}} \right) S(x_t) |u_t| \right),
\end{aligned}$$

and subsequently observe that

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

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

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

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

where we set $S_i(x) = \sup_{\pi_i \in \Pi_i} \|F_i(x,\pi_i)\|$ 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

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

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

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

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

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

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).

Finally, for part (d), we let

$$F_i(\lambda x, \pi_i) = \kappa_i(\lambda s, \pi_i)E_i(\lambda(x-s), \pi_i) + R_i(x-s, \lambda, s, \pi_i),$$

for $i = 1, 2$. If we let

$$\kappa(\cdot, \pi) = \kappa_1(\cdot, \pi_1) \otimes \kappa_2(\cdot, \pi_2) \quad \text{and} \quad E(\cdot, \pi) = E_1(\cdot, \pi_1) \otimes E_2(\cdot, \pi_2),$$

then it follows that

$$F(\lambda x, \pi) = \kappa(\lambda s, \pi) E(\lambda(x - s), \pi) + R(x - s, \lambda, s, \pi),$$

with

$$\begin{aligned} R(x-s, \lambda, s, \pi) \;=\; & R_1(x-s, \lambda, s, \pi_1) \otimes R_2(x-s, \lambda, s, \pi_2) \\ & + \kappa_1(\lambda s, \pi_1) E_1\left(\lambda(x-s), \pi_1\right) \otimes R_2(x-s, \lambda, s, \pi_2) \\ & + \kappa_2(\lambda s, \pi_2) E_2\left(\lambda(x-s), \pi_2\right) \otimes R_1(x-s, \lambda, s, \pi_1). \end{aligned}$$

We may now easily deduce from Lemma A5(a) that $R_-(x - s, \lambda, s, \pi)$ is of order smaller than $\kappa(\lambda s, \pi)/\lambda$, since E_{i-}, which is defined from E_i in a similar way as E_-, is I-regular for $i = 1, 2$, and hence bounded. ■

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

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

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

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} F\left(\frac{x_t}{\sqrt{n}}, \pi_0\right) u_t = O_p(1), \tag{32}$$

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 arbirarily small a.s. uniformly in $\pi \in N_0$, which we now set out to do.

Using Cauchy–Schwarz we have

$$\begin{aligned} & \left| \frac{1}{n} \sum_{t=1}^{n} \left(F\left(\frac{x_t}{\sqrt{n}}, \pi\right) - F\left(\frac{x_t}{\sqrt{n}}, \pi_0\right)\right) u_t \right| \\ & \quad \leq \left(\frac{1}{n} \sum_{t=1}^{n} \left(F\left(\frac{x_t}{\sqrt{n}}, \pi\right) - F\left(\frac{x_t}{\sqrt{n}}, \pi_0\right)\right)^2 \right)^{1/2} \left(\frac{1}{n} \sum_{t=1}^{n} u_t^2 \right)^{1/2}. \end{aligned} \tag{33}$$

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

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

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 (33) that there exists a neighborhood N_0 of π_0 such that

$$\sup_{\pi \in N_0} \left| \frac{1}{n} \sum_{t=1}^n \left(F\left(\frac{x_t}{\sqrt{n}}, \pi\right) - F\left(\frac{x_t}{\sqrt{n}}, \pi_0\right) \right) u_t \right| \; < \; \varepsilon \;\; \text{a.s.} \tag{34}$$

for any $\varepsilon > 0$ given. We may now easily deduce (31) from the results in (32) and (34). 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

$$\sup_{\pi \in N_0} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n F(x_t,\pi) \, u_t \right| = o_p(1). \tag{35}$$

Since it follows from Theorem 3.2 that

$$\frac{1}{\sqrt[4]{n}} \sum_{t=1}^n F(x_t,\pi_0) \, u_t = O_p(1), \tag{36}$$

it suffices to show that

$$\sup_{\pi \in N_0} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n (F(x_t,\pi) - F(x_t,\pi_0)) \, u_t \right| = o_p(1) \tag{37}$$

to deduce (35).

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

$$\sum_{t=1}^n |F(x_t,\pi) - F(x_t,\pi_0)||u_t| \le \|\pi - \pi_0\| \left(\sigma \sum_{t=1}^n |T(x_t)| + \sum_{t=1}^n |T(x_t)| w_t \right), \tag{38}$$

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

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

and

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

It is therefore clear from (38) that we may choose N_0 such that (37) 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,

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

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.

To prove part (d), we observe that

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

uniformly in $\pi \in \Pi$, which is due to Lemma A5(b). The rest of the proof is essentially identical to that of part (b), and is omitted. ■

Proof of Lemma A8 For the proof of part (a), it suffices to show that

$$\int_{-\infty}^{\infty}F(s,\cdot)L(1,s)ds \tag{39}$$

is continuous a.s., due to the occupation time formula. The continuity of (39), 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 $\sup_{\pi\in\Pi}|F(\cdot,\pi)|$. ■

9. 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 $(\Omega, \mathcal{F}, \Pr)$ 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 $(\Omega, \mathcal{F}, \Pr)$, from which the desired (U_n°, V_n°) is then defined. In the subsequent construction, we let

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

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

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

$$U\left(\frac{\tau_{nt}}{n}\right)\Big|\, \mathcal{F}_{n,t-1}^\circ =_d U_n\left(\frac{t}{n}\right)\Big|\, \mathcal{F}_{n,t-1} \quad \text{and} \quad V_{nt}|\, \mathcal{G}_{nt}^\circ =_d V_n\left(\frac{t}{n}\right)\Big|\, \mathcal{G}_{nt}$$

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

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

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

It is shown by Park and Phillips (1997) and Phillips and Ploberger (1996) that we may choose the stopping times τ_{nt} so that they satisfy condition (9). In particular, it follows from the Hölder continuity of the sample path of U that

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

a.s., for some constant c and any $\varepsilon > 0$. Now we may easily deduce from (9) that

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

for $2/q < \delta < 1$ and any $\varepsilon > 0$. In particular, $U_n^\circ \to_{a.s.} U$ uniformly on $[0,1]$. Moreover, since $(U_n^\circ, V_n^\circ) \to_d (U, V)$, we may redefine V_n°, if necessary, so that the distribution of (U_n°, V_n°) is unchanged and $V_n^\circ \to_{a.s.} V$ 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). ■

Proof of Lemma 2.2 See corollary 1.6, p. 215, of Revuz and Yor (1994). ■

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

$$\int_0^1 F(V_n(r),\pi)dr \to_{a.s.} \int_0^1 F(V(r),\pi)dr, \tag{40}$$

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

$$\int_0^1 \sup_{\pi\in N} F(V_n(r),\pi)dr \;\to_{a.s.}\; \int_0^1 \sup_{\pi\in N} F(V(r),\pi)dr, \tag{41}$$

$$\int_0^1 \inf_{\pi\in N} F(V_n(r),\pi)dr \;\to_{a.s.}\; \int_0^1 \inf_{\pi\in N} F(V(r),\pi)dr, \tag{42}$$

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

$$\begin{aligned}
&\int_0^1 \left(\sup_{\pi\in N} F(V(r),\pi) - \inf_{\pi\in N} F(V(r),\pi)\right) dr \\
&\quad = \int_{-\infty}^{\infty} \left(\sup_{\pi\in N_\delta} F(s,\pi) - \inf_{\pi\in N_\delta} H(s,\pi)\right) L(1,s)\, ds \to_{a.s.} 0,
\end{aligned} \tag{43}$$

as $\delta \to 0$. We may now easily deduce from (41)–(43) that there exists a neighborhood of π_0 where (40) holds uniformly in π. Since π_0 was chosen arbitrary and Π is compact, (40) holds uniformly on Π, as was to be shown. The sample covariance asymptotics are given in Lemma A2. ■

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

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sup_{\pi\in N}F(x_t,\pi)\ \to_p\ \left(\int_{-\infty}^{\infty}\sup_{\pi\in N}F(s,\pi)\,ds\right)L(1,0), \tag{44}$$

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\inf_{\pi\in N}F(x_t,\pi)\ \to_p\ \left(\int_{-\infty}^{\infty}\inf_{\pi\in N}F(s,\pi)\,ds\right)L(1,0), \tag{45}$$

due to Theorem 5.1 of Park and Phillips (1997).

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,

$$\int_{-\infty}^{\infty}\sup_{\pi\in N_\delta}F(s,\pi)\,ds-\int_{-\infty}^{\infty}\inf_{\pi\in N_\delta}F(s,\pi)\,ds\to 0, \tag{46}$$

as $\delta\to 0$. We may now easily deduce from (44)–(46) 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

$$\begin{aligned} M_n(r) \;=\; & \sqrt[4]{n}\sum_{t=1}^{k-1}F\left(\sqrt{n}V_n\left(\frac{t-1}{n}\right),\pi\right)\left(U\left(\frac{\tau_{nt}}{n}\right)-U\left(\frac{\tau_{n,t-1}}{n}\right)\right) \\ & +\sqrt[4]{n}F\left(\sqrt{n}V_n\left(\frac{k-1}{n}\right),\pi\right)\left(U(r)-U\left(\frac{\tau_{n,k-1}}{n}\right)\right), \end{aligned} \tag{47}$$

for $\tau_{n,k-1}/n<r\le\tau_{nk}/n$, where τ_{nk}, $k=1,\ldots,n$, are the stopping times introduced in Lemma 1.2. One may easily see that M_n is a continuous martingale such that

$$\frac{1}{\sqrt[4]{n}}\sum_{t=1}^{n}F(x_t,\pi)\,u_t\ =M_n\left(\frac{\tau_{nn}}{n}\right), \tag{48}$$

and that

$$\sup_{1\le t\le n}\left|\left(\frac{\tau_{nt}}{n}-\frac{\tau_{n,t-1}}{n}\right)-\frac{1}{n}\right|=o_{a.s.}(1), \tag{49}$$

by Lemma 1.2.

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

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

due to (49), and therefore,

$$[M_n]_r \to_p \left(\int_{-\infty}^{\infty} F(s,\pi)^2 ds\right) L(r,0), \tag{50}$$

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

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

uniformly in $r \in [0,1]$, due to (49). However, for all $r \in [0,1]$,

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

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

$$[M_n, V]_{\rho_n(r)} \to_p 0, \tag{51}$$

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 (47) is completely determined by (50) and (51), 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 (48), completes the proof for the second part. ■

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. ■

Proof of Theorem 3.4 Let x_t^* be as in Assumption 2.3, and define $s_n=\max_{1\le t\le n}\frac{x_t}{\sqrt{n}}$. We have

$$\begin{aligned}&\frac{1}{\sqrt{n}}\kappa\left(\max_{1\le t\le n}x_t,\pi\right)^{-1}\sum_{t=1}^{n}F(x_t,\pi)\\&=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}E(x_t^*,\pi)+\sqrt{n}\kappa(\sqrt{n}s_n,\pi)^{-1}\frac{1}{n}\sum_{t=1}^{n}R\left(\frac{x_t^*}{\sqrt{n}},\sqrt{n},s_n,\pi\right)\\&=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}E(x_t^*,\pi)+o_{a.s.}(1),\end{aligned}$$

uniformly in $\pi\in\Pi$. This is because $s_n\to_{a.s.}s_{\max}>0$ and

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

uniformly in $\pi\in\Pi$, by Lemma A5(b). It therefore suffices to show

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}E(x_t^*,\pi)\to_p\left(\int_{-\infty}^{0}E(s,\pi)\,ds\right)L(1,s_{\max}),\qquad(52)$$

uniformly in $\pi\in\Pi$, for the proof of sample mean asymptotics.

To deduce (52), we first define $V_n^*(r) = V_n(r) - s_n$ and write

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^n E(x_t^*, \pi) &= \sqrt{n}\int_0^1 E(\sqrt{n}V_n^*(r), \pi)\, dr \\ &= \sqrt{n}\int_0^1 E_-(\sqrt{n}V_n^*(r), \pi)\, dr.\end{aligned}$$

However, by Theorem 3.2,

$$\sqrt{n}\int_0^1 E_-(\sqrt{n}V_n^*(r), \pi)\, dr \to_p \left(\int_{-\infty}^{\infty} E_-(s, \pi)\, ds\right) L^*(1, 0), \tag{53}$$

uniformly in $\pi \in \Pi$, where L^* is the local time of the process V^* given by $V^*(r) = V(r) - s_{\max}$. As noted by Park and Phillips (1997), we have the same result for V_n^* and V^* as for V_n and V. Notice that

$$\begin{aligned}L^*(1, 0) &= L(1, s_{\max}) \\ \int_{-\infty}^{\infty} E_-(s, \pi)\, ds &= \int_{-\infty}^{0} E(s, \pi)\, ds,\end{aligned}$$

to deduce (52) from (53).

For sample covariance asymptotics, we write

$$\begin{aligned}&\frac{1}{\sqrt{[4]n}}\kappa\left(\max_{1\le t\le n} x_t, \pi\right)^{-1}\sum_{t=1}^n F(x_t, \pi)\, u_t \\ &= \frac{1}{\sqrt[4]{n}}\sum_{t=1}^n E(x_t^*, \pi)\, u_t + \sqrt[4]{n}\kappa(\sqrt{n}s_n, \pi)^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^n R\left(\frac{x_t^*}{\sqrt{n}}, \sqrt{n}, s_n, \pi\right) u_t \\ &= \frac{1}{\sqrt[4]{n}}\sum_{t=1}^n E(x_t^*, \pi)\, u_t + o_{a.s.}(1),\end{aligned}$$

upon noticing that

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

due to Lemma A5(b). The rest of the proof is identical to the second part of the proof of Theorem 3.2. Notice that V_n^* and V^* behave under Assumption 2.3 in exactly the same way as V_n and V do under Assumption 2.2. ■

Proof of Theorem 4.1 It follows readily from Lemma A7(b) that

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

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). ■

Proof of Theorem 4.2 Due to Lemma A7(c), we have

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

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

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

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). ■

Proof of Theorem 4.3 Let

$$\begin{aligned}m(\sqrt{n},\theta)^2 &= \frac{1}{n\kappa(\sqrt{n},\theta)^2}\sum_{t=1}^{n}f(x_t,\theta)^2,\\ m(\theta)^2 &= \int_{-\infty}^{\infty}h(s,\theta)^2L(1,s)\,ds.\end{aligned}$$

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

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

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,

$$\inf_{\substack{|p-\bar p|<\varepsilon\\|q-\bar q|<\varepsilon}}\inf_{\theta\in N}|\,p\kappa(\lambda,\theta)-q\kappa(\lambda,\theta_0)\,|\leq\left|\,\kappa(\sqrt{n},\theta)m(\sqrt{n},\theta)-\kappa(\sqrt{n},\theta_0)m(\sqrt{n},\theta_0)\right| \quad (54)$$

for large n.

Define

$$\begin{aligned}A_n(\theta,\theta_0) &= \frac{1}{n}\sum_{t=1}^{n}\left(f(x_t,\theta)-f(x_t,\theta_0)\right)^2,\\ B_n(\theta,\theta_0) &= \frac{1}{n}\sum_{t=1}^{n}\left(f(x_t,\theta)-f(x_t,\theta_0)\right)u_t.\end{aligned}$$

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

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

and, therefore,

$$\left(A_n^{-1}|B_n|\right)(\theta,\theta_0) \le A_n(\theta,\theta_0)^{-1/2}\left(\sigma^2 + o_p(1)\right), \tag{55}$$

uniformly in $\theta \in \Theta$. However, using (54) and the inequality

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

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

$$A_n(\theta,\theta_0) \ge \left(\kappa(\sqrt{n},\theta)m(\sqrt{n},\theta) - \kappa(\sqrt{n},\theta_0)m(\sqrt{n},\theta_0)\right)^2 \to_{a.s.} \infty, \tag{56}$$

uniformly in $\theta \in N$.

Now it follows from (55) and (56) that

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

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. ■

Proof of Theorem 4.4 Let

$$\begin{aligned} m(\sqrt{n},\theta)^2 &= \frac{1}{\sqrt{n}\kappa(\sqrt{n}s_n,\theta)^2}\sum_{t=1}^n f(x_t,\theta)^2 \\ m(\theta)^2 &= \left(\int_{-\infty}^0 e(s,\theta)^2\,ds\right) L(1,s_{\max}), \end{aligned}$$

where s_n is defined in the proof of Theorem 3.4. In view of Lemma A6(d) and Theorem 3.4, we have

$$m\left(\sqrt{n},\theta\right) \to_p m(\theta),$$

uniformly in $\theta \in \Theta$. It follows from Lemma A8(b) that m is continuous, and since $L(1,s_{\max}) > 0$ a.s., $m > 0$ a.s. for all $\theta \in \Theta$ when and only when condition (b) holds.

We let A_n and N be defined as in the proof of Theorem 4.3. Then we have

$$A_n(\theta,\theta_0) \ge n^{-1/2}\left(\kappa(\sqrt{n}s_n,\theta)\,m(\sqrt{n},\theta) - \kappa(\sqrt{n}s_n,\theta_0)m(\sqrt{n},\theta_0)\right)^2 \to_{a.s.} \infty,$$

uniformly in $\theta \in N$, by condition (a). Note that $s_n \to_{a.s.} s_{\max} > 0$ a.s. The rest of the proof is essentially identical to the proof of Theorem 4.3 and is omitted. ■

Proof of Corollary 4.5 Let

$$\sigma_n^2 = \frac{1}{n}\sum_{t=1}^{n} u_t^2. \tag{57}$$

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 $\kappa_n^2 = O_p(1)$. ■

Proof of Theorem 4.6 Redefine

$$Q_n(\theta) = \sum_{t=1}^{n} \left(y_{nt} - f(x_{nt},\theta)\right)^2,$$

and $D_n(\theta,\theta_0) = Q_n(\theta) - Q_n(\theta_0)$. Also, let

$$D(\theta,\theta_0) = \int_{-\infty}^{\infty} \left(f_0(s,\theta) - f_0(s,\theta_0)\right)^2 L(1,s)\, ds,$$

where $f_0(x,\theta) = f(\sqrt{n}_0 x,\theta)$. By successive application of Lemma A7(a), Lemma A6(a), Theorem 3.1 and the occupation formula, we obtain

$$\frac{1}{n}D_n(\theta,\theta_0) = \frac{1}{n}\sum_{t=1}^{n} \left(f(x_{nt},\theta) - f(x_{nt},\theta_0)\right)^2 + o_p(1) \to_p D(\theta,\theta_0),$$

uniformly in $\theta \in \Theta$. Moreover, as in the proof of Theorem 4.2, we may deduce that $D(\cdot,\theta_0)$ has unique minimum θ_0 when and only when the given identifying condition is satisfied. We have therefore shown that CN1 holds, which is sufficient to establish the consistency of $\bar{\theta}_n$. The consistency of $\bar{\sigma}^2$ follows immediately from the uniform convergence of $n^{-1}D_n(\cdot,\theta_0)$ to $D(\cdot,\theta_0)$, which is continuous a.s. due to Lemma A8(a).
■

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

$$\ddot{Q}_n(\theta)=\sum_{t=1}^n \dot{f}(x_t,\theta)\dot{f}(x_t,\theta)'+\sum_{t=1}^n \ddot{F}(x_t,\theta)\left(f(x_t,\theta)-f(x_t,\theta_0)\right)-\sum_{t=1}^n \ddot{F}(x_t,\theta)u_t. \quad (58)$$

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

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n \dot{f}(x_t,\theta)\dot{f}(x_t,\theta)' \to_p L(1,0)\int_{-\infty}^{\infty} \dot{f}(s,\theta)\dot{f}(s,\theta)'ds, \quad (59)$$

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

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n \ddot{f}(x_t,\theta)\,u_t \to_p 0, \quad (60)$$

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

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

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 (59) and (60), completes the proof. ■

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,

$$\ddot{Q}(\theta_0) = \int_0^1 \dot{h}(V(r),\theta_0)\dot{h}(V(r),\theta_0)'dr = \int_{-\infty}^{\infty} \dot{h}(s,\theta_0)\dot{h}(s,\theta_0)'L(1,s)\,ds, \quad (61)$$

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

$$\nu_n^{-1}\sum_{t=1}^n \dot{f}(x_t,\theta)\dot{f}(x_t,\theta)'\nu_n^{-1\prime} \to_{a.s} \int_0^1 \dot{h}(V(r),\theta)\dot{h}(V(r),\theta)'dr, \quad (62)$$

and from Lemma A7(c) that

$$(\nu_n \otimes \nu_n)^{-1} \sum_{t=1}^{n} \ddot{f}(x_t, \theta)\, u_t = (\dot{\kappa}_n \otimes \dot{\kappa}_n)^{-1} \ddot{\kappa}_n \left(\frac{\ddot{\kappa}_n^{-1}}{n} \sum_{t=1}^{n} \ddot{f}(x_t, \theta)\, u_t \right) \to_p 0, \qquad (63)$$

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

$$\begin{aligned} &(\nu_n \otimes \nu_n)^{-1} \sum_{t=1}^{n} \ddot{f}(x_t, \theta_n)\, (f(x_t, \theta_n) - f(x_t, \theta_0)) \qquad (64) \\ &\quad = \left((\dot{\kappa}_n \otimes \dot{\kappa}_n)^{-1} \kappa_n \ddot{\kappa}_n \right) \frac{1}{n} \sum_{t=1}^{n} \left(\ddot{\kappa}_n^{-1} \ddot{f}(x_t, \theta_n) \right) \left(\kappa_n^{-1} \left(f(x_t, \theta_n) - f(x_t, \theta_0) \right) \right) = o_p(1). \end{aligned}$$

It is easily seen from (58) that (62)-(64) imply AD6.

To prove (64), 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

$$\begin{aligned} &\left\| \frac{1}{n} \sum_{t=1}^{n} \left(\ddot{\kappa}_n^{-1} \ddot{f}(x_t, \theta) \right) \left(\kappa_n^{-1} \left(f(x_t, \theta) - f(x_t, \theta_0) \right) \right) \right\| \\ &\qquad \leq \left\| \ddot{h} \right\|_K \frac{1}{n\kappa_n} \sum_{t=1}^{n} |f(x_t, \theta) - f(x_t, \theta_0)| \\ &\qquad \to_{a.s.} \left\| \ddot{h} \right\|_K \int_0^1 |h(V(r), \theta) - h(V(r), \theta_0)|\, dr, \end{aligned}$$

uniformly in $\theta \in \Theta$, where $K = [s_{\min} - 1, s_{\max} + 1] \times \Theta$. To deduce (64), simply note that the limit function is continuous in θ, due to Lemma A8(a). ■

Proof of Theorem 5.3 We show that AD1–AD4 and AD7 hold to establish (14). 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 (15)

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

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 (61), 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

$$\ddot{Q}_n(\theta) - \ddot{Q}_n(\theta_0) = \left(\ddot{D}_{1n}(\theta) + \ddot{D}_{1n}(\theta)' \right) + \ddot{D}_{2n}(\theta) + \ddot{D}_{3n}(\theta) + \ddot{D}_{4n}(\theta), \qquad (65)$$

where

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

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

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

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

and define

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

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

$$\varpi_{1n}^2(\theta) \leq \sum_{t=1}^{n} \left\|\mu_n^{-1}\dot{f}(x_t,\theta_0)\right\| \left\|(\mu_n \otimes \mu_n)^{-1}\ddot{f}(x_t,\bar{\theta})\right\|, \tag{66}$$

$$\varpi_{2n}^2(\theta) \leq \sum_{t=1}^{n} \left\|(\mu_n \otimes \mu_n)^{-1}\ddot{f}(x_t,\bar{\theta})\right\|^2, \tag{67}$$

$$\varpi_{3n}^2(\theta) \leq \sum_{t=1}^{n} \left\|\mu_n^{-1}\dot{f}(x_t,\theta_0)\right\| \left\|(\mu_n \otimes \mu_n)^{-1}\ddot{f}(x_t,\theta)\right\| + \frac{1}{2}\sum_{t=1}^{n} \left\|(\mu_n \otimes \mu_n)^{-1}\ddot{f}(x_t,\bar{\theta})\right\| \left\|(\mu_n \otimes \mu_n)^{-1}\ddot{f}(x_t,\theta)\right\|, \tag{68}$$

$$\varpi_{4n}^2(\theta) \leq \sum_{t=1}^{n} \left\|(\mu_n \otimes \mu_n \otimes \mu_n)^{-1}\dddot{f}(x_t,\bar{\theta})\right\| |u_t|, \tag{69}$$

where $\bar{\theta}$ 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 (66)–(69) that

$$\varpi_{1n}^2(\theta) \leq \frac{n^{3\delta}}{\sqrt{n}}\left\|(\dot{\kappa}_n \otimes \dot{\kappa}_n)^{-1}\left(\sup_{|s|\leq\bar{s}} \sup_{\theta\in N_n} |\ddot{f}(\sqrt{n}s,\theta)|\right)\right\| \frac{1}{n}\sum_{t=1}^{n}\left\|\dot{\kappa}_n^{-1}\dot{f}(x_t,\theta_0)\right\|, \tag{70}$$

$$\varpi_{2n}^2(\theta) \leq \frac{n^{4\delta}}{n}\left\|(\dot{\kappa}_n \otimes \dot{\kappa}_n)^{-1}\left(\sup_{|s|\leq\bar{s}} \sup_{\theta\in N_n} |\ddot{f}(\sqrt{n}s,\theta)|\right)\right\|^2, \tag{71}$$

$$\varpi_{3n}^2(\theta) \;\le\; \frac{n^{3\delta}}{\sqrt{n}}\left\|(\dot\kappa_n\otimes\dot\kappa_n)^{-1}\left(\sup_{|s|\le\bar s}\sup_{\theta\in N_n}|\ddot f(\sqrt{n}s,\theta)|\right)\right\|\frac{1}{n}\sum_{t=1}^n\left\|\dot\kappa_n^{-1}\dot f(x_t,\theta_0)\right\|$$
$$+\frac{n^{4\delta}}{2n}\left\|(\dot\kappa_n\otimes\dot\kappa_n)^{-1}\left(\sup_{|s|\le\bar s}\sup_{\theta\in N_n}|\ddot f(\sqrt{n}s,\theta)|\right)\right\|^2, \tag{72}$$
$$\varpi_{4n}^2(\theta) \;\le\; \frac{n^{3\delta}}{\sqrt{n}}\left\|(\dot\kappa_n\otimes\dot\kappa_n\otimes\dot\kappa_n)^{-1}\left(\sup_{|s|\le\bar s}\sup_{\theta\in N_n}|\dddot f(\lambda s,\theta)|\right)\right\|\frac{1}{n}\sum_{t=1}^n|u_t|, \tag{73}$$

from which we may easily deduce that $\varpi_{in}^2(\theta)=o_{a.s.}(1)$, $i=1,\ldots,4$, uniformly in $\theta\in N_n$, due to (16) and (17). Now AD7 follows immediately from (70) – (73). This completes the proof. ∎

Proof of Theorem 5.4 The proof is entirely analogous to that of Theorem 5.3. We show AD1–AD4 and AD7 to establish (14). It follows directly from Theorem 3.4 that AD1 holds. For AD2, note that

$$\left\|(\nu_n\otimes\nu_n)^{-1}\sum_{t=1}^n\ddot f(x_t,\theta_0)\,u_t\right\| \;\le\; \sqrt{n}\left\|(\dot\kappa_n\otimes\dot\kappa_n)^{-1}\left(\sup_{|s|\le\bar s}|\ddot f(\sqrt{n}s,\theta_0)|\right)\right\|$$
$$\frac{1}{n}\sum_{t=1}^n|u_t|\to_{a.s.}0,$$

which follows from (21). We then have from Lemma A6(c) and Theorem 3.4 that AD3 holds with

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

Therefore, in particular, AD4 holds under condition (c).

We now show AD7. Let $\bar s=s_{\max}$, and let s_n be defined as in the proof of Theorem 3.4. We have $s_n\to_{a.s.}\bar s$ as $n\to\infty$. Fix δ such that $0<\delta<\varepsilon/3$, and write for notational brevity $\dot\kappa_n=\dot\kappa_0(\sqrt{n}s_n)$. Define $\mu_n=n^{1/4-\delta}\dot\kappa_n$ and $\nu_n=n^{1/4}\dot\kappa_n$. Obviously, $\mu_n\nu_n^{-1}\to_{a.s.}0$. Moreover, if we let N_n be defined as in AD7, then

$$N_n\subset N(\bar s,\varepsilon,\sqrt{n}),$$

for all large n.

We decompose $\ddot Q_n(\theta)-\ddot Q_n(\theta_0)$ as in (65). We have from (66)–(69) that for all $\theta\in N_n$

$$\varpi_{1n}^2(\theta) \;\le\; \frac{n^{3\delta}}{\sqrt[4]{n}}\left\|(\dot\kappa_n\otimes\dot\kappa_n)^{-1}\left(\sup_{s\le\bar s+\varepsilon}\sup_{\theta\in N_n}|\ddot f(\sqrt{n}s,\theta)|\right)\right\|\frac{1}{\sqrt{n}}\sum_{t=1}^n\left\|\dot\kappa_n^{-1}\dot f(x_t,\theta_0)\right\|, \tag{74}$$
$$\varpi_{2n}^2(\theta) \;\le\; n^{4\delta}\left\|(\dot\kappa_n\otimes\dot\kappa_n)^{-1}\left(\sup_{s\le\bar s+\varepsilon}\sup_{\theta\in N_n}|\ddot f(\sqrt{n}s,\theta)|\right)\right\|^2, \tag{75}$$

$$\varpi_{3n}^2(\theta) \;\le\; \frac{n^{3\delta}}{\sqrt[4]{n}} \left\| (\dot\kappa_n \otimes \dot\kappa_n)^{-1} \left(\sup_{s \le \bar s + \varepsilon} \sup_{\theta \in N_n} |\ddot f(\sqrt{n}s, \theta)| \right) \right\| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left\| \dot\kappa_n^{-1} \dot f(x_t, \theta_0) \right\|$$

$$+ \frac{n^{4\delta}}{2} \left\| (\dot\kappa_n \otimes \dot\kappa_n)^{-1} \left(\sup_{s \le \bar s + \varepsilon} \sup_{\theta \in N_n} |\ddot f(\sqrt{n}s, \theta)| \right) \right\|^2, \tag{76}$$

$$\varpi_{4n}^2(\theta) \;\le\; n^{1/4+3\delta} \left\| (\dot\kappa_n \otimes \dot\kappa_n \otimes \dot\kappa_n)^{-1} \left(\sup_{s \le \bar s + \varepsilon} \sup_{\theta \in N_n} | \dddot f(\lambda s, \theta)| \right) \right\| \frac{1}{n} \sum_{t=1}^n |u_t|, \tag{77}$$

from which it follows that $\varpi_{1n}^2(\theta), \varpi_{3n}^2(\theta) = o_p(1)$ and $\varpi_{2n}^2(\theta), \varpi_{4n}^2(\theta) = o_{a.s.}(1)$, uniformly in $\theta \in N_n$, due to (22) and (23). We may now easily deduce AD7 from (74) – (77), and the proof is complete. ■

Proof of Corollary 5.5 Let σ_n^2 be given as (57). 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

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

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

so that $D_n(\hat\theta_n, \theta_0) = A_n - 2B_n$.

First, let the assumptions in Theorem 5.3 be satisfied. 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

$$A_n \le \left\| \nu_n'(\hat\theta_n - \theta_0) \right\|^2 \frac{1}{n} \sum_{t=1}^n \left\| \dot\kappa_n^{-1} \dot f(x_t, \theta_n) \right\|^2 = O_p(1),$$

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

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.

If the assumptions in Theorem 5.4 hold, we let $\nu_n = n^{1/4}\dot\kappa_n$, where $\dot\kappa_n = \dot\kappa_0(\sqrt{n}s_n)$, as in the proof of Theorem 5.4. Then we have

$$A_n \le \left\| \nu_n'(\hat\theta_n - \theta_0) \right\|^2 \frac{1}{\sqrt{n}} \sum_{t=1}^n \left\| \dot\kappa_n^{-1} \dot f(x_t, \theta_n) \right\|^2 = O_p(1),$$

$$\frac{|B_n|}{\sqrt[4]{n}} \le \left\| \nu_n'(\hat\theta_n - \theta_0) \right\| \left\| \frac{\dot\kappa_n^{-1}}{\sqrt{n}} \sum_{t=1}^n \dot f(x_t, \theta_n) u_t \right\| = o_p(1),$$

similarly as above, by Theorem 3.5 and Lemma A7. We may now easily deduce that $n^{-1/4} D_n(\hat\theta_n, \theta_0) = o_p(1)$, and therefore, $\hat\sigma_n^2 = \sigma_n^2 + o_p(n^{-3/4})$, which implies the stated result. ■

Proof of Theorem 5.6 For the redefined $Q_n(\theta)$ in the proof of Theorem 4.6, AD1 – AD3 follow immediately from Lemma A6(a) and Theorem 3.1 with $\nu_n = \sqrt{n}$ and

$$\ddot{Q}(\theta_0) = \int_0^1 \dot{f}_0(V(r),\theta_0)\dot{f}_0(V(r),\theta_0)'dr = \int_{-\infty}^{\infty} \dot{f}_0(s,\theta_0)\dot{f}_0(s,\theta_0)'L(1,s)\,ds.$$

Obviously, AD4 and AD5 hold. To prove AD6, we note that

$$\frac{1}{n}\sum_{t=1}^{n} \dot{f}(x_{nt},\theta)\dot{f}(x_{nt},\theta)' \;\to_{a.s}\; \int_0^1 \dot{f}_0(V(r),\theta)\dot{f}_0(V(r),\theta)'dr,$$

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \ddot{f}(x_{nt},\theta)\,u_t \;\;\to_p\;\; 0,$$

uniformly in $\theta \in \Theta$, due to Lemma A6(a), Theorem 3.1 and Lemma A7(a). It therefore remains to show that

$$\left\|\frac{1}{n}\sum_{t=1}^{n} \ddot{f}(x_{nt},\theta_n)\,(f(x_{nt},\theta_n) - f(x_{nt},\theta_0))\right\|$$
$$\leq \left\|\ddot{f}_0\right\|_K \int_0^1 |f_0(V(r),\theta) - f_0(V(r),\theta_0)|\,dr = o_{a.s.}(1),$$

uniformly in θ near θ_0, where $\ddot{f}_0(x,\theta) = \ddot{f}(\sqrt{n}_0,\theta)$ and $K = [s_{\min} - 1, s_{\max} + 1] \times \Theta$. However, this follows directly from Theorem 3.1, Lemma A3(b) and Lemma A8(a).■

Proof of Theorem 6.1 The stated result follows from (14) if we establish AD1–AD4 and AD7. Here we show AD7. The rest of the proof is entirely analogous to that of Theorem 5.1. We let $\nu_n = n^{1/4}$ and $\mu_n = n^{1/4-\delta}$ for $0 < \delta < 1/12$, and let N_n be given as in AD7. To deduce AD7, write $\ddot{Q}_n(\theta) - \ddot{Q}_n(\theta_0)$ as in (65) and note that

$$\varpi_{1n}^2(\theta) \;\leq\; \frac{n^{3\delta}}{\sqrt[4]{n}} \left\|\ddot{f}\right\| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\|\dot{f}(x_t,\theta_0)\right\| \to_p 0,$$

$$\varpi_{2n}^2(\theta) \;\leq\; \frac{n^{4\delta}}{\sqrt{n}} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\|\ddot{f}(x_t,\bar{\theta})\right\|^2 \to_p 0,$$

$$\varpi_{3n}^2(\theta) \;\leq\; \frac{n^{3\delta}}{\sqrt[4]{n}} \left\|\ddot{f}\right\| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\|\dot{f}(x_t,\theta_0)\right\| + o_p(1) \to_p 0,$$

$$\varpi_{4n}^2(\theta) \;\leq\; \frac{n^{3\delta}}{\sqrt[4]{n}} \left\|\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \dddot{f}\,(x_t,\bar{\theta})\,u_t\right\| \to_p 0,$$

for all $\theta \in N_n$. Again, $\bar{\theta}$ lies in the line segment connecting θ and θ_0. ■

Proof of Theorem 6.2 As in the proof of Theorem 6.1, we need only show AD7. The rest of the proof is essentially identical to that of Theorem 5.2. Write $\dot{\kappa}_n = \dot{\kappa}(\sqrt{n})$, and let $\nu_n = n^{1/2}\dot{\kappa}_n$ and $\mu_n = n^{1/2-\delta}\dot{\kappa}_n$ for $0 < \delta < 1/6$. Also, let N_n be defined as in AD7. If we let $K = [s_{\min} - 1, s_{\max} + 1] \times \Theta$, then we have from (66) – (69) that

$$\begin{aligned}
\varpi_{1n}^2(\theta) &\leq \frac{n^{3\delta}}{\sqrt{n}} \left\|(\dot{\kappa} \otimes \dot{\kappa})^{-1}\ddot{\kappa}\right\| \left\|\ddot{h}\right\|_K \frac{1}{n}\sum_{t=1}^{n} \left\|\dot{\kappa}_n^{-1}\dot{f}(x_t, \theta_0)\right\| + o_{a.s.}(1), \\
\varpi_{2n}^2(\theta) &\leq \frac{n^{4\delta}}{n} \left\|(\dot{\kappa} \otimes \dot{\kappa})^{-1}\ddot{\kappa}\right\|^2 \left\|\ddot{h}\right\|_K^2 + o_{a.s.}(1), \\
\varpi_{3n}^2(\theta) &\leq \frac{n^{3\delta}}{\sqrt{n}} \left\|(\dot{\kappa} \otimes \dot{\kappa})^{-1}\ddot{\kappa}\right\| \left\|\ddot{h}\right\|_K \frac{1}{n}\sum_{t=1}^{n} \left\|\dot{\kappa}_n^{-1}\dot{f}(x_t, \theta_0)\right\| + o_{a.s.}(1), \\
\varpi_{4n}^2(\theta) &\leq \frac{n^{3\delta}}{\sqrt{n}} \left\|(\dot{\kappa} \otimes \dot{\kappa} \otimes \dot{\kappa})^{-1}\dddot{\kappa}\right\| \left\|\dddot{h}\right\|_K \frac{1}{n}\sum_{t=1}^{n} |u_t| + o_{a.s.}(1),
\end{aligned}$$

all coverge a.s. to zero, uniformly on a neighborhood of θ_0. AD7 now follows directly from (65). ■

Proof of Lemma 6.3 It is obvious that AD1 holds. Also, we may easily deduce AD2, since

$$\left\|(\nu_n \otimes \nu_n)^{-1}\sum_{t=1}^{n} \ddot{f}(x_t, \theta_0)\, u_t\right\| \leq \sum_{i=1}^{m}\left(\sum_{t=1}^{n}\left\|(\nu_{in} \otimes \nu_{in})^{-1}\ddot{f}_i(x_t, \theta_{i0})\, u_t\right\|\right) \to_p 0,$$

as $n \to \infty$. To establish AD7, we introduce some additional notation. Let ε_i be given by Theorem 5.3 or 5.4 if f_i satisfies the assumptions there, and otherwise let ε_i be positive real numbers satisfying $0 < \varepsilon_i < 1/2$, for $i = 1, \ldots, m$. Subsequently, we define $\varepsilon = \min(\varepsilon_1, \ldots, \varepsilon_m)$ and δ to be a number such that $0 < \delta < \min(1/12, \varepsilon/3)$. Also, let

$$\mu_n = \mathrm{diag}\,(\,\mu_{1n}, \ldots, \mu_{mn}),$$

with $\mu_{in} = n^{-\delta}\nu_{in}$ for $i = 1, \ldots, m$, so that $\mu_n\nu_n^{-1} \to_{a.s.} 0$. Finally, we let $N_{in} = \{\theta_i : \|\mu_{in}'(\theta_i - \theta_{i0})\| \leq 1/m\}$, for $i = 1, \ldots, m$. Since $\|\mu_n'(\theta - \theta_0)\| \leq \sum_{i=1}^{m} \|\mu_{in}'(\theta_i - \theta_{i0})\|$, we have $\theta \in N_n$ whenever $\theta_i \in N_{in}$ for all $i = 1, \ldots, m$.

First note that

$$\left\|\,\mu_n^{-1}\dot{f}(x_t, \theta)\right\| \;\leq\; \sum_{i=1}^{m}\left\|\,\mu_{in}^{-1}\dot{f}_i(x_t, \theta_i)\right\|, \tag{78}$$

$$\left\|(\mu_n \otimes \mu_n)^{-1}\ddot{f}(x_t, \theta)\right\| \;\leq\; \sum_{i=1}^{m}\left\|(\mu_{in} \otimes \mu_{in})^{-1}\ddot{f}_i(x_t, \theta_i)\right\|, \tag{79}$$

Moreover, we have from the earlier results

$$\begin{aligned} \sum_{i=1}^{n}\left\|\nu_{in}^{-1}\dot{f}_i(x_t,\theta_{i0})\right\|^2 &= O_p(1), \\ n^{\varepsilon}\sup_{\theta_i\in N_{in}}\sum_{t=1}^{n}\left\|(\nu_{in}\otimes\nu_{in})^{-1}\ddot{f}_i(x_t,\theta_i)\right\|^2 &= o_p(1), \end{aligned}$$

and it follows that for all $i = 1, \ldots, m$

$$\begin{aligned} &\sum_{t=1}^{n}\left\|\mu_{in}^{-1}\dot{f}_i(x_t,\theta_{i0})\right\|\left\|(\mu_{jn}\otimes\mu_{jn})^{-1}\ddot{f}_j(x_t,\theta_j)\right\| \\ &\leq \left(\sum_{t=1}^{n}\left\|\mu_{in}^{-1}\dot{f}_i(x_t,\theta_{i0})\right\|^2\right)^{1/2}\left(\sum_{t=1}^{n}\left\|(\mu_{jn}\otimes\mu_{jn})^{-1}\ddot{f}_j(x_t,\theta_j)\right\|^2\right)^{1/2} \\ &= n^{3\delta}\left(\sum_{t=1}^{n}\left\|\nu_{in}^{-1}\dot{f}_i(x_t,\theta_{i0})\right\|^2\right)^{1/2}\left(\sum_{t=1}^{n}\left\|(\nu_{jn}\otimes\nu_{jn})^{-1}\ddot{f}_j(x_t,\theta_j)\right\|^2\right)^{1/2} \to_p 0, \end{aligned} \tag{80}$$

and

$$\sum_{t=1}^{n}\left\|(\mu_{in}\otimes\mu_{in})^{-1}\ddot{f}_i(x_t,\theta_i)\right\|^2 = n^{4\delta}\sum_{t=1}^{n}\left\|(\nu_{in}\otimes\nu_{in})^{-1}\ddot{f}_i(x_t,\theta_i)\right\|^2 \to_{a.s.} 0, \tag{81}$$

uniformly in $\theta_i \in N_{in}$.

If we define $\ddot{Q}_n(\theta) - \ddot{Q}_n(\theta_0)$ as in (65), then we have from (78)–(81) that

$$\begin{aligned} \varpi_{1n}^2(\theta) &\leq \sum_{t=1}^{n}\left\|\mu_n^{-1}\dot{f}(x_t,\theta_0)\right\|\left\|(\mu_n\otimes\mu_n)^{-1}\ddot{f}(x_t,\bar{\theta})\right\| \\ &\leq \sum_{i,j=1}^{m}\left(\sum_{t=1}^{n}\left\|\mu_{in}^{-1}\dot{f}_i(x_t,\theta_{i0})\right\|\left\|(\mu_{jn}\otimes\mu_{jn})^{-1}\ddot{f}_j(x_t,\bar{\theta}_j)\right\|\right) \to_p 0, \quad (82) \\ \varpi_{2n}^2(\theta) &\leq \sum_{t=1}^{n}\left\|(\mu_n\otimes\mu_n)^{-1}\ddot{f}(x_t,\bar{\theta})\right\|^2 \quad (83) \\ &\leq 2^{m-1}\sum_{i=1}^{m}\left(\sum_{t=1}^{n}\left\|(\mu_{in}\otimes\mu_{in})^{-1}\ddot{f}_i(x_t,\bar{\theta}_i)\right\|^2\right) \to_p 0, \quad (84) \\ \varpi_{3n}^2(\theta) &\leq \sum_{t=1}^{n}\left\|\mu_n^{-1}\dot{f}(x_t,\theta_0)\right\|\left\|(\mu_n\otimes\mu_n)^{-1}\ddot{f}(x_t,\theta)\right\| \\ &\quad +\frac{1}{2}\sum_{t=1}^{n}\left\|(\mu_n\otimes\mu_n)^{-1}\ddot{f}(x_t,\theta)\right\|\left\|(\mu_n\otimes\mu_n)^{-1}\ddot{f}(x_t,\bar{\theta})\right\| \\ &\leq \sum_{i,j=1}^{m}\left(\sum_{t=1}^{n}\left\|\mu_{in}^{-1}\dot{f}_i(x_t,\theta_{i0})\right\|\left\|(\mu_{jn}\otimes\mu_{jn})^{-1}\ddot{f}_j(x_t,\theta_j)\right\|\right) \end{aligned}$$

$$+2^{m-2}\sum_{i=1}^{m}\left(\sup_{\theta_i\in N_{in}}\sum_{t=1}^{n}\left\|(\mu_{in}\otimes\mu_{in})^{-1}\ddot{f}_i(x_t,\theta_i)\right\|^2\right)\to_p 0, \qquad (85)$$

for all $\theta\in N_n$. As in other proofs, $\bar{\theta}$ and $\bar{\theta}_i$ lie on the line segments connecting θ and θ_0, and θ_i and θ_{i0}, respectively. Moreover,

$$\varpi_{4n}^2(\theta)\leq\sum_{i=1}^{m}\left\|\sum_{t=1}^{n}(\mu_{in}\otimes\mu_{in})^{-1}\left(\ddot{f}_i(x_t,\theta_i)-\ddot{f}_i(x_t,\theta_{i0})\right)u_t\right\|\to_p 0, \qquad (86)$$

for all $\theta_i\in N_{in}$, $i=1,\ldots,m$, as shown earlier. We may now deduce from (82)–(86) that AD7 holds. ■

Proof of Theorem 6.4 For separability, it suffices to show that AD1–AD4 and AD7 hold with block diagonal $\ddot{Q}(\theta_0)$. Lemma 6.3, however, establishes AD1, AD2 and AD7, and we need only show that $\nu_n^{-1}\ddot{Q}_n^{\circ}(\theta_0)\nu_n^{-1\prime}$ converges in probability to a block-diagonal matrix $\ddot{Q}(\theta_0)$. Let $\dot{\kappa}_{1n}=\dot{\kappa}_1(\sqrt{n},\theta_{10})$, $\dot{f}_{i0}(\cdot)=\dot{f}_i(\cdot,\theta_{i0})$ for $i=1,2$ to simplify notation. Also, write $\dot{h}_1(\cdot,\theta_{10})=\dot{h}_{10}(\cdot)$ for short. Then we have

$$\nu_{1n}^{-1}\sum_{t=1}^{n}\dot{f}_{10}(x_t)\dot{f}_{20}(x_t)'\nu_{2n}^{-1\prime}=n^{-3/4}\sum_{t=1}^{n}\dot{\kappa}_{1n}^{-1}\dot{f}_{10}(x_t)\dot{f}_{20}(x_t)',$$

and

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\|\dot{\kappa}_{1n}^{-1}\dot{f}_{10}(x_t)\dot{f}_{20}(x_t)'\right\| &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\|\dot{h}_{10}\left(\frac{x_t}{\sqrt{n}}\right)\dot{f}_{20}(x_t)'\right\|+o_p(1)\\ &\leq \left\|\dot{h}_{10}\right\|_K\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\|\dot{f}_{20}(x_t)\right\|+o_p(1)=O_p(1),\end{aligned}$$

from which the block diagonality of $\ddot{Q}(\theta_0)$ follows immediately. ■

Proof of Theorem 6.5 As explained in the proof of Theorem 6.4, it suffices to show that $\ddot{Q}(\theta_0)$ is block-diagonal. Let $\dot{\kappa}_{1n}=\dot{\kappa}_1(\sqrt{n},\theta_{10})$ and $\dot{\kappa}_{2n}=\dot{\kappa}_2(\sqrt{n}s_n,\theta_{20})$ for notational brevity, and use the other notation defined in the proof of Theorem 6.4. We have

$$\nu_{1n}^{-1}\sum_{t=1}^{n}\dot{f}_{10}(x_t)\dot{f}_{20}(x_t)'\nu_{2n}^{-1\prime}=n^{-3/4}\sum_{t=1}^{n}\dot{\kappa}_{1n}^{-1}\dot{f}_{10}(x_t)\dot{f}_{20}(x_t)'\dot{\kappa}_{2n}^{-1\prime},$$

and that

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\|\dot{\kappa}_{1n}^{-1}\dot{f}_{10}(x_t)\dot{f}_{20}(x_t)'\dot{\kappa}_{2n}^{-1\prime}\right\| &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\|\dot{h}_{10}\left(\frac{x_t}{\sqrt{n}}\right)\dot{f}_{20}(x_t)'\dot{\kappa}_{2n}^{-1\prime}\right\|+o_p(1)\\ &\leq \left\|\dot{h}_{10}\right\|_K\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\|\dot{\kappa}_{2n}^{-1}\dot{f}_{20}(x_t)\right\|+o_p(1)=O_p(1),\end{aligned}$$

from which we may easily deduce the block diagonality of $\ddot{Q}(\theta_0)$. ■

Proof of Theorem 6.6 We will show that

$$\nu_n^{-1}\sum_{t=1}^n \dot f(x_t,\theta_0)u_t \;=\; \nu_n^{-1}\sum_{t=1}^n \dot f_*(x_t,\theta_0)\,u_t + o_p(1), \tag{87}$$

$$(\nu_n\otimes\nu_n)^{-1}\sum_{t=1}^n \ddot f(x_t,\theta_0)\,u_t \;=\; (\nu_n\otimes\nu_n)^{-1}\sum_{t=1}^n \ddot f_*(x_t,\theta_0)\,u_t + o_p(1), \tag{88}$$

$$\nu_n^{-1}\sum_{t=1}^n \dot f(x_t,\theta_0)\dot f(x_t,\theta_0)'\,\nu_n^{-1\prime} \;=\; \nu_n^{-1}\sum_{t=1}^n \dot f_*(x_t,\theta_0)\dot f_*(x_t,\theta_0)'\,\nu_n^{-1\prime} + o_p(1), \tag{89}$$

and subsequently establish AD7. Given AD7, it follows from (87)–(89) that the regression on f is asymptotically identical to that on f_*, for which the asymptotic theory is given by Theorem 6.4. The stated result, therefore, follows immediately.

Write $\dot\kappa_{1n} = \dot\kappa_1(\sqrt{n})$ as before, and let $c_n = \|n^{-1/4}\dot\kappa_{1n}^{-1}\|$. By (c), $c_n \to 0$ as $n\to\infty$. Also, we define $\dot f_i$ and $\ddot F_i$, for $i=1,2$, in the same way as $\dot f$ and $\ddot F$. Further, let

$$\dot f_2 = \begin{pmatrix} \dot f_{2\alpha} \\ \dot f_{2\beta} \end{pmatrix} \quad\text{and}\quad \ddot F_2 = \begin{pmatrix} \ddot F_{2\alpha\alpha} & \ddot F_{2\alpha\beta} \\ \ddot F_{2\beta\alpha} & \ddot F_{2\beta\beta} \end{pmatrix}.$$

Then it follows that

$$\dot f = \begin{pmatrix} \dot f_1 + \dot f_{2\alpha} \\ \dot f_{2\beta} \end{pmatrix} \quad\text{and}\quad \ddot F = \begin{pmatrix} \ddot F_1 + \ddot F_{2\alpha\alpha} & \ddot F_{2\alpha\beta} \\ \ddot F_{2\beta\alpha} & \ddot F_{2\beta\beta} \end{pmatrix}, \tag{90}$$

and

$$\dot f_* = \begin{pmatrix} \dot f_1 \\ \dot f_{2\beta} \end{pmatrix} \quad\text{and}\quad \ddot F_* = \begin{pmatrix} \ddot F_1 & 0 \\ 0 & \ddot F_{2\beta\beta} \end{pmatrix}. \tag{91}$$

Let $\ddot f_* = \operatorname{vec}\ddot F_*$ and $\ddot f_i = \operatorname{vec}\ddot F_i$.

To show (87)–(89), we note that

$$\dot f - \dot f_* = \begin{pmatrix} \dot f_{2\alpha} \\ 0 \end{pmatrix},\quad \ddot F - \ddot F_* = \begin{pmatrix} \ddot F_{2\alpha\alpha} & \ddot F_{2\alpha\beta} \\ \ddot F_{2\beta\alpha} & 0 \end{pmatrix},$$

and

$$\dot f\dot f' - \dot f_*\dot f_*' = \begin{pmatrix} \dot f_1\dot f_{2\alpha}' + \dot f_{2\alpha}\dot f_1' + \dot f_{2\alpha}\dot f_{2\alpha}' & \dot f_{2\alpha}\dot f_{2\beta} \\ \dot f_{2\beta}\dot f_{2\alpha} & 0 \end{pmatrix},$$

which follow directly from (90) – (91). Now we easily deduce (87), since

$$\left\|\nu_n^{-1}\sum_{t=1}^n (\dot f - \dot f_*)(x_t,\theta_0)\,u_t\right\| \le c_n \left\|\frac{1}{\sqrt[4]{n}}\sum_{t=1}^n \dot f_{2\alpha}(x_t,\theta_0)\,u_t\right\| \to_p 0.$$

It also follows that

$$\begin{aligned}\left\|(\nu_n\otimes\nu_n)^{-1}\sum_{t=1}^n (\ddot f - \ddot f_*)(x_t,\theta_0)\,u_t\right\| &\le \left\|(\nu_n\otimes\nu_n)^{-1}\sum_{t=1}^n \ddot f_2(x_t,\theta_0)\,u_t\right\| \\ &\le \frac{1+c_n+c_n^2}{\sqrt[4]{n}}\left\|\frac{1}{\sqrt[4]{n}}\sum_{t=1}^n \ddot f_2(x_t,\theta_0)\,u_t\right\| = o_p(1),\end{aligned}$$

which proves (88). Note that $\|(\nu_n \otimes \nu_n)^{-1}\| \leq n^{-1/2}(1+c_n+c_n^2)$. Moreover, we have

$$
\begin{aligned}
\left\|\nu_{1n}^{-1} \sum_{t=1}^{n} \dot{f}_1(x_t,\theta_0)\dot{f}_{2\alpha}(x_t,\theta_0)'\,\nu_{1n}^{-1\prime}\right\| &\leq \frac{c_n}{\sqrt[4]{n}} \left\|\dot{h}_1\right\|_K \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\|\dot{f}_{2\alpha}(x_t,\theta_0)\right\| + o_p(1),\\
\left\|\nu_{1n}^{-1} \sum_{t=1}^{n} \dot{f}_{2\alpha}(x_t,\theta_0)\dot{f}_{2\alpha}(x_t,\theta_0)'\,\nu_{1n}^{-1\prime}\right\| &\leq c_n^2 \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\|\dot{f}_{2\alpha}(x_t,\theta_0)\dot{f}_{2\alpha}(x_t,\theta_0)'\right\|,\\
\left\|\nu_{1n}^{-1} \sum_{t=1}^{n} \dot{f}_{2\alpha}(x_t,\theta_0)\dot{f}_{2\beta}(x_t,\theta_0)'\,\nu_{2n}^{-1\prime}\right\| &\leq c_n \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\|\dot{f}_{2\alpha}(x_t,\theta_0)\dot{f}_{2\beta}(x_t,\theta_0)'\right\|,
\end{aligned}
$$

all of which are $o_p(1)$, and from which (89) follows easily.

Now we show that AD7 holds for some μ_n. Let δ be any number such that $0 < \delta < 1/12$, and let $\mu_n = n^{-\delta}\nu_n$ as before. Also, define N_n as given in AD7. Clearly, $\mu_n \nu_n^{-1} \to 0$, as required. Furthermore, we have

$$
\left\|\mu_n^{-1}\dot{f}(x,\theta)\right\| \leq n^{-1/2+\delta}\left\|\dot{\kappa}_{1n}^{-1}\dot{f}_1(x,\alpha)\right\| + n^{-1/4+\delta}\left\|\dot{f}_2(x,\alpha,\beta)\right\|, \quad (92)
$$

$$
\begin{aligned}
\left\|(\mu_n \otimes \mu_n)^{-1}\ddot{f}(x,\theta)\right\| &\leq n^{-1+2\delta}\left\|(\dot{\kappa}_1 \otimes \dot{\kappa}_1)^{-1}\ddot{\kappa}_1\right\|\left\|\ddot{\kappa}_{1n}^{-1}\ddot{f}_1\right\|\\
&\quad + n^{-1/2+2\delta}(1+c_n+c_n^2)\left\|\ddot{f}_2(x,\alpha,\beta)\right\|, \quad (93)
\end{aligned}
$$

for all $\theta \in N_n$. If we write $\ddot{Q}_n(\theta) - \ddot{Q}_n(\theta_0)$ as in (65), then it follows immediately from (92) and (93) that $\varpi_{1n}^2(\theta), \varpi_{2n}^2(\theta), \varpi_{3n}^2(\theta) \to_p 0$ as $n \to \infty$, uniformly in $\theta \in N_n$.

Now

$$
\varpi_{4n}^2(\theta) \leq \left\|(\mu_n \otimes \mu_n \otimes \mu_n)^{-1}\sum_{t=1}^{n} \dddot{f}\,(x_t,\bar{\theta})\,u_t\right\|,
$$

with $\bar{\theta}$ between θ and θ_0. However,

$$
\begin{aligned}
\left\|(\mu_n \otimes \mu_n \otimes \mu_n)^{-1}\sum_{t=1}^{n} \dddot{f}\,(x_t,\theta)\,u_t\right\| &\leq \frac{n^{3\delta}}{\sqrt{n}}\left\|(\dot{\kappa}_1 \otimes \dot{\kappa}_1 \otimes \dot{\kappa}_1)^{-1}\,\dddot{\kappa}_1\right\|\left\|\frac{1}{n}\sum_{t=1}^{n}\dddot{\kappa}_{1n}^{-1}\dddot{f}_1\,(x_t,\alpha)\,u_t\right\|\\
&\quad + \frac{n^{3\delta}(1+c_n+c_n^2+c_n^3)}{\sqrt[4]{n}}\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\dddot{f}_2\,(x_t,\alpha,\beta)\,u_t\right\|.
\end{aligned}
$$

Notice that

$$
\left\|(\mu_n \otimes \mu_n \otimes \mu_n)^{-1}\right\| \leq n^{-3/4+3\delta}(1+c_n+c_n^2+c_n^3).
$$

Since we have by Lemma A7

$$
\left\|\frac{1}{n}\sum_{t=1}^{n}\dddot{\kappa}_{1n}^{-1}\dddot{f}_1\,(x_t,\alpha)\,u_t\right\|, \left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\dddot{f}_2\,(x_t,\alpha,\beta)\,u_t\right\| \to_p 0,
$$

uniformly in α and β, it follows that $\varpi_{4n}^2(\theta) \to_p 0$ uniformly on N_n. The condition AD7 therefore holds, as was to be shown. ■

References

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

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.* Wiley, New York.

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

Chung, K.L. and R.J. Williams (1990). *Introduction to Stochastic Integration*, 2nd ed. Birkhäuser, Boston.

Granger, C. W. J. (1995). "Nonlinear realtionships between extended-memory variables". *Econometrica*, 63, 265-280.

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.

Karatzas, I. and S.E. Shreve (1988). *Brownian Motion and Stochastic Calculus.* Springer-Verlag, New York.

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.

Park, J.Y. and P.C.B. Phillips (1997). Asymptotics for nonlinear transformations of integrated time series. Yale University, mimeographed.

Phillips, P. C. B. (1986). "Understanding spurious regressions in econometrics," *Journal of Econometrics* 33, 311–340.

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

Phillips, P. C. B. (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 W. Ploberger (1996). An asymptotic theory of Bayesian inference for time series. *Econometrica* **64**, 381–412.

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.* Springer–Verlag, New York.

Revuz, D. and M. Yor (1994). *Continuous Martingale and Brownian Motion*, 2nd ed. Springer–Verlag, New York.

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 R.F. Engle and D.L. McFadden (eds.) *Handbook of Econometrics,* Vol. IV, pp. 2639–2738. Elsevier, Amsterdam.

Wu, C.F. (1981). Asymptotic theory of nonlinear least squares estimation. *Annals of Statistics* **9**, 501–513.