Nonlinear Econometric Models with Cointegrated and Deterministically Trending Regressors[1]

Yoosoon Chang
Department of Economics
Rice University

Joon Y. Park
School of Economics
Seoul National University

and

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

Abstract

This paper develops an asymptotic theory for a general class of nonlinear nonstationary regressions. The model considered accommodates a linear time trend and stationary regressors, as well as multiple I(1) regressors. We establish consistency and derive the limit distribution of the nonlinear least squares estimator. The estimator is consistent under fairly general conditions but the convergence rate and the limiting distribution are critically dependent upon the type of the regression function. For integrable regression functions, the parameter estimates converge at a reduced $n^{1/4}$ rate and have mixed normal limit distributions. On the other hand, if the regression functions are homogeneous at infinity, the convergence rates are determined by the degree of the asymptotic homogeneity and the limit distributions are non-Gaussian. It is shown that nonlinear least squares generally yields inefficient estimators and invalid tests, just as in linear nonstationary regressions. The paper proposes a methodology to overcome such difficulties. The approach is simple to implement, produces efficient estimates and leads to tests that are asymptotically chi-square.

Keywords: Nonlinear regressions, integrated time series, nonlinear least squares, Brownian motion, Brownian local time

[1] Chang and Park gratefully acknowledge the financial support from Rice University through CSIV fund. Park thanks the Department of Economics at Rice University, where he is an Adjunct Professor, for its continuing hospitality. Phillips thanks the NSF for support under Grant No. SBR 97-30295. The paper was completed while Chang and Park visited the Cowles Foundation, Yale University in the fall of 1999. The paper was typed by the authors in SW2.5.

Original Draft: January, 1999
Completed Version: October, 1999

1. Introduction

Most of the work done on nonlinear econometric models since the early 1970's has involved nontrending variables and a substantial body of theory has been developed. Much of the asymptotic theory for such models is reliant on strong laws and central limit theory for weakly dependent time series. GMM estimation theory, in particular, has been especially reliant on such conditions for its development. Two exceptions to the general thrust of this research are Wooldridge (1994) and Andrews and McDermott (1995). Wooldridge developed asymptotics under high level conditions that encompass some interesting cases of trending variables, although much work is needed in verifying the conditions and it is only in doing so that the effects of the trends are understood. Andrews and McDermott sought to extend the theory of extremum estimation to situations where deterministic trends (but not stochastic trends) appear in the nonlinear model by using triangular array asymptotics. Both papers gave qualitatively similar results, indicating that the asymptotic distributions of extremum estimators are generally the same (i.e. normal and chi-squared) with deterministically trending variables as for nontrending variables (Andrews and McDermott, 1995, p.343).

One of the main difficulties of dealing with general nonlinear functions of trending variables is that the asymptotic behavior of sample partial sums of the functions is no longer immediately apparent. For instance, if x_t is a strictly stationary and ergodic time series, then the function $1/(1+\theta x_t^2)$ is measurable and, therefore, stationary and ergodic. Since the function is also an integrable function, it follows by the ergodic theorem that

$$\frac{1}{n}\sum_{t=1} \frac{1}{1+\theta x_t^2} \to_{a.s.} E\left(\frac{1}{1+\theta x_t^2}\right), \tag{1}$$

and the convergence is clearly uniform over $\theta \in \Theta$, for any compact set Θ in $\mathbb{R}^+$. On the other hand, when $x_t = t$ is a linear trend we have the following quite different behavior

$$\sum_{t=1} \frac{1}{1+\theta t^2} < \frac{1}{\theta}\sum_{t=1} \frac{1}{t^2} < \frac{1}{\theta}\sum_{t=1} \frac{1}{t^2} < \infty, \tag{2}$$

for all $\theta > 0$, and uniformly so for θ in a compact set $\Theta \subset \mathbb{R}^+$. The situation is much more complex when x_t is a stochastic trend and very different results apply. The analysis of sample mean asymptotics in this case was done recently in Park and Phillips (1998) using some new techniques of spatial analysis for nonstationary processes.

When the nonlinear function of a trend is homogeneous, like a polynomial, an automatic restandardization is possible. Thus, if $f(\lambda t) = \lambda^k f(t)$, then the sample partial sums behave like Riemann sums and can be approximated as follows

$$\frac{1}{n^{1+k}}\sum_{t=1} f(t) = \frac{1}{n}\sum_{t=1} f\left(\frac{t}{n}\right) \approx \int_0^1 f(r)dr.$$

This suggests that one approach to developing an asymptotic theory for trending series is to 'force' a nonlinear function into a framework where the sample moments behave like Riemann sums. The triangular array approach of Andrews and McDermott (1995) can be viewed in this light. Some of the disadvantages of this approach are mentioned in their paper. In the present case, it suffices to point out that the approach implies that sample means like (2) can, for some given n_0, be written in the approximate form (for large n and large fixed n_0)

$$\sum_{t=1} \frac{1}{1+\theta n_0^2\left(\frac{t}{n}\right)^2}. \tag{3}$$

Then,

$$\frac{1}{n}\sum_{t=1} \frac{1}{1+\theta n_0^2\left(\frac{t}{n}\right)^2} \to \int_0^1 \frac{1}{1+\theta n_0^2 r^2}dr,$$

as $n \to \infty$, implying that (3) is of order $O(n)$, whereas the original sample moment (2) is $O(1)$. Thus, one possible effect of forcing sample partial sums into a Riemann sum framework is to materially change their order of magnitude. Clearly, to the extent that such changes influence the excitation property in a regression, they can affect properties, like the consistency of an estimator. This is indeed what happens in a nonlinear regression model of the form

$$y_t = \frac{1}{1+\theta t^2} + u_t, \tag{4}$$

where the parameter θ is to be estimated and the errors u_t are iid $(0,1)$. For, in (4), when t is large the model becomes asymptotically equivalent to $y_t = \frac{1}{\theta t^2} + u_t$, there is little information in the data y_t about the parameter θ, and the persistent excitation condition fails because $\sum_{t=1}^{\infty} t^{-4} < \infty$, just like (2). On the other hand, when we replace the true model with the approximate formulation

$$y_t = \frac{1}{1+\theta n_0^2\left(\frac{t}{n}\right)^2} + u_t, \tag{5}$$

it is apparent that, for large t, the model retains its form and there continues to be information in the mean of y_t about θ. Indeed, the marginal information is contained in the derivative

$$-\frac{n_0^2\left(\frac{t}{n}\right)^2}{\left[1+\theta n_0^2\left(\frac{t}{n}\right)^2\right]^2}$$

for which we have

$$\sum_{t=1}^{n} \frac{n_0^2 \left(\frac{t}{n}\right)^2}{\left[1+\mu n_0^2 \left(\frac{t}{n}\right)^2\right]^2} \sim n \int_0^1 \frac{n_0^2 r^2 dr}{\left[1+\mu n_0^2 r^2\right]^2} = O(n);$$

and persistent excitation follows. In short, μ is consistently estimable in (5), but not consistenty estimable in (4). Thus, the net effect of forcing sample partial sum functions into a Riemann summable form is to lose some of the essential features of the nonlinearity. In the case of (4), what is lost is that the trend is evaporating (i.e., effectively temporary), so that the information that it carries about μ also evaporates as t becomes large. What (5) does, is reformulate the function so that every observation continues to count, just as it does in the stationary and ergodic case (1). Thus, when we force sample mean functions into a Riemann Stieltjes form, we effectively force the asymptotics into the framework that applies in the case of weak dependence and little that is new is learnt about the effect of the nonlinearity and, in some cases, like (2), some important characteristics are lost.

This paper seeks to develop an asymptotic theory for a general type of nonlinear nonstationary regression without using the device of a triangular array. It may be considered as continuation of the research program of Phillips and Durlauf (1986), Park and Phillips (1988, 1989) and Phillips and Hansen (1990) to the nonlinear case. In particular, the paper builds on the theory for nonlinear regression with integrated time series constructed recently in Park and Phillips (1999). The theory there was developed for scalar I(1) regressors and no deterministic functions of time were included. That theory is extended here to allow for multiple integrated regressors and for the presence of both deterministic trends and stationary regressors. In doing so, it leads to a framework of efficient nonlinear regression for time series that have deterministic and stochastic nonstationarity as well as stationary components. As such, the theory should be applicable in many practical cases of nonlinear econometric models with deterministic and stochastic trends.

In nonlinear regressions with integrated time series, the asymptotic theory for multiple regression turns out to be different from that of simple regression in a non-trivial way. This is so, since integrated time series behave like Brownian motion asymptotically, and, as is known in the theory of stochastic processes, the behavior of nonlinear functionals of a vector Brownian motion can be drastically different from that of a one dimensional Brownian motion. It transpires that these differences show up in important ways in regression asymptotics with nonlinear functions of I(1) processes. These differences are explored in the present paper.

The plan of the paper is as follows. Section 2 lays out the model and assumptions. Section 3 gives some preliminary asymptotic theory for sample moments that provides the foundation of our analysis. The nonlinear regression theory is developed in Section 4. Section 5 considers issues of efficient estimation, and Section 6 reports some simulations. Section 7 concludes and proofs are collected together in Section 8.

A word about our notation. For a vector $x = (x_i)$, the notation $\|x\|$ stands for the standard Euclidean norm, i.e., $\|x\|^2 = \sum_i x_i^2$. When applied to a matrix, $\|A\|$

signifies the operator norm defined by $\|A\| = \sup_x \|Ax\|/\|x\|$. The same notation is also used to denote the supremum norm for a function, and therefore, we write $\|f\| = \sup_x |f(x)|$ if $f : \mathbb{R} \to \mathbb{R}$. Tensor division by vectors is denoted by $\otimes/$; so that $a \otimes/ b = (a_1/b_1, \ldots, a_1/b_q, a_2/b_1, \ldots, a_p/b_q)$ for p-vector a and q-vector b.

2. The Model and Assumptions

We consider the nonlinear regression model for y_t given by

$$\begin{aligned} y_t &= f(z_t, \theta_0) + u_t \\ &= \tau(t, \pi_0) + p(w_t, \alpha_0) + q(x_t, \beta_0) + u_t, \end{aligned} \tag{6}$$

where w_t and x_t are stationary and integrated regressors, respectively, and the functions τ, p and q are assumed to be all known. The regression (6) generalizes the model studied previously by Park and Phillips (1998,1999) in two important directions. While those papers concentrate on bivariate nonlinear regression, we allow for multiple integrated regressors, and for the presence of a deterministic trend and stationary regressors. When specialized to linear functions, (6) reduces to a conventional cointegrating regression, possibly with a time trend and other stationary regressors. The statistical theory for such regressions was developed in Phillips and Durlauf (1986) and Park and Phillips (1988,1889), and is now heavily utilized in both theoretical and practical work.

The theory of regressions on time trends and stationary time series is well established and can be found, e.g., in Wooldridge (1994). Here, we include such regressors in our expanded model to study the effect of their inclusion in nonlinear regressions which also involve integrated regressors. To highlight the additional effects, we take the case where both the deterministic and stationary components of the regression function are linear, and we specify τ and p simply as

$$\tau(t, \pi) = \pi' d_t \quad \text{and} \quad p(w_t, \alpha) = \alpha' w_t, \tag{7}$$

where d_t is a deterministic sequence such as constant and a linear time trend. It should be emphasized here, however, that our subsequent theory may well apply to models with more general specifications of τ and p. Of course, much more general nonadditive functions of t and w_t together would be accommodated if we were to use the simplifying approach outlined in the introduction (as in (5)) and used in Andrews and McDermott (1995).

For our subsequent theory to be applicable, it is essential to have additivity between the parts of the regression function driven by integrated processes and by other deterministic and stationary processes. In particular, we cannot accommodate nonadditive functions of t, w_t and x_t in our present framework. In our model (6), we further specify q as

$$q(x, \beta) = \sum_{i \in I} q_i(x_i, \beta_i) + \sum_{i \in H} q_i(x_i, \beta_i) \tag{8}$$

where $x = (x_1, \ldots, x_m)'$ and $\beta = (\beta_1', \ldots, \beta_m')'$ with $\beta_0 = (\beta_1^{0\prime}, \ldots, \beta_m^{0\prime})'$. The index sets I and H are mutually exclusive and exhaustive, i.e., it is assumed that either $i \in I$ or $i \in H$ for each and every $i = 1, \ldots, m$. These sets refer, respectively, to groups of integrable and groups of asymptotically homogeneous functions. That is, $q_i(\cdot, \beta_i), i \in I$, is an integrable function, while $q_i(\cdot, \beta_i), i \in H$, is a function which behaves asymptotically as a homogeneous function.

The concept of an asymptotically homogeneous function was first introduced by Park and Phillips (1999) as part of their asymptotic analysis of nonlinear transformations of integrated processes. If $T : \mathbb{R} \to \mathbb{R}^k$ is asymptotically homogeneous, it has the representation

$$T(\lambda s) \simeq \kappa(\lambda) H(s),$$

for large λ. The transformation T is thus expected to behave asymptotically like a homogeneous function, as $\lambda \to \infty$. We call κ and H, the asymptotic order and the limit homogeneous function of T, respectively. The class of asymptotically homogeneous functions includes, among others, polynomials, logarithmic functions and all distribution function-like functions. In particular, the asymptotically homogeneous functions $T(s) = |s|^k, \log|s|, 1/(1 + e^{-s})$ have asymptotic orders $\kappa(\lambda) = \lambda^k, \log \lambda, 1$ and limit homogeneous functions $H(s) = |s|^k, 1, 1\{s \geq 0\}$, respectively. The reader is referred to Park and Phillips (1999) for further details and discussion.

In our specification (8), the nonstationary stochastic part of the model itself is also assumed to be additively separable. The assumption of additive separability here is also important for our subsequent results to hold, and asymptotic results for models with nonadditive functions of two integrated regressors can be expected to be quite different. This is because, moments of functions of integrated processes behave asymptotically as functionals of Brownian motions and, as is well known, the theory of functionals of a vector Brownian motion is generally very different from that of a scalar Brownian motion. Note in (8) that x_i is scalar, but β_i may be a vector. Therefore, each additive term may only include a single integrated regressor, but can have multiple parameters.

For regression with integrated regressors, the I- and H-regularities introduced in Park and Phillips (1998) define appropriate regularity conditions on a regression function, which is a function of both a regressor and a parameter. As a function of the regressor, these conditions require that the regression function be integrable or asymptotically homogeneous, while, as a function of the parameter, some continuity and identifying conditions are needed. Roughly speaking, all regression functions that are integrable as functions of the regressor are I-regular, if they are bounded and piecewise smooth. Indicators over compact intervals such as $q_i(x_i, \beta_i) = \beta_i 1\{0 \leq x_i \leq 1\}$ and other smooth functions like $q_i(x_i, \beta_i) = e^{-\beta_i x_i^2}$ are some obvious examples. In general, we may allow for less smooth functions, provided we require the existence of higher moments of the underlying process. On the other hand, regression functions that are asymptotically homogeneous as functions of the regressor are H-regular under very mild conditions. Examples include regression functions such as $q_i(x_i, \beta_i) = x_i^{\beta_i}, |x_i|^{\beta_i}, \beta_i \log|x_i|, (|x_i|^{\beta_i} - 1)/\beta_i, \beta_i 1\{x_i \geq 0\}$, among many others. For more examples, see Park and Phillips (1998).

To introduce our assumptions on q_i, we let $\dot q_i$, $\ddot q_i$ and $\dddot q_i$ denote, respectively, the first, second and third derivatives of q_i with respect to β_i. We assume that they are all vectorized and arranged by lexicographic ordering of their indices. For H-regular q_i, we denote by $\dot\kappa_i$, $\ddot\kappa_i$ and $\dddot\kappa_i$ the asymptotic orders of $\dot q_i$, $\ddot q_i$ and $\dddot q_i$, respectively, and $\dot h_i$ signifies the limit homogeneous function of $\dot q_i$. In general, the asymptotic orders of H-regular $\dot q_i$, $\ddot q_i$ and $\dddot q_i$ depend on β_i. If they are the same for all values of β_i, then we say that they are H_0-regular.

Assumption I The functions $q_i, i \in I$, satisfy the following conditions.

(a) $\dot q_i, \ddot q_i, \dddot q_i$ are I-regular.

(b) $\int_{-\infty}^{\infty} \dot q_i(s, \beta_i^0)\dot q_i(s, \beta_i^0)' ds > 0$.

Assumption H The functions $q_i, i \in H$, satisfy either (a_1) or (a_2), and (b) below.

(a_1) $\dot q_i, \ddot q_i, \dddot q_i$ are H_0-regular with $\|\ddot\kappa_i / (\dot\kappa_i \otimes \dot\kappa_i)\|, \|\dddot\kappa_i / (\dot\kappa_i \otimes \dot\kappa_i \otimes \dot\kappa_i)\| < \infty$.

(a_2) if we let $N_i(\delta) = \{\beta_i : \|\beta_i - \beta_i^0\| < \delta\}$ for $\delta > 0$, then there exists $\varepsilon > 0$ such that $\lambda^{-1+\varepsilon}\|\dot\kappa_i(\lambda, \beta_i^0)^{-1}\| \to 0$ and

$$\lambda^{\varepsilon}\left\| (\dot\kappa_i \otimes \dot\kappa_i)(\lambda, \beta_i^0)^{-1} \left(\sup_{|s| \le \bar{s}} \sup_{\beta_i \in N_i(\delta)} |\ddot q_i(\lambda s, \beta_i)| \right) \right\| \to 0,$$

for any $\bar{s} > 0$.

(b) $\int_{|s|<\delta} \dot h_i(s, \beta_i^0)\dot h_i(s, \beta_i^0)' ds > 0$ for all $\delta > 0$.

In both Assumption I and H, condition (a) gives the regularity conditions on the regression function, while condition (b) is for identification purposes. Throughout the paper, we assume that $q_i, i \in I$ and $q_i, i \in H$ satisfy the conditions in Assumptions I and H above. The conditions are mild enough to accommodate virtually all functions used in practical nonlinear analyses, except for exponential functions. We could allow for exponential functions in our model and this would involve no new technical difficulties, as shown in Park and Phillips (1998). However, it is not done in the present paper to make our model, assumptions and theoretical results simple and more presentable. The gains from including exponential functions seem slight, as exponential functions of an integrated process have exaggerated explosive behavior and they seem to be of little empirical relevance.

We now introduce precise assumptions on the data generating processes. As mentioned earlier, x_t is assumed to be an integrated process, so we let $v_t = \Delta x_t$ be a stationary process with certain properties. We specify v_t and w_t as general linear processes given by

$$v_t = \Phi(L)\varepsilon_t = \sum_{k=0}^{\infty} \Phi_k \varepsilon_{t-k} \quad \text{and} \quad w_t = \Psi(L)\eta_t = \sum_{k=0}^{\infty} \Psi_k \eta_{t-k}; \tag{9}$$

with $\Phi_0 = I$ and $\Psi_0 = I$. Define $\xi_t = (\xi_{i,t}) = (u_t, \varepsilon_{t+1}', \eta_{t+1}')'$ and the filtration $\mathcal{F}_t = \sigma\left((\xi_s)_{-\infty}^{t}\right)$, i.e., the σ-field generated by $(\xi_s)_{s\le t}$. The following conditions are made on the innovation process ξ_t:

2.1 Assumption

(a) $(\xi_t, \mathcal{F}_t)$ is a stationary and ergodic martingale difference sequence,
(b) $E(\xi_t\xi_t'|\mathcal{F}_{t-1}) = \Sigma$,
(c) $\sup_{t\ge 1} E(\|\xi_t\|^r|\mathcal{F}_{t-1}) < \infty$ for some $r > 4$.
(d) $E(\xi_{i,t}^2\xi_{j,t-l}) = 0$ for all i, j and for all $l \ge 1$.

Condition (a) implies that the regressors x_t and w_t are predetermined, i.e., $E(x_t|\mathcal{F}_{t-1}) = E(w_t|\mathcal{F}_{t-1}) = 0$. We therefore have $E(y_t|\mathcal{F}_{t-1}) = f(z_t, \theta_0)$, as in usual nonlinear regression theory. The moment conditions in (b) and (c) are fairly standard. The third moment condition (d) holds fairly generally, for example, for all ARCH processes with $E(\xi_{i,t}^3|\mathcal{F}_{t-1}) = 0$ (see Guo and Phillips, 1998). We partition Σ as

$$\Sigma = \begin{pmatrix} \sigma_u^2 & \sigma_{u\varepsilon} & \sigma_{u\eta} \\ \sigma_{\varepsilon u} & \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon\eta} \\ \sigma_{\eta u} & \Sigma_{\eta\varepsilon} & \Sigma_{\eta\eta} \end{pmatrix};$$

conformably with ξ_t.

The martingale difference assumption on the error process can be relaxed under certain circumstances. For instance, the assumption would be unnecessary for the consistency of the LS estimator if there were only linear functions of integrated regressors in which case our model (6) would reduce to a simple cointegrating regression. This is, of course, well known. Also, consistency of the NLS estimator under more general errors can be established if the regression includes only certain classes of asymptotically homogeneous regression functions such as polynomials. On the other hand, a full generalization of our theory allowing for correlated errors would involve a substantial additional level of complexity and is not attempted here. However, it does not seem overly restrictive at this point to assume the absence of serial correlation in the errors, especially given our flexible nonlinear specification of the regression function and the presence of stationary regressors in the model.

2.2 Assumption

(a) $\Phi(1) \neq 0$, $\sum_{k=0}^{\infty} k\|\Phi_k\| < \infty$, and
(b) $\sum_{k=0}^{\infty} k^{1/2}\|\Psi_k\| < \infty$.

Summability conditions like those for Φ_k and Ψ_k in Assumption 2.2 are standard, and are routinely imposed in stationary time series analysis. Of course, 1/2–summability in (b) is weaker than 1–summability in (a). The condition $\Phi(1) \neq 0$ ensures that x_t is integrated of order one. For the innovation process ε_t of x_t, we impose some stronger conditions.

2.3 Assumption ε_t is iid with $E\|\varepsilon_t\|^r < \infty$ for some $r > 8$, and its distribution is absolutely continuous with respect to Lebesque measure and has characteristic function φ for which $\varphi(\lambda) = o(\|\lambda\|^{-\delta})$ as $\lambda \to \infty$ for some $\delta > 0$.

Assumption 2.3 is strong, but is still satisfied by a wide class of data generating processes, including all invertible Gaussian ARMA models.

For u_t and v_t, we define stochastic processes

$$U_n(r) = \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} u_t \quad \text{and} \quad V_n(r) = \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} v_t \tag{10}$$

on $[0,1]$, where $[s]$ denotes the largest integer not exceeding s. The process (U_n, V_n) takes values in $D[0,1]^{1+m}$, where $D[0,1]$ is the space of cadlag functions on $[0,1]$. Under Assumptions 2.1 and 2.2(a), an invariance principle holds for (U_n, V_n). That is, we have as $n \to \infty$

$$(U_n, V_n) \to_d (U, V) \tag{11}$$

where (U, V) is $(1+m)$-dimensional vector Brownian motion, as shown in Phillips and Solo (1992). For more general invariance principles relevant to the analysis of models with integrated processes, the reader is referred to Phillips and Durlauf (1986) and Hansen (1992), and the references cited there.

The covariance matrix of the limit Brownian motion (U, V) is written as

$$\Omega = \begin{pmatrix} \omega_u^2 & \omega_{uv} \\ \omega_{vu} & \Omega_{vv} \end{pmatrix}.$$

Note that $\omega_u^2 = \sigma_u^2$, since u_t is a martingale difference sequence. We also have

$$\omega_{vu} = \Phi(1)\sigma_{\varepsilon u} \quad \text{and} \quad \Omega_{vv} = \Phi(1)\Sigma_{\varepsilon\varepsilon}\Phi(1)' \tag{12}$$

as is easily checked.

2.4 Assumption There exists a nonsingular sequence of normalizing matrices κ_{nd} such that if $d_n(r) = \kappa_{nd}^{-1} d_{[nr]}$ on $[0,1]$, then

(a) $\sup_{n \geq 1} \sup_{0 \leq r \leq 1} \|d_n(r)\| < \infty$, and

(b) $d_n \to_{L^2} d$ as $n \to \infty$ for some $d \in L^2[0,1]$ such that $\int_0^1 d(r)d(r)' dr > 0$.

The conditions in Assumption 2.4 are general enough to allow for deterministic regressors such as constant and time polynomials, possibly with breaks, which are commonly used in time series analyses (see Park, 1992, for the asymptotics of integrated processes with such time trends). Condition (b) is quite weak, as in most cases of practical interest we will have uniform convergence $\|d_n - d\| \to 0$.

3. Preliminary Results

This section outlines some background theory and gives some preliminary results that are crucial for the asymptotic analysis of our model. The first asymptotic theory for nonlinear regressions with integrated processes was developed in Park and Phillips (1998,1999). Their results provide the essential tools for the analysis of the asymptotic properties of the NLS estimator, but are applicable only to regressions with a single integrated regressor. Here, we extend those results to allow for multiple integrated processes as well as a time trend and stationary regressors.

The theory relies heavily on the local time of Brownian motion, which is described briefly for convenience here, while referring the reader to a standard source such as Revuz and Yor (1994) for a detailed discussion. The local time of a Brownian motion B is a two parameter process, written as $L_B(t,s)$, with t and s respectively being the time and spatial parameters, satisfying the important (so-called occupation time) formula

$$\int_0^t T(B(r))\,d[B]_r = \int_{-\infty}^{\infty} T(s)L_B(t,s)\,ds, \tag{13}$$

for locally integrable $T : \mathbb{R} \to \mathbb{R}^k$, where $[B]_r$ is the quadratic variation process of B. If we apply (13) to the function $T(x) = 1\{a \leq x \leq b\}$ for $a, b \in \mathbb{R}$, then

$$\int_0^t 1\{a \leq B(r) \leq b\}\,d[B]_r = \int_a^b L_B(t,s)\,ds,$$

and, correspondingly, when the local time $L_B(t,s)$ is treated as a function of its spatial parameter s, it can be viewed as an occupation time (or soujourn) density. The time that B stays in the interval $[a,b]$ is measured by $d[B]_r$, which can be thought of as a natural time scale for B. Also, due to the continuity of $L_B(t,\cdot)$, we have

$$L_B(t,s) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^t 1\{|B(r) - s| < \varepsilon\}\,d[B]_r.$$

Therefore, $L_B(t,s)$ measures the time (in units of quadratic variation) that B spends in the neighborhood of s, up to time t.

Write $V(r) = (V_1(r), \ldots, V_m(r))'$, and denote by L_{V_i} the local time of V_i, for $i = 1, \ldots, m$. Define

$$L_i(t,s) = (1/\omega_i^2)L_{V_i}(t,s) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^t 1\{|V_i(r) - s| < \varepsilon\}\,dr,$$

where ω_i^2 is the variance of V_i, for $i = 1, \ldots, m$. Clearly, L_i is just a scaled local time of V_i that measures time in chronological units. Our asymptotic results will be presented using L_i, instead of L_{V_i}. Using L_i, the occupation time formula (13) is rewritten as

$$\int_0^t T(V_i(r))\,dr = \int_{-\infty}^{\infty} T(s)L_i(t,s)\,ds, \tag{14}$$

since $d[V_i]_r = \omega_i^2 dr$. In the rest of the paper, we refer to (14) as the occupation time formula.

In addition to the Brownian motions U and $V = (V_1, \ldots, V_m)'$, we need to introduce another set of independent Brownian motions $W_1, \ldots, W_m$. Throughout the paper, the Brownian motions W_i will all be independent of U and V_i, and all will have the same variance $\sigma^2 = \sigma_u^2$.

Much of our subsequent theory is based on the following lemma, which contains several new results and characterizes the asymptotic behavior of various sample product moments involving nonlinear functions of integrated processes.

3.1 Lemma Suppose that x_t, u_t, w_t and d_t satisfy Assumptions 2.1 – 2.4, and $a_i, b_i : \mathbb{R} \to \mathbb{R}^{k_i}$ for $i = 1, \ldots, m$. Let a_i be I-regular, and b_i be H-regular with asymptotic order κ_i and limit homogeneous function h_i. Assume that $\kappa_{ni} = \kappa_i(\sqrt{n})$ is nonsingular, and h_i is piecewise differentiable with locally bounded derivative. Define $b_{ni} = \kappa_{ni}^{-1} b_i$ and $d_{nt} = \kappa_{nd}^{-1} d_t$. Then, the following hold as $n \to \infty$:

(a) $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} a_i(x_{it}) \to_d L_i(1,0) \int_{-\infty}^{\infty} a_i(s)\, ds.$

(b) $\frac{1}{n} \sum_{t=1}^{n} b_{ni}(x_{it}) \to_d \int_0^1 h_i(V_i(r))\, dr.$

(c) $\frac{1}{\sqrt[4]{n}} \sum_{t=1}^{n} a_i(x_{it}) u_t \to_d \left(L_i(1,0) \int_{-\infty}^{\infty} a_i(s) a_i(s)'\, ds \right)^{1/2} W_i(1).$

(d) $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} b_{ni}(x_{it}) u_t \to_d \int_0^1 h_i(V_i(r))\, dU(r).$

(e) $\frac{1}{n^{3/4}} \sum_{t=1}^{n} a_i(x_{it}) w_t' \to_p 0.$

(f) $\frac{1}{n} \sum_{t=1}^{n} b_{ni}(x_{it}) w_t' \to_p 0.$

(g) $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} d_{nt} a_i(x_{it}) = O_p(1).$

(h) $\frac{1}{n} \sum_{t=1}^{n} d_{nt} b_{ni}(x_{it}) \to_d \int_0^1 d(r) h_i(V_i(r))\, dr.$

(i) $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} a_i(x_{it}) a_i(x_{it})' \to_d L_i(1,0) \int_{-\infty}^{\infty} a_i(s) a_i(s)'\, ds.$

(j) $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} a_i(x_{it}) a_j(x_{jt})' \to_p 0$ for $i \neq j$.

(k) $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} a_i(x_{it}) b_{nj}(x_{jt})' = O_p(1).$

$$(\ell)\quad \frac{1}{n}\sum_{t=1}^{n} b_{ni}(x_{it})b_{nj}(x_{jt})' \to_d \int_0^1 h_i(V_i(r))h_j(V_j(r))'dr.$$

The weak convergence in (a), (b), (c), (d), (h), (i) and (ℓ) holds jointly.

The results in parts (a), (b), (d) and (i) of Lemma 3.1 are shown in Park and Phillips (1998,1999), and are included here for completeness. For each i, part (c) is also derived there. Our result shows that, in addition to the weak convergence, the limit processes W_i are independent of U for all i, and W_i and W_j are independent for $i \neq j$ in the representation of the limiting distribution. As indicated above, there are also some useful new results in Lemma 3.1. In particular, parts (e) – (h) are needed to obtain our limit theory for models with time trends and stationary regressors. Also, parts (i) – (ℓ) are used in dealing with multiple integrated regressors. Each of these is new.

Park and Phillips (1989) show that integrated regressors are asymptotically orthogonal to stationary regressors in linear regression models. That this result also applies between nonlinear functions of integrated regressors and stationary regressors is an interesting consequence of parts (e) and (f) of Lemma 3.1. Asymptotic orthogonality applies not only between integrated regressors and stationary regressors. In nonlinear regressions, it also applies between different types of functions of integrated regressors. Indeed, the result in part (k) shows that integrable and asymptotically homogeneous functions do not interact in the limit. If the functions involved are integrable, we even have asymptotic orthogonality between any transformations of two integrated regressors, regardless of how correlated these individual regressors may be. This follows from part (j). As will become apparent in the next section of the paper, these orthogonalities help to simplify the asymptotics of nonlinear regressions with multiple integrated and stationary processes.

The limiting distribution in part (c) is mixed normal. Note that, for each i, W_i is independent of V_i, as mentioned earlier, and hence W_i is also independent of L_i. However, the limiting distribution given in part (d) is a normal mixture only in the special case where V_i is independent of U. If this is the case, then we have

$$\int_0^1 h_i(V_i(r))dU(r) =_d \left(\int_0^1 h_i(V_i(r))h_i(V_i(r))'dr\right)^{1/2} W_i(1).$$

Independence of U and V_i holds when u_t is uncorrelated with future values of v_t, as well as its present and past values. This will, of course, rarely be satisfied in practical applications. So, in general, the distribution is not mixture normal and will depend on the correlation between U and V_i.

In addition to the results listed in Lemma 3.1, our subsequent theory also relies on the limits

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} d_{nt}u_t \to_d \int_0^1 d(r)\,dU(r) =_d N\left(0, \sigma^2\int_0^1 d(r)d(r)'dr\right) \tag{15}$$

and

$$\frac{1}{\sqrt{n}}\sum_{t=1} w_t u_t \to_d N(0;\sigma^2\Sigma_{ww}). \quad (16)$$

In view of Assumptions 2.1, (15) and (16) follow directly from the martingale central limit theorem of McLeish (1974). Moreover, since u_t is a martingale difference, and w_t is $\mathcal{F}_{t-1}$ measurable we have $E(w_t u_t u_s) = 0$ for all $t, s \geq 1$ in view of Assumption 2.1 (d). Also, under Assumption 2.1(a) and (b), we have $E(u_t v_{t+1+k}|\mathcal{F}_{t-1}) = \Phi_k \sigma_{\varepsilon u}$ for all $k \geq 0$ and 0 otherwise, and it follows from this and Assumption 2.1 (d) that $E(w_t u_t v_s) = 0$ for all $t, s \geq 1$. The limit distribution in (16) and $(U(r), V(r)), 0 \leq r \leq 1$, are therefore uncorrelated and, being Gaussian, are independent. Consequently, the limit distribution in (16) is independent of the limit distributions in (c), (d), (h), (i) and (j) of Lemma 3.1, and that in (15).

4. Asymptotic Theory

We suppose that model (6) is estimated by nonlinear least squares (NLS) and let

$$\hat{\theta}_n = (\hat{\pi}_n', \hat{\alpha}_n', \hat{\beta}_n')'$$

be the NLS estimator of $\theta = (\pi', \alpha', \beta')'$, i.e.,

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

obtained in the usual way by numerical optimization. Following convention, we assume here that the parameter set Θ is convex and compact, and that the true value, θ_0, is an interior point of Θ.

Let

$$Q_n(\theta) = \frac{1}{2}\sum_{t=1} (y_t - f(z_t;\theta))^2,$$

and define $\dot{Q}_n = \partial Q_n/\partial\theta$ and $\ddot{Q}_n = \partial^2 Q_n/\partial\theta\partial\theta'$. The first order Taylor expansion of $\dot{Q}_n(\hat{\theta}_n)$ around θ_0 yields

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

where θ_n lies on the line connecting θ_0 and $\hat{\theta}_n$. From this expansion we may obtain the asymptotic distribution of $\hat{\theta}_n$, under suitable regularity conditions.

To derive the asymptotic distribution of $\hat{\theta}_n$, we need to introduce some additional notation. Let $\dot{f}$ and $\dot{q} = (\dot{q}_i)$ be the derivatives of f, q and q_i respectively with respect to θ, β and β_i. Notice that

$$\dot{f}(z_t;\theta) = (d_t', w_t', \dot{q}(x_t;\beta)')'.$$

We let $\nu_{ni} = \sqrt{n}$ and $\sqrt{n}\,\kappa_{ni}$ respectively for $i \in I$ and $i \in H$. Let

$$\nu_n = \operatorname{diag}(\nu_{n1}, \ldots, \nu_{nm}),$$

and subsequently define

$$D_n = \operatorname{diag}\left(\sqrt{n}\,\kappa_{nd}, \sqrt{n} I, \nu_n\right),$$

where the partition is made conformably with $\dot{F}$ given above.

Define

$$M_n = D_n^{-1} \sum_{t=1}^{n} \dot{F}(z_t;\theta_0)\dot{F}(z_t;\theta_0)' D_n^{-1} \quad \text{and} \quad R_n = D_n^{-1} \sum_{t=1}^{n} \ddot{F}(z_t;\theta_0)\, u_t D_n^{-1},$$

where $\ddot{F} = \partial^2 F/\partial\theta\partial\theta'$, so that $M_n = D_n^{-1}\ddot{Q}_n(\theta_0)D_n^{-1} + R_n$, and let $Z_n = -D_n^{-1}Q_n(\theta_0)$. Moreover, let $D_{n\delta} = n^{-\delta}D_n$ for $\delta > 0$, and define

$$\Theta_n = \left\{\theta : \left\| D_{n\delta}^{-1}(\theta - \theta_0)\right\| \le 1\right\}.$$

Our model (6) includes integrated regressors, but the approach by Wooldridge (1994) for nonstationary regressions is still applicable. A trivial modification of his result is introduced below as a lemma for easy reference.

4.1 Lemma Suppose that

(a) $(M_n, Z_n) \to_d (M, Z)$,

(b) $R_n = o_p(1)$,

(c) $M > 0$ a.s., and

(d) $D_{n\delta}^{-1}\left(\ddot{Q}_n(\theta) - \ddot{Q}_n(\theta_0)\right)D_{n\delta}^{-1\prime} \to_p 0$ uniformly in $\theta \in \Theta_n$ for some $\delta > 0$.

Then we have $D_n(\hat{\theta}_n - \theta_0) \to_d M^{-1}Z$ as $n \to \infty$.

Once we show that the conditions of Lemma 4.1 are met for our model, the limiting distribution of $\hat{\theta}_n$ can be obtained readily. Conditions (a) and (b) follow immediately from the results in Lemma 3.1. Condition (c) is guaranteed by a suitable identification condition. The most difficult part is to establish condition (d).

To represent the limiting distributions of $(\hat{\beta}_i)_{i\in H}$ compactly, we let

$$\beta_H = (\beta_i)_{i\in H} \quad \text{and} \quad q_H(x;\beta) = (q_i(x_i;\beta_i))_{i\in H}$$

which are vectorized into column vectors by vertical stacking. Also, we define $\hat{\beta}_H^n$ and β_H^0 accordingly from $\hat{\beta}_i^n$ and β_i^0, respectively, for $i \in H$. Moreover, for $i \in H$ let

$$q_i(\lambda s;\beta_i) \approx \kappa_i(\lambda;\beta_i)h_i(s;\beta_i);$$

and let $\kappa_{ni} = \kappa_i(\sqrt{n};\beta_i^0)$. Then we define

$$\kappa_n = \operatorname{diag}(\kappa_{ni})_{i\in H} \quad \text{and} \quad h(x;\beta) = \left(h_i(x_i;\beta_i)\right)_{i\in H}.$$

The following theorem presents the asymptotic distribution of $\hat{\theta}_n$.

4.2 Theorem Under Assumptions 2.1 – 2.4, we have as $n \to \infty$

(a) $\sqrt{n}(\hat{\alpha}_n - \alpha_0) \to_d N\left(0, \sigma^2 \Sigma_{ww}^{-1}\right)$

(b) $\sqrt[4]{n}(\hat{\beta}_i^n - \beta_i^0) \to_d \left(L_i(0,1) \int_{-\infty}^{\infty} \dot{q}_i(s,\beta_i^0)\dot{q}_i(s,\beta_i^0)' ds \right)^{-1/2} W_i(1)$

(c) $\sqrt{n} \begin{pmatrix} \kappa_{nd}(\hat{\pi}_n - \pi_0) \\ \dot{\kappa}_n(\hat{\beta}_H^n - \beta_H^0) \end{pmatrix} \to_d \left(\int_0^1 N(r)N(r)' dr \right)^{-1} \int_0^1 N(r)\, dU(r)$

where $i \in I$ and $N(r) = (d(r)', \dot{h}(V(r),\beta_0)')'$. The convergences in (a) – (c) hold jointly, and the limit distribution in (a) is independent of those in (b) and (c).

The convergence rates of the NLS estimates for the coefficients of the stationary and deterministic regressors are given respectively by $\sqrt{n}$ and $\sqrt{n}\kappa_{nd}$, as in standard regressions. Those for the NLS estimates of the parameters in functions of the integrated regressors are different, however, and they are dependent upon the types of functions involved. For integrable functions of integrated regressors, the rate is $\sqrt[4]{n}$, i.e., an order of magnitude smaller than the usual rate $\sqrt{n}$ for the NLS estimates in the standard stationary regressions. For the asymptotically homogeneous functions, the convergence rates are determined by their asymptotic orders, and are given by $\sqrt{n}\dot{\kappa}_n$. For functions with increasing asymptotic orders, they are therefore faster than the usual rate $\sqrt{n}$.

The limiting distribution of the NLS estimate of the coefficients of the stationary regressors is normal. The NLS estimators for the parameters in the integrable functions of the integrated regressors have asymptotic distributions that are mixed normal with mixing variates essentially given by the local times of their limit Brownian motions. That the estimators have limiting normal or mixed normal distributions has important implication for asymptotic inference, i.e., it implies that the usual chi-square tests are possible. On the other hand, the asymptotic distributions of the NLS estimators of time trend coefficients or the parameters in the asymptotically homogeneous functions of the integrated regressors are generally non-Gaussian. Only in the very special case where the integrated regressors are strictly exogenous, do they have mixed normal distributions. The limiting distributions are thus biased, and the usual chi-squared approach to inference is not possible. The critical values of the tests are dependent upon nuisance parameters.

To look at the asymptotic results in Theorem 4.2 more closely, we consider the following regressions

$$\begin{aligned} y_t &= \alpha' w_t + u_t \\ y_t &= q_i(x_{it},\beta_i) + u_t \quad \text{for each } i \in I \\ y_t &= \pi' d_t + \sum_{i \in H} q_i(x_{it},\beta_i) + u_t. \end{aligned} \tag{17}$$

The set of regressions introduced in (17) are nothing but regressions on some of the additive terms in (6) separately. The first regression is one exclusively on stationary regressors, the second set of regressions are those on each I-regular function of an integrated regressor, and the third is the multiple regression on a deterministic trend and all H-regular functions of integrated regressors. We denote by $\tilde{\pi}_n$ the LS estimators of the parameters in the regressions (17).

4.3 Corollary Under Assumptions 2.1 – 2.4, the limiting distributions of $\hat{\pi}_n$ and $\tilde{\pi}_n$ are identical.

Corollary 4.3 implies that various asymptotic orthogonalities hold for the regressors included in our model (6). First, the stationary regressors are asymptotically orthogonal to any function of the integrated regressors, as well as to the deterministic trends. Therefore, the NLS estimator of the coefficients of stationary regressors in our model (6) behave asymptotically as if it were from a regression with only these variables. Naturally, the usual normal asymptotics apply to such regressions. Second, integrable functions of integrated processes are orthogonal in the limit not only to other asymptotically homogeneous functions of integrated regressors, but they themselves also become orthogonal each another. The NLS estimator of a parameter in any integrable function of an integrated regressor thus behaves as if there were no other functions of integrated regressors in the regression.

As an illustrative example, we look at the regression with deterministic part $d_t'\pi = \pi_1 + \pi_2 t$ and

$$q_1(x_1;\beta_1) = e^{-\beta_1 x_1^2}; \quad q_2(x_2;\beta_2) = \beta_2 x_2; \quad q_3(x_3;\beta_3) = \frac{1}{1+e^{-\beta_3 x_3}}; \tag{18}$$

where $\beta_1;\beta_2;\beta_3 \in \mathbb{R}_+$. We may assume w.l.o.g. that there are no stationary regressors, since they are asymptotically independent of the integrated parts of the model. To derive the asymptotics for the $\hat{\beta}_i$'s, it is more convenient to consider the regression with $q_1; q_2$ and

$$\begin{aligned} q_3^*(x_3;\beta_3) &= q_3(x_3;\beta_3) - 1\{x_3 \geq 0\} \\ &= \frac{1}{1+e^{-\beta_3 x_3}} 1\{x_3 < 0\} - \frac{e^{-\beta_3 x_3}}{1+e^{-\beta_3 x_3}} 1\{x_3 \geq 0\} \end{aligned} \tag{19}$$

in place of q_3.

The regression function q_3, indeed, does not satisfy the regularity conditions in Park and Phillips (1998). As one can easily see, $q_3(\cdot;\beta_3)$ is asymptotically homogeneous with asymptotic order 1 and limit homogeneous function $1\{\cdot \geq 0\}$. This is true for all values of β_3. The limit homogeneous function thus does not depend upon β_3, which implies that β_3 is not asymptotically identified in q_3. On the other hand, β_3 is identified in q_3^*. Obviously, the NLS estimators from the regression with $q_1; q_2$ and q_3 are identical to those from the regression with $q_1; q_2$ and q_3^*.

Now it is immediate from Theorem 4.2(b) that

$$\sqrt[4]{n}(\hat{\beta}_{1n} - \beta_1^0) \to_d \left(\frac{3\sqrt{2\pi}}{32\beta_{10}^{5/2}} L_1(1;0) \right)^{-1/2} W_1(1);$$

$$\sqrt{n}(\hat{\beta}_{3n} - \beta_3^0) \to_d \left(\frac{\pi^2 - 6}{18\beta_{30}^3} L_3(1,0) \right)^{-1/2} W_3(1),$$

where we write $\beta_{10} = \beta_1^0$ and $\beta_{30} = \beta_3^0$. Moreover, we have

$$\begin{pmatrix} n^{1/2}(\hat{\pi}_{1n} - \pi_1^0) \\ n^{3/2}(\hat{\pi}_{2n} - \pi_2^0) \\ n(\hat{\beta}_{2n} - \beta_2^0) \end{pmatrix} \to_d \left(\int_0^1 N(r)N(r)' dr \right)^{-1} \int_0^1 N(r)\, dU(r),$$

from Theorem 4.2(c), where $N(r) = (1, r, V_2(r)')'$.

Nonlinear models often include a function like

$$\beta_4 q_3(x_3, \beta_3) = \frac{\beta_4}{1 + e^{-\beta_3 x_3}}, \tag{20}$$

instead of q_3. Park and Phillips (1998) show that the regression with $\beta_4 q_3(x_3, \beta_3)$ in (20) is the same as the regression with two functions

$$\begin{aligned} \beta_4^0 q_3^*(x_3, \beta_3) &= \beta_4^0 \left(\frac{1}{1 + e^{-\beta_3 x_3}} 1\{x_3 < 0\} - \frac{e^{-\beta_3 x_3}}{1 + e^{-\beta_3 x_3}} 1\{x_3 \geq 0\} \right) \\ q_4(x_3, \beta_4) &= \beta_4 1\{x_3 \geq 0\} \end{aligned}$$

where q_3^* is defined in (19). Their result easily extends to our general regression, which one can easily see from the proof of Theorem 4.2. That is, the regression with $q_1(x_1, \beta_1), q_2(x_2, \beta_3)$ and $\beta_4 q_3(x_3, \beta_3)$ respectively in (18) and (20) behaves asymptotically the same as the regression on $q_1(x_1, \beta_1), q_2(x_2, \beta_2), \beta_4^0 q_3^*(x_3, \beta_3)$ and $q_4(x_3, \beta_4)$. The limiting distributions of the parameters can therefore be easily obtained.

5. Efficient Estimation

In this section, we develop an efficient method of estimating our model (6). The usual NLS estimator $\hat{\mu}_n$ considered in Section 4 is generally not efficient, because it does not utilize the presence of the unit root in the regressors. We may obtain a more efficient estimator if such information is used, in the same manner as shown by Phillips (1991) for linear cointegrating regressions. Inefficiency of the estimator also results in invalidity of the usual t- or chi-squared test on the parameter in the asymptotically homogeneous regression functions. We now introduce a more efficient estimator (in the sense of Phillips, 1991) along the lines of Phillips and Hansen (1990) and Park (1992). The estimator has a mixed normal limit distribution and thus yields asymptotically valid t- or chi-square tests in the usual manner.

Here we concentrate on the parameter in the nonlinear functions of integrated regressors, together with those of the deterministic regressors. As shown in Section 4, the limiting distribution for the estimated coefficient of the stationary regressors is not affected by the inclusion of the regressors with trends. Due to this asymptotic orthogonality, we may not increase the efficiency of the coefficient estimator for the stationary regressors by utilizing information on the presence of the unit root in the nonstationary regressors. In what follows, we simply assume that $p(w_t; \alpha_0) = 0$ and is known in (6). This is just to make our exposition simple. Our subsequent methodology is applicable for the regression with stationary regressors, if we run the first step regression to get $\hat\alpha_n$ and consider the regression

$$y_t - p(w_t; \hat\alpha_n) = \tau(t; \pi) + q(x_t; \beta) + u_t$$

to efficiently estimate π and β.

For the development of our method, we need to introduce some additional assumptions on the innovation process $v_t = \Phi(L)\varepsilon_t$ of the integrated regressor x_t defined in (9).

5.1 Assumption We assume that

(a) $\Phi(z)$ is bounded and bounded away from zero for $|z| \leq 1$, and

(b) if we write $\Phi(z)^{-1} = 1 - \sum_{k=1}^{\infty} \Pi_k z^k$, then $\ell^s \sum_{k=\ell+1}^{\infty} \|\Pi_k\|^2 < \infty$ for some $s \geq 9$.

To estimate our model efficiently, we first run the regression

$$v_t = \hat\Pi_1 v_{t-1} + \cdots + \hat\Pi_\ell v_{t-\ell} + \hat\varepsilon_{\ell,t}$$

where we let ℓ increase as $n \to \infty$. More precisely, we let $\ell = n^\delta$, and select δ so that

$$\frac{r+2}{2r(s-3)} < \delta < \frac{r}{6+8r}, \tag{21}$$

where r is given by the moment condition for (ε_t), i.e., $E\|\varepsilon_t\|^r < \infty$ for some $r > 8$ as given in Assumption 2.3. It is easy to see that δ satisfying condition (21) exists for all $r > 8$, if $s \geq 9$ as is assumed in Assumption 5.1. For Gaussian ARMA models, Assumptions 2.3 and 5.1 hold for any finite r and s. Then, we may choose any δ such that $0 < \delta < 1/8$.

We define

$$y_t^* = y_t - \hat\sigma_{u\varepsilon} \hat\Sigma_{\varepsilon\varepsilon}^{-1} \hat\varepsilon_{\ell,t+1},$$

where

$$\hat\sigma_{u\varepsilon} = \frac{1}{n} \sum_{t=1}^{n} \hat u_t \hat\varepsilon'_{\ell,t+1} \quad \text{and} \quad \hat\Sigma_{\varepsilon\varepsilon} = \frac{1}{n} \sum_{t=1}^{n} \hat\varepsilon_{\ell,t} \hat\varepsilon'_{\ell,t},$$

with the first step NLS residual $\hat u_t$. Then consider the regression

$$y_t^* = f(z_t; \theta_0) + u_t^*, \tag{22}$$

in place of (6).

The efficient estimator that we propose is the NLS estimator $\hat\theta_n^*$ of θ_0 computed from the transformed regression (22), which we call efficient nonstationary nonlinear least squares (EN-NLS) estimator. In its motivation, the EN-NLS is closely related to the FM-OLS method by Phillips and Hansen (1990) and the CCR method by Park (1992), which both yield efficient estimates for the coefficients in the linear cointegrating regression. Just as for the FM-OLS and CCR methods, the EN-NLS corrects the long-run dependency between the regression errors and the innovations of the integrated regressors. However, the EN-NLS achieves the goal while maintaining the martingale difference condition on the regression errors. The latter condition is more important for nonlinear nonstationary regressions, in contrast to linear cointegrating regressions where we may allow for more general error processes in FM-OLS and CCR regressions.

Theorem 5.2 below presents the limit theory for the EN-NLS estimator $\hat\theta_n^*$. Let $\hat\theta_n^* = (\hat\pi_n^{*\prime}, \hat\beta_n^{*\prime})'$ and define $\hat\beta_{nI}^*$ and $\hat\beta_{nH}^*$ similarly as before. Recall that we assume there are no stationary regressors in the regression. Also, define

$$U^* = U - \omega_{uv}\omega_{vv}^{-1}V.$$

The process U^* is independent of V, and its variance is given by

$$\sigma_*^2 = \sigma_u^2 - \omega_{uv}\omega_{vv}^{-1}\omega_{vu},$$

i.e., the long-run conditional variance of U given V. In the same way as $W_1, \ldots, W_m$, let $W_1^*, \ldots, W_m^*$ be an independent set of Brownian motions that are independent of V and have the variance σ_*^2.

5.2 Theorem Under Assumptions 2.1 – 2.4 and 5.1, we have

(a) $\sqrt[4]{n}(\hat\beta_{ni}^* - \beta_{i0}) \to_d \left(L_i(0,1)\int_{-\infty}^{\infty} \dot g_i(s,\beta_{i0})\dot g_i(s,\beta_{i0})'ds\right)^{-1/2} W_i^*(1)$

(b) $\sqrt{n}\begin{pmatrix}\kappa_{nd}(\hat\pi_n^* - \pi_0)\\ \nu_n(\hat\beta_{nH}^* - \beta_{H0})\end{pmatrix} \to_d \left(\int_0^1 N(r)N(r)'dr\right)^{-1}\int_0^1 N(r)\,dU^*(r)$

where $i \in I$ and $N(r) = (d(r)', \dot h(V(r),\beta_0)')'$. The convergences in (a) and (b) hold jointly.

Asymptotic variances are seen to be reduced for the EN-NLS estimates of the parameters in the integrable functions of integrated regressors. Strict variance reduction occurs whenever the regression errors are correlated with the future or past as well as the present values of the innovations of the integrated regressors. We have similar reductions in the asymptotic variances of the EN-NLS estimators for the time trend coefficients and the parameters in the asymptotically homogeneous functions of integrated regressors. In particular, the estimators have limit distributions that are unbiased and mixed normal. Their asymptotic bias and non-normality are thus removed by the EN-NLS method. This, in particular, implies that the usual t- and chi-square tests are now possible for all parameters using these EN-NLS estimators.

There are other ways of obtaining asymptotically efficient estimators for the parameters in the nonlinear regression (6). It is indeed easy to see that we would get an estimator which is asymptotically equivalent to the EN-NLS estimator if we were to fit the nonlinear regression

$$y_t = f(z_t;\theta_0) + \varepsilon'_{\cdot,t+1}\delta_0 + e_t \tag{23}$$

Note that we may set $e_t \approx u_t^*$ with $\delta_0 = \Sigma^{-1}\sigma_{\cdot u}$, since $u_t = \varepsilon'_{\cdot,t+1}\delta_0 + e_t$. However, fitting (23) involves the estimation of a nonlinear regression with an additional regressor, and seems less preferable to EN-NLS which directly corrects the regressand. It is also possible, at least theoretically, to get an asymptotically efficient estimator by including leads and lags of Δx_t in the regression (as in Phillips and Loretan, 1991, Saikkonen, 1991, and Stock and Watson, 1993). Such a scheme, however, requires that the number of included leads and lags of Δx_t increases as the sample size $n \to \infty$. This makes it necessary to estimate large dimensional nonlinear regressions, which seems undesirable and impractical especially when the sample size is large.

6. Simulations

This section investigates the finite sample performance of the NLS and the EN-NLS estimators in a specific nonlinear regression model. We choose the following additive nonlinear model that combines both integrable and asymptotically homogeneous functions, thereby having some of the elements of our general theory. The model has the form

$$y_t = \pi_0 + \exp(-\beta_1^0 x_{1t}^2) + \frac{\beta_2^0}{1+\exp(-x_{2t})} + u_t. \tag{24}$$

The regression error u_t and integrated regressor x_t, with $x_t = (x_{1t}, x_{2t})'$, are generated from

$$u_t = \varepsilon_{0,t+1}/\sqrt{2} + (\varepsilon_{1,t+1} + \varepsilon_{2,t+1})/2;$$

and

$$\Delta x_t = v_t = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} + \begin{pmatrix} 0.2 & 0 \\ 0 & 0.5 \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{pmatrix},$$

where (ε_{0t}), (ε_{1t}) and (ε_{2t}) are drawn from independent $N(0,\sigma^2)$ distributions with $\sigma^2 = 0.1^2$. The true values of the parameters are set at $\pi_0 = 0$, $\beta_1^0 = 1$ and $\beta_2^0 = 0$.

By construction, the regression error u_t is a martingale difference sequence and is asymptotically correlated with the innovations that generate the integrated processes x_t. The Gaussian function $\exp(-\beta_1 x_{1t}^2)$ appearing in the simulation model (24) belongs to the I-regular class. The logistic function $\beta_2/(1+e^{-x_{2t}})$ in the model is asymptotically homogeneous with asymptotic order 1 and with limit homogeneous function $\beta_2 1\{x_2 \geq 0\}$ and it belongs to the H-regular class. As shown in Theorem 4.2, the NLS estimator for the parameter β_1 inside the integrable function converges at a slower rate $n^{1/4}$ to a mixed normal distribution, even when the corresponding integrated process x_{1t} is asymptotically correlated with the regression error. However, due to the asymptotic correlation of u_t and x_{2t}, the NLS estimator for the constant term π and the coefficient β_2 on the logistic function converge at the rate $\sqrt{n}$ to non-Gaussian limit distributions, which are biased.

As shown in Theorem 5.2, the limit distributions of the EN-NLS estimators for π and β_2, as well as that for β_1, are mean-zero mixed normal. Moreover, the EN-NLS estimators for all three parameters π, β_1 and β_2 have reduced variances, compared with the NLS estimators. More precisely, for the DGP used for our simulation, we have $\sigma^2 = 0.1^2$ and $\sigma_{\cdot}^2 = 0.1^2/2$. This implies that the variance of the errors in the transformed regression (22) is one-half in magnitude of that in the original regression (6). The utilization of the information on the presence of the unit root in the regressors may thus reduce the error variance in half.

In the simulation, the samples of sizes 250 and 500 are drawn 5000 times to estimate the NLS and EN-NLS estimators and t-statistics based on these estimators. For the construction of the EN-NLS correction terms, we use one-period ahead fitted innovations $\hat{\varepsilon}_{\ell,t+1}$, which is obtained from the ℓ-th order vector autoregressions of v_t with $\ell = 1, 2$, respectively for $n = 250$ and $n = 500$. For the nonlinear estimation, we use the GAUSS Optimization routine and the Gauss-Newton algorithm. The simulation results are summarized in Figures 1–3. Figures 1 and 2 present the density estimates for the NLS and EN-NLS estimators for sample sizes $n = 250$ and $n = 500$, respectively. The estimators are scaled with their respective convergence rates, and the density estimates for the scaled estimators are included also in Figures 1 and 2.

The finite sample performances of the NLS and the EN-NLS estimators are much as would be expected from the limit theory. As can be seen from Figures 1 and 2, the sampling distributions of the estimators well reflect their theoretical convergence rates. The estimators for the parameter β_1 inside the integrable function converge slower than the estimators for the intercept π and the coefficient β_2 on the logistic function. As expected, the finite sample distribution of the NLS estimator for β_1 is symmetric and well centered. However, those for π and β_2 are skewed and suffer from bias that does not seem to vanish as the sample size increases. On the other hand, the empirical distributions of the EN-NLS for π and β_2 as well as for β_1 are all symmetric and noticeably more concentrated around zero both in the small and large samples. Our EN-NLS estimators thus seem more efficient than their NLS counterparts.

The sampling behavior of the t-statistics based on the NLS and EN-NLS estimators also corroborate our limit theory. As can be seen clearly from Figure 3, the finite sample distributions of the t-statistics based on the NLS estimators for π and β_2 are noticeably biased, while those based on the NLS estimators for β_1 are symmetric and well centered, reasonably well approximating their limit $N(0, 1)$ distribution. This is as expected from the limit theory of the NLS estimators. On the other hand, the sampling distributions of the t-statistics based on all of the EN-NLS estimators approximate closely their limit $N(0, 1)$ distribution, and the approximation becomes even closer as the sample size increases. This indicates that our EN-NLS correction works in nonstationary nonlinear regressions, which in turn implies that we can conduct conventional hypothesis testing using statistics, such as t-statistics or χ^2 tests, constructed from the EN-NLS estimators.

7. Conclusions

This paper develops an asymptotic theory of regression for models with deterministic trends, stationary regressors, and nonstationary integrated regressors. In many respects, this work continues the program of research that started with the Phillips and Durlauf (1986) study of multiple regressions with integrated time series. One of the conclusions of their study was that while least squares regression in models with nonstationary regressors produces consistent estimates, and at the faster $O_p(n^{-1})$ rate than in models with stationary regressors, the limit distribution theory is nonnormal and standard tests no longer typically yield asymptotic χ^2 criteria. The present paper shows that, when the models are nonlinear, least squares regression continues to be consistent, but the rates of convergence can be slower as well as faster than the conventional $O_p(n^{-1/2})$ rate of stationary time series regression. Thus, the nature of the nonlinearity plays a major role in the asymptotic theory, sometimes attenuating the signal and sometimes strengthening the signal from a nonstationary regressor. Moreover, as in the linear theory, least squares regression yields estimators that are generally inefficient and produce invalid statistical tests. The next step forward from the Phillips-Durlauf regression theory was the development of modified linear regressions that produced efficient estimates and asymptotic χ^2 test criteria. That step was taken in the work of Phillips and Hansen (1990) and Park (1992) with fully modified least squares regression and CCR estimation, and it has had many empirical applications. The present paper proposes a related methodology for nonlinear regression, leading to the EN-NLS estimator discussed in Section 5. The approach is simple to implement, produces efficient estimates and leads to test statistics that are asymptotical χ^2 test criteria.

The analytic framework of Phillips and Durlauf (1986) also made headway by allowing for the nonparametric treatment of weak dependence in the regression errors. The theory was therefore applicable in the context of quite general cointegrating regressions with stationary errors. The present theory for the nonlinear case is more delimited. Our analytic framework allows for martingale difference regression errors and is therefore more directly suited to the estimation of nonlinear equations that arise from the solution of rational expectations or dynamic stochastic general equilibrium models. In such cases, it is more natural to take the innovations on an efficient macroeconomic equilibrium, for example, to be martingale differences. Extensions of our theory to accommodate the more general context of broad empirical relationships between time series variables that move together over time, but possibly in a nonlinear manner, is important and it is an ongoing area of research for the authors.

8. Mathematical Proofs

8.1 Proof of Lemma 3.1

In the proofs of parts (e), (f), (g), (h), (j) and (k), we assume that a_i and b_i are scalar-valued. The proofs for the vector-valued a_i and b_i can be done simply by looking at each component of them separately.

Proofs of (a)-(d) The proofs are in Park and Phillips (1999). For (c), we must show that W_i is independent of V_j, $j \neq i$, as well as of V_i. This, however, is straightforward from their derivation. ■

Proof of (e) Under Assumption 2.1(c), we have

$$E\|w_t\|^r < \infty,$$

as is well known. Let $c_n = n^{\delta}$ with some δ such that $1/4 < \delta < 1/4(s-1)$, and write

$$\frac{1}{n^{3/4}}\sum_{t=1}^{n} a_i(x_{it})w_t = \frac{1}{n^{3/4}}\sum_{t=1}^{n} a_i(x_{it})w_t 1\{\|w_t\| \le c_n\} + \frac{1}{n^{3/4}}\sum_{t=1}^{n} a_i(x_{it})w_t 1\{\|w_t\| > c_n\}. \tag{25}$$

We have

$$\frac{1}{n^{3/4}}\sum_{t=1}^{n} \|a_i(x_{it})w_t\| 1\{\|w_t\| \le c_n\} \le \frac{c_n}{\sqrt[4]{n}}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} |a_i(x_{it})| \to_p 0; \tag{26}$$

since $\delta < 1/4$. Moreover,

$$\begin{aligned}\frac{1}{n^{3/4}}\sum_{t=1}^{n} \|a_i(x_{it})w_t\| 1\{\|w_t\| > c_n\} &\le \|a_i\|\frac{1}{n^{3/4}}\sum_{t=1}^{n} \|w_t\| 1\{\|w_t\| > c_n\} \\ &\lesssim \sqrt[4]{n}\,\frac{\sup_{1\le t\le n}\|w_t\|^r}{c_n^{r-1}} \to 0;\end{aligned} \tag{27}$$

since $\delta > 1/4(r-1)$. The stated result follows immediately from (26) and (27), due to (25). ■

Proof of (f) First assume that h_i is differentiable with locally bounded derivative $\dot{h}_i$, and define

$$K_i = \left[\min_{0\le r\le 1} V_i(r) - 1, \max_{0\le r\le 1} V_i(r) + 1\right]. \tag{28}$$

We have

$$\frac{1}{n}\sum_{t=1}^{n} b_{ni}(x_{it})w_t = \frac{1}{n}\sum_{t=1}^{n} h_i\left(\frac{x_{it}}{\sqrt{n}}\right) w_t + o_p(1); \tag{29}$$

as shown in Park and Phillips (1998). Also, if we let

$$M_n = \frac{1}{n}\sum_{t=1}^{n} h_i\left(\frac{x_{i,t-1}}{\sqrt{n}}\right) w_t;$$

then it follows that

$$\frac{1}{n}\sum_{t=1}^{n} h_i\left(\frac{x_{it}}{\sqrt{n}}\right) w_t = M_n + o_p(1); \tag{30}$$

since

$$\frac{1}{n}\sum_{t=1}^{n}\left\|\left(h_i\left(\frac{x_{it}}{\sqrt{n}}\right) - h_i\left(\frac{x_{i,t-1}}{\sqrt{n}}\right)\right) w_t\right\| \le \frac{1}{\sqrt{n}}\left\|\dot{h}_i\right\|_{K_i}\frac{1}{n}\sum_{t=1}^{n}\|v_{it}w_t\| \to_p 0;$$

as $n \to \infty$.

Due to (29) and (30), it suffices to show

$$M_n \to_p 0 \tag{31}$$

to establish the stated result. To prove (31), we let

$$w_t = \Psi(1)\eta_t + (\bar{w}_{t-1} - \bar{w}_t);$$

as in Phillips and Solo (1992), and write $M_n = A_n + B_n$ with

$$\begin{aligned} A_n &= \Psi(1)\frac{1}{n}\sum_{t=1}^{n} h_i\left(\frac{x_{i,t-1}}{\sqrt{n}}\right)\eta_t \\ B_n &= \frac{1}{n}\sum_{t=1}^{n} h_i\left(\frac{x_{i,t-1}}{\sqrt{n}}\right)(\bar{w}_{t-1} - \bar{w}_t) \\ &= \frac{1}{n}\sum_{t=1}^{n}\left(h_i\left(\frac{x_{it}}{\sqrt{n}}\right) - h_i\left(\frac{x_{i,t-1}}{\sqrt{n}}\right)\right)\bar{w}_t - \frac{1}{n}h_i\left(\frac{x_{in}}{\sqrt{n}}\right)\bar{w}_n. \end{aligned}$$

It follows directly from Park and Phillips (1998) that $A_n \to_p 0$. Moreover, it is not difficult to see that $B_n \to_p 0$, since

$$\frac{1}{n}\sum_{t=1}^{n}\left\|\left(h_i\left(\frac{x_{it}}{\sqrt{n}}\right) - h_i\left(\frac{x_{i,t-1}}{\sqrt{n}}\right)\right)\bar{w}_t\right\| \le \frac{1}{\sqrt{n}}\left\|\dot{h}_i\right\|_{K_i}\frac{1}{n}\sum_{t=1}^{n}\|v_{it}\bar{w}_t\| \to_p 0;$$

and

$$\left\|\frac{1}{n}h_i\left(\frac{x_{in}}{\sqrt{n}}\right)\bar{w}_n\right\| \le \|h_i\|_{K_i}\frac{1}{n}\|\bar{w}_n\| \to_p 0.$$

The result in (31) is therefore proved.

Next, we show that (29) – (31) hold for $h_i(x) = 1\{x \ge 0\}$. Clearly, the stated result would then follow for piecewise differentiable functions with locally bounded support. It follows from Park and Phillips (1998) that (29) holds. To deduce (30), it suffices to show that $R_n \to_p 0$, where

$$R_n = \frac{1}{n}\sum_{t=1}^{n} 1\{x_{it} \ge 0, x_{i,t-1} < 0\} + 1\{x_{it} < 0, x_{i,t-1} \ge 0\}.$$

This is because

$$\left\|\frac{1}{n}\sum_{t=1}^{n}\left(h_i\left(\frac{x_{it}}{\sqrt{n}}\right)-h_i\left(\frac{x_{i,t-1}}{\sqrt{n}}\right)\right)w_t\right\| \le \left(\frac{1}{n}\sum_{t=1}^{n}\left|h_i\left(\frac{x_{it}}{\sqrt{n}}\right)-h_i\left(\frac{x_{i,t-1}}{\sqrt{n}}\right)\right|^2\right)^{1/2}\left(\frac{1}{n}\sum_{t=1}^{n}\|w_t\|^2\right)^{1/2};$$

by Cauchy-Schwarz, and $(1/n)\sum_{t=1}^{n}\|w_t\|^2 = O_p(1)$.

To show $R_n \to_p 0$, we let $c_n = n^{-\delta}$ for some $0 < \delta < 1/2$, and bound R_n by $R_n \le S_n + T_n$, where

$$S_n = \frac{1}{n}\sum_{t=1}^{n} 1\left\{\left|\frac{x_{i,t-1}}{\sqrt{n}}\right| < c_n\right\} \quad \text{and} \quad T_n = \frac{1}{n}\sum_{t=1}^{n} 1\left\{\left|\frac{v_{it}}{\sqrt{n}}\right| \ge c_n\right\};$$

since, if $|v_{it}| < c_n\sqrt{n}$, x_{it} can change sign from $x_{i,t-1}$ only when $|x_{i,t-1}| < c_n\sqrt{n}$. Note that

$$\begin{aligned} S_n &=_d \int_0^1 1\{|V_{ni}(r)| \le c_n\}dr \\ &= \int_0^1 1\{|V_i(r)| \le c_n\}dr + o_p(1) \to_p 0; \end{aligned}$$

due, in particular, to Lemma 2.5(b) in Park and Phillips (1999). Moreover,

$$E|T_n| \le \frac{1}{n}\sum_{t=1}^{n}\Pr\left\{\left|\frac{v_{it}}{\sqrt{n}}\right| \ge c_n\right\} \le \frac{\sup_{1\le t\le n} E|v_{it}|}{c_n\sqrt{n}} \to 0;$$

since $c_n\sqrt{n} \to \infty$.

Finally, we show that $A_n, B_n \to_p 0$ to deduce (31) for $h_i(x) = 1\{x \ge 0\}$. That $A_n \to_p 0$ follows directly from Park and Phillips (1998). It is also not difficult to see that $B_n \to_p 0$. We then have, by Cauchy-Schwarz

$$\left\|\frac{1}{n}\sum_{t=1}^{n}\left(h_i\left(\frac{x_{it}}{\sqrt{n}}\right)-h_i\left(\frac{x_{i,t-1}}{\sqrt{n}}\right)\right)w_t\right\| \le \left(\frac{1}{n}\sum_{t=1}^{n}\left|h_i\left(\frac{x_{it}}{\sqrt{n}}\right)-h_i\left(\frac{x_{i,t-1}}{\sqrt{n}}\right)\right|^2\right)^{1/2}\left(\frac{1}{n}\sum_{t=1}^{n}\|w_t\|^2\right)^{1/2};$$

which, together with $R_n \to_p 0$, implies that the first term of B_n converges in probability to zero. Note that $E\|w_t\|^2 < \infty$ under Assumptions 2.1 and 2.2, and we have $(1/n)\sum_{t=1}^{n}\|w_t\|^2 = O_p(1)$. Obviously, the second term of B_n is of order $o_p(1)$ since $h_i(x) = 1\{x \ge 0\}$ is bounded. ∎

Proof of (g) This is immediate since

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\|d_{nt}a_i(x_{it})\| \le \left(\sup_{n\ge 1}\sup_{1\le t\le n}\|d_{nt}\|\right)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}|a_i(x_{it})|;$$

due to Assumption 2.4. ■

Proof of (h) We have

$$\frac{1}{n}\sum_{t=1}^{n} d_{nt}b_{ni}(x_{it}) = \frac{1}{n}\sum_{t=1}^{n} d_{nt}h_i\left(\frac{x_{it}}{\sqrt{n}}\right) + o_p(1);$$

which follows easily from the proof of Park and Phillips (1999), since $\sup_{1\le t\le n}\|d_{nt}\| < \infty$. Moreover, we have

$$\frac{1}{n}\sum_{t=1}^{n} d_{nt}h_i\left(\frac{x_{it}}{\sqrt{n}}\right) = \int_0^1 d(r)h_i(V_{ni}(r))\,dr + o_p(1);$$

because d_n converges in L^2. The stated result now follows immediately from the continuous mapping theorem. ■

Proof of (i) and (ℓ) The proofs are in Park and Phillips (1998). ■

Proof of (j) The proof heavily relies on the results in the proof of Theorem 5.1 in Park and Phillips (1998), which will simply be referred to PP in what follows. Let κ_n and δ_n be given as in PP, and denote by a_{ni} the simple function approximating a_i over the truncated support $[-\kappa_n, \kappa_n)$, i.e.,

$$a_{ni}(x) = \sum_{k=-\kappa_n}^{\kappa_n - 1} a_i(k\delta_n)1\{k\delta_n \le x < (k+1)\delta_n\}.$$

We may easily deduce

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} a_i(x_{it})a_j(x_{jt}) &=_d \sqrt{n}\int_0^1 a_i(\sqrt{n}V_{ni}(r))a_j(\sqrt{n}V_{nj})\,dr \\ &= \sqrt{n}\int_0^1 a_{ni}(\sqrt{n}V_{ni}(r))a_{nj}(\sqrt{n}V_{nj})\,dr + o_p(1); \qquad (32)\end{aligned}$$

as in (31)–(33) of PP, since a_i and a_j are both bounded.

To simplify notation in the subsequent derivation of our result, we make the convention $|\int 1(A) - \int 1(B)| = \int |1(A) - 1(B)|$ for indicators $1(A)$ and $1(B)$ on A and B. All the approximation results on the integrals of indicators in PP, which are in turn based on Akonom (1993), hold in the sense of $\int |1(A) - 1(B)|$ as well as that of $|\int 1(A) - \int 1(B)|$. If we let

$$\begin{aligned} N_{ni}(k;\delta_n) &= \int_0^1 1\{k\delta_n \le \sqrt{n}V_{ni}(r) < (k+1)\delta_n\} \\ N_n(k;\ell;\delta_n) &= \int_0^1 1\{k\delta_n \le \sqrt{n}V_{ni}(r) < (k+1)\delta_n\} 1\{\ell\delta_n \le \sqrt{n}V_{nj}(r) < (\ell+1)\delta_n\} dr; \end{aligned}$$

then we have

$$\begin{aligned} |N_n(k;\ell;\delta_n) - N_n(0;0;\delta_n)| \;\le\; & |N_{ni}(k;\delta_n) - N_{ni}(0;\delta_n)|^{1/2} N_{nj}(\ell;\delta_n)^{1/2} \\ & + N_{ni}(0;\delta_n)^{1/2} |N_{nj}(\ell;\delta_n) - N_{nj}(0;\delta_n)|^{1/2}; \end{aligned} \tag{33}$$

due to our convention. Moreover, if we define

$$\begin{aligned} N_i(k;\delta_n) &= \int_0^1 1\{k\delta_n \le \sqrt{n}V_i(r) < (k+1)\delta_n\} \\ N(k;\ell;\delta_n) &= \int_0^1 1\{k\delta_n \le \sqrt{n}V_i(r) < (k+1)\delta_n\} 1\{\ell\delta_n \le \sqrt{n}V_j(r) < (\ell+1)\delta_n\} dr; \end{aligned}$$

then it follows that

$$\begin{aligned} |N_n(0;0;\delta_n) - N(0;0;\delta_n)| \;\le\; & |N_{ni}(0;\delta_n) - N_i(0;\delta_n)|^{1/2} N_{nj}(0;\delta_n)^{1/2} \\ & + N_i(0;\delta_n)^{1/2} |N_{nj}(0;\delta_n) - N_j(0;\delta_n)|^{1/2}; \end{aligned} \tag{34}$$

under the convention.

It is tedious, but rather straightforward, to show that

$$\begin{aligned} & \sqrt{n}\int_0^1 a_{ni}(\sqrt{n}V_{ni}(r))a_{nj}(\sqrt{n}V_{nj}(r))\,dr \\ &= \sqrt{n}\sum_{k=-\kappa_n}^{\kappa_n-1}\sum_{\ell=-\kappa_n}^{\kappa_n-1} a_i(k\delta_n)a_j(\ell\delta_n)N_n(k;\ell;\delta_n) \\ &= \left(\sum_{k=-\kappa_n}^{\kappa_n-1}\sum_{\ell=-\kappa_n}^{\kappa_n-1} a_i(k\delta_n)a_j(\ell\delta_n)\right)\sqrt{n}N_n(0;0;\delta_n) + o_p(1) \\ &= \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a_i(r)a_j(s)\,dr\,ds\right)(\sqrt{n}/\delta_n^2)N_n(0;0;\delta_n) + o_p(1); \end{aligned} \tag{35}$$

in the same way as (34) and the first line of (35) in PP, using (33).

Now choose δ_n so that

$$(\sqrt{n}/\delta_n^3)\,|N_{ni}(0;\delta_n) - N_i(0;\delta_n)| = o_p(1):$$

Then since

$$(\sqrt{n}/\delta_n)\,|N_{ni}(0;\delta_n)|;\ (\sqrt{n}/\delta_n)\,|N_i(0;\delta_n)| = O_p(1);$$

we have from (34) that

$$(\sqrt{n}/\delta_n^2)\,|N_n(0,0;\delta_n) - N(0,0;\delta_n)| \to_p 0. \tag{36}$$

However, for large n,

$$N(0,0;\delta_n) \le N(0,0;1) = O_p(\log n/n) \tag{37}$$

as shown in Kasahara and Kotani (1979). The stated result now follows easily from (32), (35), (36) and (37). ■

Proof of (k) Let K_i be defined as in the proof of (f), and note that

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^n |a_i(x_{it})b_{nj}(x_{jt})| &= \frac{1}{\sqrt{n}}\sum_{t=1}^n |a_i(x_{it})|\left|h_j\left(\frac{x_{jt}}{\sqrt{n}}\right)\right| + o_p(1)\\ &\le \|h_j\|_{K_j}\frac{1}{\sqrt{n}}\sum_{t=1}^n |a_i(x_{it})| + o_p(1) = O_p(1);\end{aligned}$$

as was to be shown. ■

8.2 Proof of Lemma 4.1

See Theorem 8.1 of Wooldridge (1994). ■

8.3 Proof of Theorem 4.2

We only need to prove that the condition in Lemma 4.1(d) holds. The conditions (a)–(c) are easy to show, given the results in Lemma 3.1. The stated asymptotic distributions then follow immediately from Lemma 3.1 and Lemma 4.1.

Let ε_i be given by Assumption H(a) if q_i satisfies the condition there, and let otherwise it be any number 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 $0 < \delta < \varepsilon/6$. It is shown in Park and Phillips (1998, proof of Lemma 6.3) that

$$\begin{aligned}\sum_{i=1}^{n}\left\|\nu_{in}^{-1}\dot{q}_i(x_{it},\theta_i^0)\right\|^2 &= O_p(1)\\ n^{\varepsilon}\sup_{\theta\in\Theta_n}\sum_{t=1}^{n}\left\|(\nu_{ni}\otimes\nu_{ni})^{-1}\ddot{q}_i(x_{it},\theta_i)\right\|^2 &= o_p(1);\end{aligned} \tag{38}$$

and that, due to Cauchy-Schwarz,

$$n^{\varepsilon/2}\sup_{\theta\in\Theta_n}\sum_{t=1}^{n}\left\|\nu_{ni}^{-1}\dot{q}_i(x_{it},\theta_i^0)\right\|\left\|(\nu_{nj}\otimes\nu_{nj})^{-1}\ddot{q}_i(x_{jt},\theta_j)\right\| = o_p(1). \tag{39}$$

Furthermore, since

$$\sum_{t=1}^{n} \left\| n^{-1/2}\kappa_{nd}^{-1} d_t \right\|^2 = O(1) \quad \text{and} \quad \sum_{t=1}^{n} \left\| n^{-1/2} w_t \right\|^2 = O_p(1),$$

we may easily deduce that

$$n^{\epsilon/2} \sup_{\theta\in\Theta_n} \sum_{t=1}^{n} \left\| n^{-1/2} w_t \right\| \left\| (\nu_{ni} \otimes \nu_{ni})^{-1} \ddot{q}_i(x_{it}, \beta_i) \right\| = o_p(1) \qquad (40)$$

$$n^{\epsilon/2} \sup_{\theta\in\Theta_n} \sum_{t=1}^{n} \left\| n^{-1/2}\kappa_{nd}^{-1} d_t \right\| \left\| (\nu_{ni} \otimes \nu_{ni})^{-1} \ddot{q}_i(x_{it}, \beta_i) \right\| = o_p(1), \qquad (41)$$

by simple applications of Cauchy-Schwarz.

We now let

$$\ddot{Q}_n(\theta) - \ddot{Q}_n(\theta_0) = (R_{n1}(\theta) + R_{n1}(\theta)') + R_{n2}(\theta) + R_{n3}(\theta) + R_{n4}(\theta)$$

where

$$\begin{aligned} R_{n1}(\theta) &= \sum_{t=1}^{n} \dot{f}(z_t, \theta_0)\left(\dot{f}(z_t, \theta) - \dot{f}(z_t, \theta_0)\right)' \\ R_{n2}(\theta) &= \sum_{t=1}^{n} \left(\dot{f}(z_t, \theta) - \dot{f}(z_t, \theta_0)\right)\left(\dot{f}(z_t, \theta) - \dot{f}(z_t, \theta_0)\right)' \\ R_{n3}(\theta) &= \sum_{t=1}^{n} \ddot{F}(z_t, \theta)\left(f(z_t, \theta) - f(z_t, \theta_0)\right) \\ R_{n4}(\theta) &= -\sum_{t=1}^{n} \left(\ddot{F}(z_t, \theta) - \ddot{F}(z_t, \theta_0)\right) u_t \end{aligned}$$

and $\ddot{F}(z_t, \theta) = \partial^2 f(z_t, \theta)/\partial\theta\partial\theta'$.

If we define $\ddot{Q}(x_t, \beta) = \partial^2 q(x_t, \beta)/\partial\beta\partial\beta'$ and $\ddot{Q}_i(x_{it}, \beta_i) = \partial^2 q_i(x_{it}, \beta_i)/\partial\beta_i\partial\beta_i'$ similarly as $\ddot{F}$, then we have

$$\ddot{F}(z_t, \theta) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \ddot{Q}(x_t, \beta) \end{pmatrix} \quad \text{and} \quad \ddot{Q}(x_t, \beta) = \begin{pmatrix} \ddot{Q}_1(x_{1t}, \beta_1) & & 0 \\ & \ddots & \\ 0 & & \ddot{Q}_m(x_{mt}, \beta_m) \end{pmatrix}.$$

It follows that

$$D_{n\delta}^{-1} \ddot{F}(z_t, \theta) D_{n\delta}^{-1} = n^{2\delta} \nu_n^{-1} \ddot{Q}(x_t, \beta) \nu_n^{-1}, \qquad (42)$$

and that

$$\left\| \nu_n^{-1} \ddot{Q}(x_t, \beta) \nu_n^{-1} \right\| \leq \sum_{i=1}^{m} \left\| (\nu_{ni} \otimes \nu_{ni})^{-1} \ddot{q}_i(x_{it}, \beta_i) \right\| \qquad (43)$$

$$\left\| \nu_n^{-1} \ddot{Q}(x_t, \beta) \nu_n^{-1} \right\|^2 \leq \sum_{i=1}^{m} \left\| (\nu_{ni} \otimes \nu_{ni})^{-1} \ddot{q}_i(x_{it}, \beta_i) \right\|^2. \qquad (44)$$

Here we use the fact that $\|A\|^2 \le \mathrm{tr}(A'A) \le \|\mathrm{vec}\,A\|^2$ for any square matrix A, and that $\|A\| \le \sum_{i=1}^m \|A_i\|$ for $A = \mathrm{diag}(A_1, \ldots, A_m)$ with any square matrices A_i.

We want to show

$$\sup_{\theta \in \Theta_n} \left\| D_{n\delta}^{-1} R_{nk}(\theta) D_{n\delta}^{-1} \right\| \to_p 0, \tag{45}$$

for $k = 1, \ldots, 4$. To establish (45) for $k = 1$, we note that

$$\left\| D_{n\delta}^{-1} R_{n1}(\theta) D_{n\delta}^{-1} \right\| \le \sum_{t=1}^n \left\| D_{n\delta}^{-1} \dot\ell(z_t;\theta_0) \right\| \left\| D_{n\delta}^{-1} \ddot\ell(z_t;\bar\theta) D_{n\delta}^{-1} \right\|,$$

for all $\theta \in \Theta_n$, where $\bar\theta$ is on the line segment connecting θ and θ_0, and that

$$\left\| D_{n\delta}^{-1} \dot\ell(z_t;\theta_0) \right\| \le n^{\delta} \left(\left\| n^{-1/2} \kappa_{nd}^{-1} d_t \right\| + \left\| n^{-1/2} w_t \right\| + \sum_{i=1}^m \left\| \nu_{ni}^{-1} \dot q_i(x_{it};\beta_i^0) \right\| \right).$$

Therefore, (45) follows immediately from (39), (40) and (41), along with (42) and (43).

To deduce (45) for $k = 2$, observe that for all $\theta \in \Theta_n$

$$\left\| D_{n\delta}^{-1} R_{n2}(\theta) D_{n\delta}^{-1} \right\| \le \sum_{t=1}^n \left\| D_{n\delta}^{-1} \ddot\ell(z_t;\bar\theta) D_{n\delta}^{-1} \right\|^2,$$

where $\bar\theta$ lies, as above, between θ and θ_0. We then have (45) from (38), (42) and (44).

To prove (45) for $k = 3$, we write

$$\left\| D_{n\delta}^{-1} R_{n3}(\theta) D_{n\delta}^{-1} \right\| \le \sum_{t=1}^n \left\| D_{n\delta}^{-1} \ddot\ell(z_t;\theta) D_{n\delta}^{-1} \right\| \left| f(z_t;\theta) - f(z_t;\theta_0) \right|,$$

and use the fact that

$$\left| f(z_t;\theta) - f(z_t;\theta_0) \right| \le n^{\delta} \left(\left\| n^{-1/2} w_t \right\| + \left\| n^{-1/2} \kappa_{nd}^{-1} d_t \right\| + \sum_{i=1}^m \left\| \nu_{ni}^{-1} \dot q_i(x_{it};\beta_i^0) \right\| + \frac{1}{2} \sum_{i=1}^m \left\| (\nu_{ni} - \nu_{ni})^{-1} \ddot q_i(x_{it};\bar\beta_i) \right\| \right),$$

together with (39) – (43). Due to (42), the result (45) for the case $k = 4$ follows immediately from Park and Phillips (1998).

8.4 Proof of Corollary 4.3

This is obvious from the proof of Theorem 4.2, and is omitted. ■

8.5 Proof of Theorem 5.2

Define

$$\varepsilon_{\ell,t} = \varepsilon_t + \sum_{k=\ell+1}^{\infty} \Pi_k v_{t-k}$$
$$v_{\ell,t} = (v'_{t-1}, \ldots, v'_{t-\ell})';$$

and

$$\Pi(\ell) = (\Pi_1, \ldots, \Pi_\ell); \quad \hat{\Pi}(\ell) = (\hat{\Pi}_1, \ldots, \hat{\Pi}_\ell).$$

Then, we have

$$\hat{\varepsilon}_{\ell,t+1} = \varepsilon_{\ell,t+1} - (\hat{\Pi}(\ell) - \Pi(\ell))v_{\ell,t+1}.$$

For the proof of the main results, we first need to show

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} d_{nt} \hat{\varepsilon}'_{\ell,t+1} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} d_{nt} \varepsilon'_{t+1} + o_p(1) \tag{46}$$

$$\frac{1}{\sqrt[4]{n}} \sum_{t=1}^{n} a_i(x_{it}) \hat{\varepsilon}'_{\ell,t+1} = \frac{1}{\sqrt[4]{n}} \sum_{t=1}^{n} a_i(x_{it}) \varepsilon'_{t+1} + o_p(1) \tag{47}$$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} b_{ni}(x_{it}) \hat{\varepsilon}'_{\ell,t+1} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} b_{ni}(x_{it}) \varepsilon'_{t+1} + o_p(1) \tag{48}$$

in the notation defined in Lemma 3.1.

We proceed by proving a set of technical results that are needed for the proof of (46) – (48). Throughout the proof, we assume that Assumptions 2.1 – 2.4 and 5.1 hold.

Lemma A1 Let $\ell = o(\sqrt{n})$. Then $\left\| \hat{\Pi}(\ell) - \Pi(\ell) \right\| = O_p(\ell^{(1-s)/2}) + O_p(n^{-1/2}\ell^{1/2})$.

Proof of Lemma A1 The stated result follows from Berk (1974, Equation 2.17, page 493). ■

Lemma A2 $E\|\varepsilon_{\ell,t} - \varepsilon_t\|^r = O(\ell^{-rs/2})$ for large ℓ.

Proof of Lemma A2 Write

$$\varepsilon_{\ell,t} - \varepsilon_t = \sum_{k=\ell+1}^{\infty} \Pi_k v_{t-k} = \sum_{k=\ell+1}^{\infty} \Gamma_k \varepsilon_{t-k}.$$

We have, as shown in Berk (1974, Proof of Lemma 2),

$$\sum_{k=\ell+1}^{\infty} \|\Gamma_k\|^2 \leq c \sum_{k=\ell+1}^{\infty} \|\Pi_k\|^2;$$

for some constant c. However, for all m,

$$
\begin{aligned}
E\left\| \sum_{k=\ell+1}^{m} \Gamma_k \varepsilon_{t-k} \right\|^r &\le c\,E\left(\sum_{k=\ell+1}^{m} \|\Gamma_k\|^2 \|\varepsilon_{t-k}\|^2 \right)^{r/2} \\
&\le c \left(\sum_{k=\ell+1}^{m} \|\Gamma_k\|^2 \right)^{r/2} E\|\varepsilon_t\|^r \\
&\le c \left(\sum_{k=\ell+1}^{m} |\Gamma_k|^2 \right)^{r/2} E\|\varepsilon_t\|^r;
\end{aligned}
$$

with some constant c, by Burkholder's inequality (Hall and Heyde (1980), page 23) and Minkowski's inequality. The stated result now follows immediately. ■

Lemma A3 (a) $\sum_{t=1} a_i(x_{it}) v'_{t-k} = O_p(n^{(1+r)/2r})$ uniformly for $k = 1, \ldots, \ell$.

(b) $\sum_{t=1} a_i(x_{it})(\varepsilon_{\ell,t+1} - \varepsilon_{t+1})' = O_p(n^{(1+r)/2r}\ell^{-s/2})$.

Proof of Lemma A3 We may assume that a_i is scalar-valued as in the proof of Lemma 3.1. Let $e_t = v_{t-k}$ and $\varepsilon_{\ell;t+1} - \varepsilon_{t+1}$ respectively for parts (a) and (b), and write

$$\sum_{t=1} \|a_i(x_{it}) e_t\| \le A_n + B_n;$$

where

$$A_n = c_n \sum_{t=1} |a_i(x_{it})| \quad \text{and} \quad B_n = \|a_i\| \sum_{t=1} \|e_t\| 1\{\|e_t\| > c_n\};$$

for a sequence c_n of numbers. It follows directly from Lemma 3.1(a) that $A_n = O_p(c_n\sqrt{n})$. Notice for B_n that

$$
\begin{aligned}
E(B_n) &= \|a_i\| \sum_{t=1} E\|e_t\| 1\{\|e_t\| > c_n\} \\
&\le n\|a_i\| \frac{\sup_t E\|e_t\|^r}{c_n^{r-1}};
\end{aligned}
$$

by Tchebyshev's inequality. Part (a) follows immediately if we let $c_n = n^{1/2r}$, since $E\|v_{t-k}\|^r < \infty$ uniformly for $k = 1, \ldots, \ell$. For part (b), we set $c_n = n^{1/2r}\ell^{-s/2}$ and note that $E\|\varepsilon_{\ell;t+1} - \varepsilon_{t+1}\|^r = O(\ell^{-rs/2})$ due to Lemma A2. ■

Lemma A4 (a) $\sum_{t=1} b_{ni}(x_{it})v'_{t-k} = O_p(n^{(4+3r)/4(1+r)}\ell^{r/2(1+r)})$ uniformly for $k = 1,\ldots,\ell$.

(b) $\sum_{t=1} b_{ni}(x_{it})(\varepsilon_{\ell,t+1} - \varepsilon_{t+1})' = O_p(n\ell^{-s/2})$.

Proof of Lemma A4 As in the proof of Lemma A3, we assume that b_i is scalar-valued. It is shown in Park and Phillips (1998) that

$$\sum_{t=1} b_{ni}(x_{it})v_{t-k} = \sum_{t=1} h_i\left(\frac{x_{it}}{\sqrt{n}}\right) v_{t-k}(1 + o_p(1));$$

uniformly in $k = 1;\ldots;\ell$. We now show that

$$\sum_{t=1} h_i\left(\frac{x_{it}}{\sqrt{n}}\right) v_{t-k} = \sum_{t=1} h_i\left(\frac{x_{it}}{\sqrt{n}}\right) v_{t+1} + O_p(n^{(3r+4)/4(r+1)}\ell^{r/2(r+1)}): \qquad (49)$$

To deduce (49), we write

$$\sum_{t=1} h_i\left(\frac{x_{it}}{\sqrt{n}}\right) v_{t+1} = \sum_{t=k+1} h_i\left(\frac{x_{i,t-k-1}}{\sqrt{n}}\right) v_{t-k} + \sum_{t=n-k} h_i\left(\frac{x_{it}}{\sqrt{n}}\right) v_{t+1};$$

with the convention that $h(x_{i0}/\sqrt{n}) = 0$. It follows that

$$\sum_{t=1} h_i\left(\frac{x_{it}}{\sqrt{n}}\right) v_{t+1} - \sum_{t=k+1} h_i\left(\frac{x_{it}}{\sqrt{n}}\right) v_{t-k}$$
$$= \sum_{t=n-k} h_i\left(\frac{x_{it}}{\sqrt{n}}\right) v_{t+1} - \sum_{t=k+1}\left(h_i\left(\frac{x_{it}}{\sqrt{n}}\right) - h_i\left(\frac{x_{i,t-k-1}}{\sqrt{n}}\right)\right) v_{t-k}:$$

Clearly,

$$\sum_{t=n-k}\left\| h_i\left(\frac{x_{it}}{\sqrt{n}}\right) v_{t+1}\right\| \le \|h\|_{K_i} \sum_{t=n-k} \|v_{t+1}\| = O_p(\ell); \qquad (50)$$

uniformly in $k = 1;\ldots;\ell$.

For h differentiable with locally bounded derivative, we have

$$\sum_{t=k+1}\left\|\left(h_i\left(\frac{x_{it}}{\sqrt{n}}\right) - h_i\left(\frac{x_{i,t-k-1}}{\sqrt{n}}\right)\right) v_{t-k}\right\|$$
$$\le \frac{1}{\sqrt{n}}\|\dot{h}_i\|_{K_i} \sum_{t=k+1}\left(\sum_{j=0}\|v_{t-j}\|\right)\|v_{t-k}\| = O_p(n^{1/2}\ell); \qquad (51)$$

uniformly in $k = 1;\ldots;\ell$. Now (49) follows immediately from (50) and (51), since $\ell = O(n^{1/4})$.

For $h_i(x) = 1\{x \ge 0\}$, we start by looking at

$$\frac{1}{n}\left\| \sum_{t=1} \left(h_i\left(\frac{x_{it}}{\sqrt{n}}\right) - h_i\left(\frac{x_{i,t-k-1}}{\sqrt{n}}\right)\right) v_{t-k} \right\| \le \left(\frac{1}{n}\sum_{t=1}\left| h_i\left(\frac{x_{it}}{\sqrt{n}}\right) - h_i\left(\frac{x_{i,t-k-1}}{\sqrt{n}}\right)\right|^2\right)^{1/2} \left(\frac{1}{n}\sum_{t=1}\|v_{t-k}\|^2\right)^{1/2}. \tag{52}$$

Since $\sum_{t=k+1}^{n}\|v_{t-k}\|^2 = O_p(n)$ uniformly in $k = 1,\ldots,\ell$, we may concentrate on the first term in (52). Define

$$R_{nk} = \frac{1}{n}\sum_{t=k+1} 1\{x_{it} \ge 0, x_{i,t-k-1} < 0\} + 1\{x_{it} < 0, x_{i,t-k-1} \ge 0\},$$

and bound it by $S_{nk} + T_{nk}$, where

$$S_{nk} = \sum_{t=k+1} 1\left\{\left|\frac{x_{i,t-k-1}}{\sqrt{n}}\right| < c_n\right\} \quad \text{and} \quad T_{nk} = \sum_{t=1} 1\left\{\left|\frac{\sum_{j=0}^{k} v_{i,t-j}}{\sqrt{n}}\right| \ge c_n\right\}.$$

We have

$$\begin{aligned} S_{nk} &\le \sum_{t=1} 1\left\{\left|\frac{x_{it}}{\sqrt{n}}\right| < c_n\right\} \\ &= (nc_n)\frac{1}{c_n}\int_0^1 1\{|V_{ni}(r)| < c_n\} = O_p(nc_n). \end{aligned}$$

Moreover,

$$ET_{nk} = \sum_{t=1} \Pr\left\{\left|\frac{\sum_{j=0}^{k} v_{i,t-j}}{\sqrt{n}}\right| \ge c_n\right\} \le n^{1-r/2}c_n^{-r}E\left|\sum_{j=0} v_{i,t-j}\right|^r,$$

and by Minkowski's inequality, we have

$$E\left|\sum_{j=0} v_{i,t-j}\right|^r \le \ell^r E|v_{it}|^r,$$

for all $k = 1,\ldots,\ell$. The stated result in part (a) now follows immediately, if we let $c_n = n^{-r/2(r+1)}\ell^{r/(r+1)}$.

To show part (b), we note that

$$\left\|\sum_{t=1} b_{ni}(x_{it})(\varepsilon_{\ell,t+1} - \varepsilon_{t+1})\right\| \le \|b_{ni}\|_{K_i}\sum_{t=1}\|(\varepsilon_{\ell,t+1} - \varepsilon_{t+1})\|,$$

from which we may easily deduce the stated result, since $\|b_{n-}\|_{K_-} = O_p(1)$ and

$$E\|\varepsilon_{\ell,t+1} - \varepsilon_{t+1}\|^2 = O(\ell^{-s});$$

as shown in Lemma A2. ■

We may now show (46) – (48). To prove (46), write

$$\hat\varepsilon_{\ell,t+1} = \varepsilon_{t+1} - (\varepsilon_{\ell,t+1} - \hat\varepsilon_{\ell,t+1}) + (\varepsilon_{\ell,t+1} - \varepsilon_{t+1});$$

and use this to rewrite (46) as

$$\frac{1}{\sqrt{n}}\sum_{t=1} d_{nt}\hat\varepsilon'_{\ell,t+1} = \frac{1}{\sqrt{n}}\sum_{t=1} d_{nt}\varepsilon'_{t+1} - A_n + B_n;$$

where

$$A_n = \frac{1}{\sqrt{n}}\sum_{t=1} d_{nt}v'_{\ell,t+1}(\hat\pi(\ell) - \pi(\ell))'$$

$$B_n = \frac{1}{\sqrt{n}}\sum_{t=1} d_{nt}(\varepsilon_{\ell,t+1} - \varepsilon_{t+1})'.$$

To establish (46), it suffices to show $A_n = B_n = o_p(1)$. It follows directly from Lemma A2 and Assumption 2.4 that

$$B_n \le \frac{1}{\sqrt{n}}\sum_{t}\|d_{nt}\|\sum_{t=1}\|\varepsilon_{\ell,t+1} - \varepsilon_{t+1}\| = O(n^{-1/2})O_p(n\ell^{-s/2}) = o_p(1);$$

for $\delta > 1/s$. To show $A_n = o_p(1)$, we may first show

$$\sum_{t=1} d_{nt}v'_{t-k} = O_p(\sqrt{n}); \quad \text{uniformly in } k = 1,\ldots,\ell. \tag{53}$$

To deduce (53), notice that

$$\sum_{t=1} d_{nt}v'_{t+1} = \sum_{t=k+1} d_{nt}v'_{t-k} + \sum_{t=n-k} d_{nt}v'_{t+1};$$

with the convention $d_{n0} = 0$. Then, it follows that

$$\sum_{t=1} d_{nt}v'_{t+1} - \sum_{t=k+1} d_{nt}v'_{t-k} = \sum_{t=n-k} d_{nt}v'_{t+1} - \sum_{t=k+1}(d_{nt} - d_{nt-k-1})v'_{t-k}. \tag{54}$$

We have

$$\sum_{t=n-k}\|d_{nt}v'_{t+1}\| \le \sup_t\|d_{nt}\|\sum_{t=n-k}\|v_{t+1}\| = O_p(\ell); \tag{55}$$

uniformly in $k = 1, \ldots, \ell$, and

$$\sum_{t=k+1}^{n} \|(d_{nt} - d_{nt-k-1}) v_{t-k}'\| \le \sum_{t=k+1}^{n} \|d_{nt} - d_{nt-k-1}\| \|v_{t-k}\| = n o(\ell/n) O_p(1) = o_p(\ell). \tag{56}$$

Moreover, following Proof of Lemma 3.1 (f), it can be shown that

$$\sum_{t=1}^{n} d_{nt} v_{t+1}' = O_p(\sqrt{n}). \tag{57}$$

Now we may easily deduce (53) from (54)–(57). Then it follows that

$$A_n = O(n^{-1/2}) O_p(\ell n^{1/2}) \left(O_p(\ell^{(1-s)/2}) + O_p(n^{-1/2}\ell^{1/2}) \right) = o_p(1),$$

for $0 < \delta < 1/3$, due to Lemma A1, and therefore $A_n = B_n = o_p(1)$ holds for $1/s < \delta < 1/3$.

For (47), we first show that

$$\frac{1}{\sqrt[4]{n}} \sum_{t=1}^{n} a_i(x_{it}) (\varepsilon_{\ell,t+1} - \varepsilon_{t+1})' \to_p 0 \tag{58}$$

$$\left(\frac{1}{\sqrt[4]{n}} \sum_{t=1}^{n} a_i(x_{it}) v_{\ell,t+1}' \right) \left(\hat\Pi(\ell) - \Pi(\ell) \right) \to_p 0. \tag{59}$$

One may easily deduce (58) from Lemma A3(b), since

$$O(n^{-1/4}) O_p(n^{(1+r)/2r} \ell^{-s/2}) = o_p(1),$$

whenever $\delta > (r+2)/2rs$. The condition holds because we set $\delta > (r+2)/2r(s-3)$. For (59), apply Lemmas A1 and A2(a), and note that

$$O(n^{-1/4}) O_p(n^{(1+r)/2r} \ell) O_p(\ell^{(1-s)/2}) = o_p(1),$$

and

$$O(n^{-1/4}) O_p(n^{(1+r)/2r} \ell) O_p(n^{-1/2} \ell^{1/2}) = o_p(1),$$

respectively if $\delta > (r+2)/2r(s-3)$ and $\delta < (r-2)/6r$, which are satisfied under our condition for δ. Note that $(r-2)/6r > r/(6+8r)$ for all $r > 8$. The stated result in (47) is immediate from (58) and (59).

To prove (48), we first deduce from Lemma A4 that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} b_{ni}(x_{it}) (\varepsilon_{\ell,t+1} - \varepsilon_{t+1})' \to_p 0 \tag{60}$$

$$\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} b_{ni}(x_{it}) v_{\ell,t+1}' \right) \left(\hat\Pi(\ell) - \Pi(\ell) \right) \to_p 0; \tag{61}$$

from which (48) readily follows. To show (60), note that

$$O(n^{-1/2})O_p(n\ell^{-s/2}) = o_p(1);$$

whenever $\delta > 1/s$, which holds because $\delta > 1/2(s-3)$ under our condition. For (61), we note that

$$O(n^{-1/2})O_p(\ell n^{(4+3r)/4(1+r)}\ell^{r/2(1+r)})O_p(\ell^{(1-s)/2}) = o_p(1);$$

whenever $(s-4+1/(1+r))\delta > (r+2)/2(r+1)$. The condition holds if $\delta > (r+2)/2r(s-4)$, which is in turn satisfied under our assumption $\delta > (r+2)/2r(s-3)$. Moreover,

$$O(n^{-1/2})O_p(\ell n^{(4+3r)/4(1+r)}\ell^{r/2(1+r)})O_p(n^{-1/2}\ell^{1/2}) = o_p(1);$$

for $\delta < r/(6+8r)$. ■

Finally, we return to the proof of the main results.

Proof of (a) From (47), we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} q_i(x_{it},\beta_i^0)\hat{\varepsilon}_{\ell,t+1} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} q_i(x_{it},\beta_i^0)\varepsilon_{\ell,t+1} + o_p(1);$$

and this gives

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} q_i(x_{it},\beta_i^0)u_t^* &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} q_i(x_{it},\beta_i^0)\left(u_t - \hat{\sigma}_{u\varepsilon}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\varepsilon}_{\ell,t+1}\right)\\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} q_i(x_{it},\beta_i^0)\left(u_t - \sigma_{u\varepsilon}\Sigma_{\varepsilon\varepsilon}^{-1}\varepsilon_{\ell,t+1}\right) + o_p(1):\end{aligned}$$

Notice that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} \varepsilon_t \to_d \Omega^{-1}(1)V(r); \qquad (62)$$

and $(u_t - \sigma_{u\varepsilon}\Sigma_{\varepsilon\varepsilon}^{-1}\varepsilon_{\ell,t+1})$ is a martingale difference sequence with variance

$$\mathrm{var}(u_t - \sigma_{u\varepsilon}\Sigma_{\varepsilon\varepsilon}^{-1}\varepsilon_{\ell,t+1}) = \sigma_u^2 - \sigma_{u\varepsilon}\Sigma_{\varepsilon\varepsilon}^{-1}\sigma_{\varepsilon u} = \sigma_u^2 - \omega_{uv}\,\omega_{vv}^{-1}\omega_{\varepsilon u};$$

due to (12). Now the stated result follows directly from Lemma 3.1 (c). ■

Proof of (b) It is immediate from (46) and (9) that

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{t=1}^{n} d_{nt}u_t^* &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} d_{nt}\left(u_t - \hat{\sigma}_{u\varepsilon}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\varepsilon}_{\ell,t+1}\right)\\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} d_{nt}\left(u_t - \sigma_{u\varepsilon}\Sigma_{\varepsilon\varepsilon}^{-1}\varepsilon_{\ell,t+1}\right) + o_p(1)\\
&\to_d \int_0^1 d(r)d\left(U(r) - \sigma_{u\varepsilon}\Sigma_{\varepsilon\varepsilon}^{-1}\Omega^{-1}(1)V(r)\right)\\
&= \int_0^1 d(r)dU^*(r), \qquad (63)
\end{aligned}
$$

since

$$U(r) - \sigma_{u\varepsilon}\Sigma_{\varepsilon\varepsilon}^{-1}\Omega^{-1}(1)V(r) = U(r) - \omega_{uv}\omega_{vv}^{-1}V(r) = U^*(r),$$

due to (62) and (12). Moreover, it follows directly from (48), Lemma 3.1(d) and (62) that

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \underline{\kappa}_n^{-1}\underline{q}_H(x_t,\beta_0)u_t^* &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \underline{\kappa}_n^{-1}\underline{q}_H(x_t,\beta_0)\left(u_t - \sigma_{u\varepsilon}\Sigma_{\varepsilon\varepsilon}^{-1}\varepsilon_{\ell,t+1}\right) + o_p(1)\\
&\to_d \int_0^1 \underline{h}(V(r),\beta_0)dU^*(r). \qquad (64)
\end{aligned}
$$

Now the stated result follows immediately from (63) and (64). ■

9. 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.

Berk, K.N. (1974). "Consistent autoregressive spectral estimates," Annals of Statistics 2: 489-502.

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

Hansen, B.E. (1992). "Convergence to stochastic integrals for dependent and heterogeneous processes," Econometric Theory 8: 489-500.

Kasahara, Y. and S. Kotani (1979). "On limit processes for a class of additive functionals of recurrent diffusion processes," Z. Wahrscheinlichkeitstheorie verw. Gebiete 49: 133-153.

McLeish, D.L. (1974). \Dependent central limit theorems and invariance principles," Annals of Probability 2: 620-628.2: 620-628.

Park, J. Y. (1992). \Canonical cointegrating regressions," Econometrica 60: 119-143.

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 (1989). \Statistical inference in regressions with integrated processes: Part 2," Econometric Theory 5: 95-131.

Park, J.Y. and P.C.B. Phillips (1998). \Nonlinear regressions with integrated time series," Yale University, mimeographed.

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

Park, J.Y. and P.C.B. Phillips (2000). \Nonstationary binary choice," Econometrica, forthcoming.

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 B.E. Hansen (1990). \Statistical inference in instrumental variables regressions with I(1) processes," Review of Economic Studies 57: 99-125.

Phillips, P. C. B. and M. Loretan (1991). \Estimating long-run economic equilibria," Review of Economic Studies 59, 407{436.

Phillips, P.C.B. and V. Solo (1992). \Asymptotics for linear processes," Annals of Statistics 20: 971-1001.

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

Saikkonen, P. (1991). \Asymptotically e±cient estimation of cointegration regressions," Econometric Theory 7, 1{21.

Stock, J. H. and M. W. Watson (1993). \A simple estimator of cointegrating vectors in higher order integrated systems," Econometrica 61, 783{820.

Wooldridge, J.M. (1994). \Estimation and inference for dependent processes," in Handbook of Econometrics Vol. IV, R. Engle and D.L. McFadden (eds.): 2639-2738, North-Holland, Amsterdam.

Figure 1: Densities of NLS and EN-NLS Estimators, n = 250

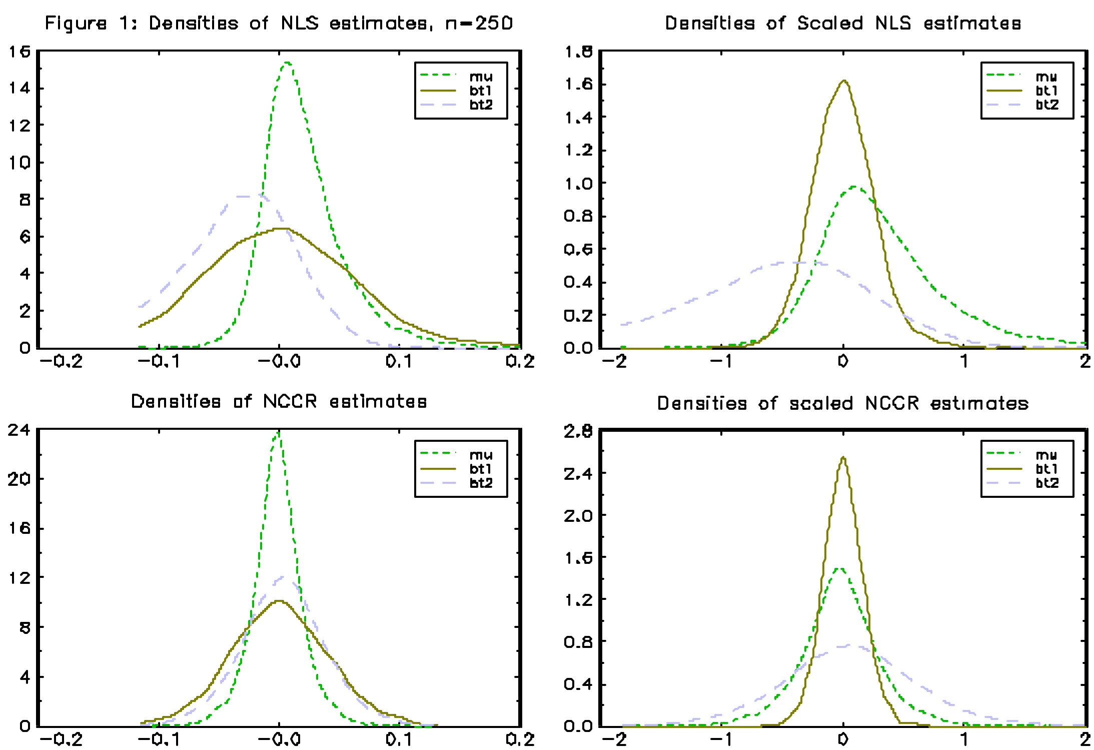

Figure 2: Densities of NLS and EN-NLS Estimators, n = 500

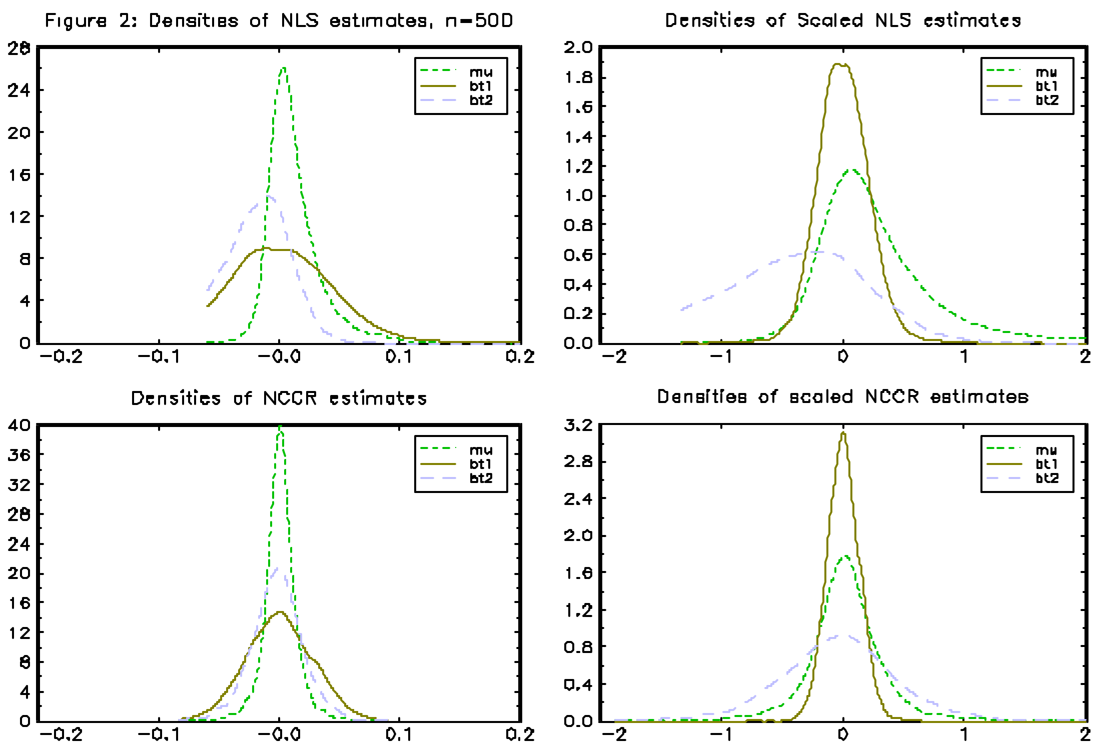

Figure 3: Densities of t-Statistics, n = 250 and 500