Endogeneity in Nonlinear Regressions with Integrated Time Series

Yoosoon Chang and Joon Y. Park

Department of Economics
Rice University

Abstract

This paper considers the nonlinear regression with integrated regressors that are contemporaneously correlated with the regression error. We, in particular, establish the consistency and derive the limiting distribution of the nonlinear least squares estimator under such endogeneity for the regressions with the integrable or asymptotically homogeneous regression function. For the regressions with both the integrable and asymptotically homogeneous regression functions, it is shown that the estimator is consistent and has the same rate of convergence as for the case of the regressions with no endogeneity. Whether or not the limiting distribution is affected by the presence of endogeneity, however, depends upon the type of the regression function. If the regression function is asymptotically homogeneous, the limiting distribution of the least squares estimator has an additional term reflecting the presence of endogeneity. On the other hand, the endogeneity does not have any effect on the least squares limit theory, if the regression function is integrable. Regardless of the presence of endogeneity, the least squares estimator has the same limiting distribution in this case. To illustrate our theory, we consider the nonlinear regressions with logistic and power regression functions with an integrated regressors that have contemporaneous correlations with the regression error.

First Draft: June 18, 2000
This version: November 30, 2005

JEL Classification: C13, C22.

Key words and phrases: nonlinear regression, integrated time series, endogeneity, consistency, limit distributions.

1The corresponding author: Department of Economics - MS 22, Rice University, 6100 Main Street, Houston, TX 77005-1892, Tel: 713-348-2796, Fax: 713-348-5278, Email: yoosoon@rice.edu.
1. Introduction

To be completed.

2. The Model and Assumptions

We consider the model given by

\[ y_t = f(x_t, \theta_0) + \varepsilon_t \]  

where \((x_t)\) is a univariate integrated regressor and given by

\[ x_t = x_{t-1} + v_t \]  

and \((\varepsilon_t)\) is the regression error that we further specify as

\[ \varepsilon_t = \sqrt{1 - \rho^2} u_t + \rho v_t \]  

for some constant \(\rho\) such that \(|\rho| \leq 1\).

We assume that the function \(f : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}\) is known, and \(\theta_0\) is an \(m\)-dimensional true parameter vector that lies inside the parameter set \(\Theta\). Moreover, it is assumed throughout the paper, unless stated otherwise, that

**Assumption 2.1** Let \((u_t)\) be independent of \((v_t)\), and independent and identically distributed with mean zero and finite variance \(\sigma_u^2\).

**Assumption 2.2** Let \((v_t)\) be independent and identically distributed with mean zero and variance \(\sigma_v^2\), and have finite \(\nu\)-th moment.

Some of our subsequent results require additional conditions on the distribution of \((v_t)\), which is given by

**Assumption 2.3** We assume that the characteristic function \(\varphi\) of \((v_t)\) satisfies either (a) \(|\varphi(s)| = 1\) if and only if \(s\) is a multiple of \(2\pi\) or (b) \(\int_{-\infty}^{\infty} |\varphi(s)|^2 ds < \infty\), depending upon whether the distribution of \((v_t)\) is of discrete type or of continuous type.

In the subsequent development of our theory, we denote by \(\mu\) the Lebesgue or counting measure, depending upon whether the distribution of \((v_t)\) is of continuous or of discrete type, and the probability density of \((v_t)\) with respect to \(\mu\) is signified by \(p\). For the expositional simplicity, we assume \(\sigma_u^2 = \sigma_v^2 = 1\) for the rest of the paper. Their variances only have the trivial scaling effect, which is unimportant for analyzing the effect of the presence of endogeneity. Under our construction, the variance of \((\varepsilon_t)\) also becomes unity regardless of the value of \(\rho\).

The nonlinear regression model in (1) is considered earlier by Park and Phillips (2001). They, however, only consider the models with exogeneous errors. This amounts to assuming
$\rho = 0$ in our specification. The lack of contemporaneous correlation between the regressor and the regression error is crucial for the derivation of their results. On the contrary, the regressor and the error are contemporaneously correlated in our model, except for the case of $\rho = 0$. In our model, $\rho$ measures the contemporaneous correlation coefficient between the regressor and the regression error. The value of $\rho$ signifies the fraction of the endogenous component of the regression error, and hence, we might say that the degree of endogeneity increases as $|\rho| \to 1$. The endogeneity becomes maximal when $|\rho| = 1$.

Our assumptions on the innovation $(v_t)$ of the regressor are rather stringent, and in particular, much stronger than those used in Park and Phillips (2001). Such stringent assumptions are introduced here to highlight and fully analyze the effects of the presence of endogeneity. Our assumption of independent and identical distribution for the innovation $(v_t)$ of the regressor is indeed unnecessarily strong for many of the subsequent results. Some of them hold under a weaker set of conditions, which allow in particular for the presence of serial correlation. Others require suitable modifications, but at least qualitatively, we may expect similar results to hold for more general $(v_t)$. These will be pointed out along the way, as we develop our theory. We may allow $(u_t)$ to be a more general martingale difference sequence with respect to a filtration $(\mathcal{F}_t)$, say, as long as $(x_t)$ is adapted to $(\mathcal{F}_{t-1})$.

Though we assume that $(v_t)$ is given by a sequence of independent and identically distributed random variables, the underlying distribution is allowed to be of discrete type, as well as of continuous type. Our assumptions on the underlying distribution are rather mild and hold for a wide class of distributions. The asymptotic theory developed in Park and Phillips (1999, 2001) and others on the asymptotics for the integrable transformations of integrated time series all require the underlying distribution to be of continuous type. This is because all the previous works rely on Akonom (1993), while our theory is built upon Borodin (1986). The former gives some basic asymptotic theories for the integrated processes driven by the linear processes having innovations with continuous distributions, but in contrast the latter provides more comprehensive asymptotics for the simple random walks generated by independent and identically distributed innovations with both discrete and continuous distributions.

Here we consider a simple regression, which has only a single regressor. This is solely for expositional purpose. Though simple, the model we introduced in (1) - (3) has all essential features that we need to analyze the effect of the presence of endogeneity in nonlinear regressions with integrated time series. Therefore, it serves our purpose very well. Our subsequent results are applicable not only for the model considered here, but also for a wide class of more general nonlinear regression models with various additional regressors, stationary as well as integrated, deterministic as well as stochastic. The required extension is indeed rather straightforward and can easily be done following the earlier work by Chang, Park and Phillips (2001). The details of the results for more general models will be given as we develop the theory for our simple model.

In the paper, we consider the estimation of (1) by nonlinear least squares (NLS). If we let

$$Q_n(\theta) = \sum_{t=1}^{n} (y_t - f(x_t, \theta))^2$$
then the NLS estimator \( \hat{\theta}_n \) of \( \theta \) is defined as the minimizer of \( Q_n(\theta) \) over \( \theta \in \Theta \), i.e.,

\[
\hat{\theta}_n = \arg \min_{\theta \in \Theta} Q_n(\theta)
\]

(4)

An error variance estimate is given by

\[
\hat{\sigma}_n^2 = \frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon}_t^2
\]

(5)

where \( \hat{\varepsilon}_t = y_t - f(x_t, \hat{\theta}_n) \).

Under Assumptions 2.1 and 2.2 with \( \nu = 2 \), an invariance principle holds jointly for \( (u_t) \) and \( (v_t) \). In particular, if we let \( U_n(r) = n^{-1/2} \sum_{t=1}^{[nr]} u_t \) and \( V_n(r) = n^{-1/2} \sum_{t=1}^{[nr]} v_t \), then we have

\[
(U_n, V_n) \rightarrow_d (U, V)
\]

(6)

where \( U \) and \( V \) are independent standard Brownian motions. These are the Brownian motions that we use to represent the asymptotics for \( \hat{\theta}_n \). Using the so-called Skorohod representation theorem, we may redefine \( (U_n, V_n) \) up to the distributional equivalence so that \( (U_n, V_n) \) and \( (U, V) \) are defined on a common probability space, and \( (U_n, V_n) \rightarrow_a.s. (U, V) \). This is well known. The stochastic process \( (U_n, V_n) \) takes values in \( D[0,1]^2 \), where \( D[0,1] \) is the space of cadlag functions defined on the unit interval \([0,1]\). For our purpose, it is convenient to endow \( D[0,1] \) with the uniform metric.

Our limit theory also involves the local time \( L \) of the Brownian motion \( V \), which is given by

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

Roughly, \( L(t, x) \) measures the rate of time spent by \( V \) in a neighborhood of \( x \), up to time \( t \). In our subsequent theory, in particular, frequently appears \( L(1, 0) \), i.e., the rate of time spent by \( V \) in an immediate vicinity of the origin in the unit time interval. The reader is referred for the details of the Brownian local time \( L \) to Park and Phillips (1999) and the references cited there.

3. Function Classes and Preliminary Asymptotics

In this section, we introduce the function classes and their asymptotics that are needed for the development of our theory on nonlinear regressions with integrated time series.

3.1 Regular and Strongly Regular Functions

First we consider the asymptotics for normalized integrated time series, and introduce the required regularity conditions. Obviously, the presence of endogeneity only affects the covariance asymptotics, which we will mostly look at here. The same results as in Park and Phillips (2001, PP henceforth) apply for the mean asymptotics.
**Definition 3.1** A transformation $T$ on $\mathbb{R}$ is said to be *regular* if
(a) it is locally bounded and Riemann-integrable, and
(b) it is differentiable on $\mathbb{R}\setminus\{0\}$ and its derivative on $\mathbb{R}\setminus\{0\}$ is (i) bounded by $c|x|^a$ for all $x \in \mathbb{R}\setminus\{0\}$ in a neighborhood of the origin for some constants $a > -1$ and $c > 0$, and (ii) continuously differentiable for all $x \in \mathbb{R}\setminus\{0\}$ with derivative bounded by $c|x|^b$ for some constants $b$ and $c > 0$ in a neighborhood of the origin.

If, in addition to (a) and (b),
(c) it either has vanishing derivative on $\mathbb{R}\setminus\{0\}$ or is continuous at the origin, then it is said to be *strongly regular*.

**Lemma 3.2** Suppose that Assumption 2.2 holds with $\nu = 4$, and let $T$ be a transformation on $\mathbb{R}$.
(a) If $T$ is regular, then
\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} T\left(\frac{x_t}{\sqrt{n}}\right) v_t = O_p(1)
\]
for all large $n$.

(b) Let $T$ be strongly regular. If $T$ has vanishing derivative on $\mathbb{R}\setminus\{0\}$ and has values $a$ and $b$ respectively on the positive and negative parts of $\mathbb{R}$, then
\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} T\left(\frac{x_t}{\sqrt{n}}\right) v_t \rightarrow_d \int_0^1 T(V(r)) dV(r) + cKL(1,0)
\]
where $c = b - a$ and $K$ is a constant given by
\[
K = \int_{-\infty}^{\infty} \int_{-\infty}^{y} xp(x)\mu(dx)\mu(dy)
\]
as $n \to \infty$. If $T$ is continuous at the origin and has nonvanishing derivative $\nabla T$ on $\mathbb{R}\setminus\{0\}$, then
\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} T\left(\frac{x_t}{\sqrt{n}}\right) v_t \rightarrow_d \int_0^1 T(V(r)) dV(r) + \int_0^1 \nabla T(V(r)) dr
\]
as $n \to \infty$.

The motivations for our definitions of regular and strongly regular functions are now clear. For all regular functions, the cross product sum of the properly normalized and transformed integrated process with its contemporaneous innovation is shown to be of order $O_p(n^{1/2})$. The regularity conditions are therefore sufficient to ensure that the required normalization factor is $\sqrt{n}$ for the covariance asymptotics under endogeneity. With the strong regularity conditions, we may obtain the explicit covariance asymptotics in the presence of endogeneity. The regularity conditions required for the regular and strongly regular functions may seem stringent. However, they are satisfied by virtually all functions that are used in practical nonlinear analyses. They hold, for instance, for all the examples considered in PP.
It can be easily deduced, e.g. as in PP, that
\[ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} T\left(\frac{x_t}{\sqrt{n}}\right)u_t \to_d \int_{0}^{1} T(V(r)) \, dU(r) \] (7)
as \( n \to \infty \),\(^2\) which may be regarded as the covariance asymptotics without endogeneity.

Comparing (7) with our results in Lemma 3.2, we may readily see that our regularity conditions are sufficient to guarantee that the same \( \sqrt{n} \)-rate is applicable for the covariance asymptotics even in the presence of endogeneity. If some additional conditions are met and the strong regularity holds, the covariance asymptotics comparable to (7) are derived. As shown in Lemma 3.2, the presence of endogeneity affects the covariance asymptotics, producing an additional term reflecting its presence. Note that for the case of \( \nabla T = 0 \) the additional bias term does not vanish unless \( a = b \), i.e., the function \( T \) is essentially constant.

The Gaussian case for which \( \mu(dx) = (1/\sqrt{2\pi}) \exp(-x^2/2) \, dx \) yields \( K = 1 \).

In the covariance asymptotics for the strongly regular functions, it is interesting to note that the stochastic processes
\[ L(r,0) \text{ and } \int_{0}^{r} \nabla T(V(s)) \, ds \]
are additive functionals whose sample paths are of bounded variation a.s., while
\[ \int_{0}^{r} T(V(s)) \, dV(s) \]
is a continuous martingale. The limit processes are therefore given by semi-martingales for the covariance asymptotics under endogeneity. This is in contrast to the exogenous case, where the limit process is represented solely by a continuous martingale. The presence of the bounded variation components, of course, implies that their limit distributions are biased. The presence of endogeneity thus introduces bias in the limit distributions of the covariance asymptotics.

So far, we have assumed that \( T \) is real-valued. We may, however, easily extend our definitions and results to the vector-valued function \( T \). From now on, we will say that a vector-valued function \( T \) is regular (strongly regular) if all its component real-valued functions are regular (strongly regular) in the sense of Definition 3.1. For such vector-valued functions, the covariance asymptotics obtained in Lemma 3.2 continue to hold. Of course, in this case, the integrals \( \int_{0}^{1} T(V(r)) \, dV(r) \) and \( \int_{0}^{1} \nabla T(V(r)) \, dr \) are understood to be vector-valued, and \( c, a \) and \( b \) to be vectors. It should however be emphasized that we continue to assume there is a single integrated time series \( (x_t) \) and \( V \) is scalar.

\(^2\)PP establishes the result under weaker martingale difference assumption on \( (u_t) \). It also uses a slightly different set of regularity conditions. The regularity concept in PP is equivalent to the local boundedness and local Riemann-integrability here. For their equivalence, see the proof of Theorem 1.1, pp. 81-82, in Ibragimov and Borodin (1995).
3.2 I-Regular and H-Regular Functions

We now consider the family of functions defined on a parameterset, say, Π. Let F be a vector-valued function defined on R × Π. We introduce two classes of such families: I- and H-regular functions. As in PP, I-regular function is a family of integrable functions satisfying certain regularity conditions, and H-regular function is a family of asymptotically homogeneous functions with required regularity conditions. The asymptotically homogeneous functions, introduced first by Park and Phillips (1999), are the functions that behave as homogeneous functions asymptotically. The parameter set Π may be a singleton set, in which case I- and H-regularities become the characteristics of functions rather than families of functions.

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

(a) for all π₀ ∈ Π, there exists a neighborhood N₀ of π₀ such that ∥F(x, π) − F(x, π₀)∥ ≤ ∥π − π₀∥T(x) for all π ∈ N₀, where T : R → R satisfies |T(x)| < c/(1 + |x|₁⁺ε) for some c > 0 and ε > 0 and ∫|T(x)|μ(dx) < ∞, and

(b) for all π ∈ Π, F(·, π) is bounded and ∫|x|₁₂⁺ε∥F(x, π)∥μ(dx) < ∞ for some ε > 0.

**Definition 3.4** Let

F(λx, π) = κ(λ, π)H(x, π) + R(x, λ, π)

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

(a) H(·, π) is strongly regular for each π ∈ Π, and for all x ∈ R, F(x, ·) is equicontinuous in a neighborhood of x, and

(b) ∥R(x, λ, π)∥ ≤ ω(λ, π)Q(x), where ω(λ, π) is such that (κ⁻¹ω)(λ, π) → 0 uniformly in π ∈ Π as λ → ∞ and Q is regular.

We call κ the asymptotic order and H the limit homogeneous function of F. If κ does not depend upon π, then F is said to be H₀-regular.

Definitions 3.3 and 3.4 are comparable to Definitions 3.3 and 3.5, respectively, in PP. The required conditions are, however, somewhat different. Here we modified the conditions in PP so that we may accommodate the integrated processes driven by innovations having discrete distributions and effectively deal with endogeneity. Our conditions for the I-regularity in Definition 3.3 are similar to those used in Definition 3.3 of PP. The differences are mainly due to the necessary modifications that we need in order to use the results of Borodin (1986), in place of Akonom (1993). The conditions for the H-regularity in Definiton 3.4 are more distinctive from those in Definition 3.5 of PP. In this paper, we impose the strong regularity on the limit homogeneous function to derive the limit distributions under endogeneity. Moreover, here we do not allow for the presence of integral functions in the residual term. This is just to simplify the exposition of our theory. Any integral component included in the asymptotically homogeneous function becomes negligible in the limit, and hence, it can be ignored in our asymptotic analyses.

Though our regularity conditions are different from those used in PP, we may easily deduce the mean and covariance asymptotics in PP for I- and H-regular functions. If F is
I-regular on a compact set $\Pi$, then
\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} F(x_t, \pi) \rightarrow_p L(1, 0) \int_{-\infty}^{\infty} F(x, \pi) \mu(dx)
\] (8)
uniformly in $\pi \in \Pi$. Moreover, if $F(\cdot, \pi)$ is I-regular,
\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} F(x_t, \pi) u_t \rightarrow_d \mathbf{MN} \left( 0, L(1, 0) \int_{-\infty}^{\infty} F(x, \pi) F(x, \pi)' \mu(dx) \right)
\] (9)
The mean and covariance asymptotics in (8) and (9) require that Assumptions 2.1, 2.2 with $\nu = 3$ and 2.3 hold.

Similarly, if $F$ is H-regular on a compact set $\Pi$, then
\[
\frac{1}{n} \kappa(\sqrt{n}, \pi)^{-1} \sum_{t=1}^{n} F(x_t, \pi) \rightarrow_a.s. \int_{0}^{1} H(V(r), \pi) dr
\] (10)
uniformly in $\pi \in \Pi$. Moreover, if $F(\cdot, \pi)$ be H-regular, then we have
\[
\frac{1}{\sqrt{n}} \kappa(\sqrt{n}, \pi)^{-1} \sum_{t=1}^{n} F(x_t, \pi) u_t \rightarrow_d \int_{0}^{1} H(V(r), \pi) dU(r)
\] (11)
The mean and covariance asymptotics in (10) and (11) hold under Assumptions 2.1 and 2.2 with $\nu = 2$.

The mean and covariance asymptotics in (8) – (11) hold under much more general conditions than we use in this paper. The mean and covariance asymptotics in (8) and (9) can be deduced for $(x_t)$ driven by general linear processes as long as the distributions of the innovations satisfy a mild set of regularity conditions, and $(u_t)$ is a martingale difference sequence with respect to the filtration $(\mathcal{F}_t)$ such that $(x_t)$ is adapted to $(\mathcal{F}_{t-1})$. Also, we may derive the mean and covariance asymptotics in (10) and (11) under the minimal condition to ensure the weak convergence in (6) and the martingale difference assumption on $(u_t)$ given above. The reader is referred to PP for the details.

We now present the covariance asymptotics for I- and H-regular functions under endogeneity.

**Lemma 3.5** Let Assumptions 2.1, 2.2 with $\nu = 3$ and 2.3 hold. If $F(\cdot, \pi)$ is I-regular,
\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} F(x_t, \pi) u_t \rightarrow_d \mathbf{MN} \left( 0, L(1, 0) \int_{-\infty}^{\infty} F(x, \pi) F(x, \pi)' \mu(dx) \right)
\]
as $n \rightarrow \infty$, independently of (9).
Lemma 3.6  Let Assumptions 2.1 and 2.2 with \( \nu = 4 \) hold. If \( F(\cdot, \pi) \) is H-regular, then we have as \( n \to \infty \)

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} F(x_t, \pi) v_t \to_d \int_{0}^{1} H(V(r), \pi) dV(r) + cKL(1,0)
\]

\[
\frac{1}{\sqrt{n}} \kappa(\sqrt{n}, \pi)^{-1} \sum_{t=1}^{n} F(x_t, \pi) v_t \to_d \int_{0}^{1} H(V(r), \pi) dV(r) + \int_{0}^{1} \nabla H(V(r), \pi) dr
\]

in the notation introduced in Lemma 3.2, depending upon the limit homogeneous function \( H(\cdot, \pi) \) has vanishing or nonvanishing derivative on \( \mathbb{R}\setminus\{0\} \).

4. Nonlinear Least Squares Asymptotics under Endogeneity

We now consider the asymptotics for the NLS estimator \( \hat{\theta}_n \) of \( \theta \) defined in (4) in the presence of the cross contemporaneous correlation between the regressor and the regression error. To present our asymptotics, we denote by \( \dot{f} \) and \( \ddot{f} \) the first and the second derivatives of the regression function \( f \) with respect to the parameter \( \theta \). We assume that they are all vectorized and arranged by lexicographic ordering of their indices. For H-regular \( f \), we denote by \( \dot{\kappa} \) and \( \ddot{\kappa} \) the asymptotic orders of \( \dot{f} \) and \( \ddot{f} \), respectively, and signify by \( \dot{h} \) the limit homogeneous function of \( \dot{f} \).

For the regressions with integrable regression functions, we have

**Theorem 4.1**  Let Assumptions 2.1, 2.2 with \( \nu = 3 \), and 2.3 hold. Assume

(a) \( f, \dot{f} \) and \( \ddot{f} \) are I-regular on \( \Theta \),
(b) \( \int_{-\infty}^{\infty} (f(x, \theta) - f(x, \theta_0))^2 \mu(dx) > 0 \) for all \( \theta \neq \theta_0 \), and
(c) \( \int_{-\infty}^{\infty} \dot{f}(x, \theta_0) \dot{f}(x, \theta_0)^\prime \mu(dx) > 0 \).

Then we have

\[
\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d MN \left( 0, \left[ L(1,0) \int_{-\infty}^{\infty} \dot{f}(x, \theta_0) \dot{f}(x, \theta_0)^\prime \mu(dx) \right]^{-1} \right)
\]

and \( \hat{\sigma}_n^2 \to_p \sigma^2 \), as \( n \to \infty \).

For the regressions with I-regular regression functions, Theorem 4.1 shows that the NLS estimator \( \hat{\theta} \) is consistent with the convergence rate \( \sqrt{n} \), and has a normal mixture limiting distribution, even under the presence of endogeneity. The consistency of \( \hat{\sigma}_n^2 \) is also established. The required regularity conditions are mild. The reader is referred to PP for a detailed discussions on them with some concrete examples.

It should be noted that the asymptotics given in Theorem 4.1 are indeed exactly the same as those obtained by PP for the regressions without endogeneity. The limiting distributions of \( \hat{\theta}_n \) are independent of \( \rho \), the parameter which measures the contemporaneous correlation between the regressor and the regression error. For all the values of \( \rho \), \( \hat{\theta}_n \) converges to its
true value $\theta_0$ at the same rate. Moreover, if properly standardized, $\hat{\theta}_n$ converges to a well defined limiting distribution, which does not depend upon the value of $\rho$. Not surprisingly, endogeneity also has no effect on the consistency of $\hat{\sigma}_n^2$. We may therefore see that the presence of endogeneity does not play any role, at least for the first order asymptotics, on the asymptotic behaviors of the NLS estimators if the regression function is integrable.

For the regressions with homogeneous regression functions, we have

**Theorem 4.2** Let Assumptions 2.1, and 2.2 with $\nu = 4$ hold. Assume

(a) $f, \hat{f}$ and $\bar{f}$ are $H_0$-regular on $\Theta$,
(b) $\kappa > 0$ at infinity, $\| (\hat{\kappa} \otimes \bar{\kappa})^{-1} \kappa \bar{\kappa} \| < \infty$,
(c) $\int_{|x| \leq \delta} (h(x, \theta) - h(x, \theta_0))^2 dx > 0$ for all $\theta \neq \theta_0$ and $\delta > 0$, and
(d) $\int_{|x| \leq \delta} \dot{h}(x, \theta_0)\bar{h}(x, \theta_0)' dx > 0$ for all $\delta > 0$.

Then we have

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow_d \left( \int_0^1 \dot{h}(V, \theta_0)\bar{h}(V, \theta_0)' \right)^{-1} \left( \int_0^1 \dot{h}(V, \theta_0) dW(\rho) + \rho cKL(1, 0) \right)$$

or

$$\sqrt{n}\kappa(\sqrt{n})'(\hat{\theta}_n - \theta_0) \rightarrow_d \left( \int_0^1 \dot{h}(V, \theta_0)\bar{h}(V, \theta_0)' \right)^{-1} \left( \int_0^1 \dot{h}(V, \theta_0) dW(\rho) + \rho \int_0^1 \nabla \dot{h}(V, \theta_0) \right)$$

where

$$W(\rho) = \sqrt{1 - \rho^2}U + \rho V$$

depending upon whether $\dot{h}(\cdot, \theta_0)$ has vanishing or nonvanishing derivative on $R \setminus \{0\}$, and $\hat{\sigma}_n^2 \rightarrow_p \sigma^2$, as $n \rightarrow \infty$.

**Theorem 4.3** Let Assumptions 2.1, and 2.2 with $\nu = 4$ hold. Assume

(a) $\hat{f}$ is $H$-regular on $\Theta$,
(b) for any $M > 0$ given, there exists $\epsilon > 0$ and $\delta$-neighborhood $N$ of $\theta_0$ such that

$$\lambda^{-1+\epsilon} \| \kappa(\lambda, \theta_0)^{-1} \| \rightarrow 0, \quad \lambda^\epsilon \left\| (\bar{\kappa} \otimes \bar{\kappa})(\lambda, \theta_0)^{-1} \left( \sup_{|x| \leq M} \sup_{\theta \in N} |\bar{f}(\lambda x, \theta)| \right) \right\| \rightarrow 0$$

as $\lambda \rightarrow \infty$, and
(c) $\int_{|x| \leq \delta} h(x, \theta_0)\bar{h}(x, \theta_0)' dx > 0$ for all $\delta > 0$.

Then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \quad \text{or} \quad \sqrt{n}\kappa(\sqrt{n})'(\hat{\theta}_n - \theta_0)$$

has the same limiting distributions as in Theorem 4.2, depending upon whether $\dot{h}(\cdot, \theta_0)$ has vanishing or nonvanishing derivative on $R \setminus \{0\}$, and $\hat{\sigma}_n^2 \rightarrow_p \sigma^2$, as $n \rightarrow \infty$.

---

3Of course, the equivalence we establish here is asymptotic. The endogeneity might still have some significant effects in finite samples. This will be investigated through simulations in a later section.
Theorems 4.2 and 4.3 provide the asymptotics for the NLS estimators \( \hat{\theta}_n \) and \( \hat{\sigma}^2_n \) for the regressions with H-regular regression functions. They both yield the same asymptotics. For the regressions with \( H_0 \)-regular regression functions, the conditions in Theorem 4.2 are required for the derived asymptotics to hold. If the regressions have more general H-regular regression functions, the conditions in Theorem 4.3 need to be satisfied. For a detailed discussions on those conditions and examples, see PP. In short, virtually all the nonlinear regression models used in practical applications satisfy the regularity conditions in Theorems 4.2 or 4.3.

Unlike the regressions with I-regular regression functions, the presence of endogeneity affects the asymptotics here. The NLS estimator \( \hat{\theta}_n \) is consistent and the convergence rate remains the same, regardless of the value of the endogeneity parameter \( \rho \). We also have the consistency of \( \hat{\sigma}^2_n \) despite of the contemporaneous correlation between the regressor and the regression error. The asymptotic distribution of \( \hat{\theta}_n \), however, is given as a function of \( \rho \). An additional term, which reflects the presence of endogeneity, appears in the limit distribution of \( \hat{\theta}_n \) as a result of endogeneity. The added term, as we discussed earlier in the previous section, contributes more towards the bias in its asymptotic distribution.

It should be pointed out that the consistency of the NLS estimator \( \hat{\theta}_n \) here does not necessarily require an accelerated convergence rate. As is well known, the least squares estimator is consistent under endogeneity in linear cointegrating regression model. This, however, has been understood as being due to the super-consistency of the least squares estimator, i.e., the convergence rate being \( n \), an order of magnitude greater than the usual \( \sqrt{n} \)-rate. Our results here make it clear that the consistency still holds under endogeneity when this is not the case. For the regressions with integrated time series, the consistency continues to hold under endogeneity for the regressions with I-regular regression functions or H-regular regression functions with vanishing or non-vanishing derivatives, for which the convergence rate is \( \sqrt{n} \) or even slower and reduced to \( \sqrt{\mathbb{N}} \). The robustness of the consistency in the regressions with integrated time series is not due to the accelerated convergence rate, but to the magnitude of the signal provided by the presence of the stochastic trend in the integrated regressor.

Nonlinear regressions with multiple regressors in an additive regression function are considered in Chang, Park and Phillips (2001). They consider nonlinear functions of stationary regressors and deterministic trends, as well as integrated regressors, and study how individual components interact in the limit. For instance, they show that in the regression

\[
y_t = \sum_{i \in \mathbb{I}} a_i(x_{it}, \theta_i) + \sum_{i \in \mathbb{H}} b_i(x_{it}, \theta_i) + u_t
\]

the I-regular terms in the regression function, i.e., \( \sum_{i \in \mathbb{I}} a_i(x_{it}, \theta_i) \), are asymptotically orthogonal to the H-regular terms, i.e., \( \sum_{i \in \mathbb{H}} b_i(x_{it}, \theta_i) \), and the individual components \( a_i(x_{it}, \theta_i) \) in the I-regular component are orthogonal each other. This implies that the NLS estimators \( \hat{\theta}_{in} \) of \( \theta_i \), for \( i \in \mathbb{I} \), obtained from running the regression (12) are asymptotically equivalent to the NLS estimators \( \tilde{\theta}_{in} \) of \( \theta_i \) obtained from running \( y_t = a_i(x_{it}, \theta_i) + u_t \) separately for each \( i \in \mathbb{I} \). The asymptotic equivalence between \( \hat{\theta}_{in} \) and \( \tilde{\theta}_{in} \) for the I-regular components provides a useful implication.
For I-regular regression functions, the individual components \( \dot{f}_i \)'s of \( \dot{f} \) are asymptotically orthogonal as shown in Chang, Park and Phillips (2001). This in particular implies that the \((m \times m)\) matrix \( \int_{-\infty}^{\infty} \dot{f}(x, \theta_0) \dot{f}(x, \theta_0)' \mu(dx) \) appearing in the definition of the limit mixing variate of the mixed normal distribution given in Theorem 4.1 is diagonal. The limit distribution given in Theorem 4.1 therefore reduces to

\[
\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d MN \left( 0, \left[ L(1, 0) \int_{-\infty}^{\infty} \dot{f}_i(x, \theta_0)^2 \mu(dx) \right]^{-1} \right) \tag{13}
\]

for each individual component \( i = 1, \ldots, m \).

5. Specific Examples

In this section, we consider some specific nonlinear regression models and obtain the explicit asymptotics as illustrations. We consider two regression functions that seem to be of some special interests in econometric applications. One is the power function given by

\[
f(x, \alpha, \beta) = \alpha x^\beta 1\{x > 0\} \tag{14}
\]

and the other is the logistic function specified as

\[
f(x, \alpha, \beta) = \alpha e^{\beta x} / (1 + e^{\beta x}) \tag{15}
\]

Following our convention, let \( \theta = (\alpha, \beta)' \) be the parameter vector, and denote by \( \hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)' \) and \( \theta_0 = (\alpha_0, \beta_0)' \) the NLS estimator and the true value, respectively.

It is straightforward to show that the power function \( f \) given in (14) is H-regular with asymptotic order \( \kappa \) and limit homogeneous function \( h \) given respectively by

\[
\kappa(\lambda, \alpha, \beta) = \alpha \lambda^\beta \quad \text{and} \quad h(x, \alpha, \beta) = x^\beta 1\{x > 0\}
\]

Moreover, we have

\[
\kappa(\lambda, \alpha, \beta) = \begin{pmatrix} \lambda^\beta & 0 \\ \alpha \lambda^\beta \log \lambda & \alpha \lambda^\beta \end{pmatrix}
\]

and

\[
\dot{h}(x, \alpha, \beta) = \begin{pmatrix} x^\beta & 0 \\ x^\beta \log x & 1 \end{pmatrix} 1\{x > 0\}, \quad \nabla \dot{h}(x, \alpha, \beta) = \begin{pmatrix} \beta x^{\beta-1} \\ x^{\beta-1} (1 + \beta \log x) \end{pmatrix} 1\{x > 0\}
\]

If we let \( \Theta \) to be a compact subset of \( \mathbb{R} \setminus \{0\} \times \mathbb{R}_+ \), then we may easily check that all the conditions in Theorem 4.3 are satisfied for the power function. We may therefore obtain the limiting distribution of \( \hat{\theta}_n \) readily from Theorem 4.3.

In particular, if we let \( \dot{h} = (\dot{h}_1, \dot{h}_2)' \) and \( \nabla \dot{h} = (\nabla \dot{h}_1, \nabla \dot{h}_2)' \), and subsequently define

\[
M_{ij} = \int_0^1 \dot{h}_i(V, \theta_0) \dot{h}_j(V, \theta_0) \quad N_i(\rho) = \int_0^1 \dot{h}_i(V, \theta_0) dW(\rho) + \rho \int_0^1 \nabla \dot{h}_i(V, \theta_0)
\]
for $i, j = 1, 2$, then we have
\[
\frac{n(1+\beta_0)/2}{\log \sqrt{n}} (\hat{\alpha}_n - \alpha_0) = -\alpha_0 n(1+\beta_0)/2(\hat{\beta}_n - \beta_0) + O_p((\log \sqrt{n})^{-1}) \tag{16}
\]
\[
\rightarrow_d \left( M_{22} - \frac{M_{12}^2}{M_{11}} \right)^{-1} \left( N_2(\rho) - \frac{M_{12}N_1(\rho)}{M_{11}} \right)
\]
as $n \to \infty$.

The logistic function $f$ given in (15) does not satisfy our conditions in any of Theorems 4.1 – 4.3. Therefore, we may not directly apply them to obtain the asymptotics in this case. However, if we write $f = f_1 + f_2$, where
\[
f_1(x, \alpha) = \alpha 1\{x \geq 0\}
\]
\[
f_2(x, \alpha, \beta) = \alpha \left( e^{\beta x} 1\{x < 0\} - \frac{1}{1 + e^{\beta x}} 1\{x \geq 0\} \right)
\]
and define $f_2^*(x, \beta) = f_2(x, \alpha_0, \beta)$, then the NLS estimators $\hat{\alpha}_n$ and $\hat{\beta}_n$ from the regression with regression function $f$ have the same limiting distributions as the NLS estimators from the two separate regressions with regression functions $f_1$ and $f_2^*$ only. This is shown in Chang, Jiang and Park (2005).

The regression functions $f_1$ and $f_2^*$ satisfy, respectively, the conditions in Theorems 4.2 and 4.1. We may readily deduce that
\[
\sqrt{n}(\hat{\alpha}_n - \alpha_0) \rightarrow_d \left( \int_0^1 1\{V \geq 0\} dW(\rho) - \rho KL(1,0) \right) \tag{17}
\]
and
\[
\sqrt{n}(\hat{\beta}_n - \beta_0) \rightarrow_d MN \left( 0, \left[ \frac{\alpha_0^2 (\pi^2 - 6)}{18 \beta_0^3} L(1,0) \right]^{-1} \right) \tag{18}
\]
as $n \to \infty$. The limiting distribution of $\hat{\alpha}_n$ is affected by the presence of endogeneity, and given as a function of $\rho$. However, the distribution of $\hat{\beta}_n$ becomes independent of $\rho$ in the limit, and hence, is asymptotically invariant with respect to the endogeneity parameter $\rho$.

6. Simulations

In this section, we conduct simulations to convey the finite sample performances of the NLS estimators in the presence of endogeneity, and in particular they are compared to those obtained under no endogeneity. For our simulations, we consider the nonlinear regression model in (1) with the regression functions given by the power function and logistic function, specified respectively in (14) and (15). The regressor $x_t$ and the regression error $\varepsilon_t$ are allowed to be contemporaneously correlated and generated as in (2) and (3). To see how varying degrees of endogeneity affect the finite sample performances of the NLS estimators and their $t$-statistics, we try three different values of the parameter $\rho$ which measures the
fraction of the endogeneous component of the regression error: \( \rho = 0, 0.5, \) and 1. The innovations \( u_t \) and \( v_t \) are drawn from independent \( N(0, \sigma^2) \) with \( \sigma = 0.1. \) The true parameters are set at \( \alpha_0 = 1, \beta_0 = 3 \) for the regressions with the logistic regression function, and at \( \alpha_0 = 1, \beta_0 = 0.5 \) and 2 for the regressions with the power regression function. For the power function, two values for the true value of the power exponent \( \beta_0 \) are tried - one less than one and the other greater than one - to see how the behaviors of the NLS estimators change as the magnitude of the exponent \( \beta \) changes, especially around the unity. Samples of sizes \( n = 200, 500 \) are considered and each simulation is run for 10,000 times.

The finite sample performances of the NLS estimators are presented via a set of figures with the estimated densities of the centered NLS estimators and the corresponding \( t \)-statistics. Figures 1 – 6 present the results from the regression with power regression function. Figures 1 – 3 provide the results for the case with \( \beta_0 = 0.5. \) Figures 1 and 2 present the estimated densities of the centered NLS estimators computed with the three different values of \( \rho \) for the cases with \( n = 200 \) and \( n = 500 \), respectively. The densities of both unscaled and scaled centered NLS estimators are provided respectively in the left and right hand side columns of each figure. The scaling factors are computed according to the convergence rates derived in (16). Panels in each row represent the density estimates from using three different values of \( \rho = 0, 0.5, 1. \) Figure 3 provides the estimated densities for the \( t \)-statistics for the both sample sizes \( n = 200, 500. \) Figures 4 – 6 provide the results for the case with \( \beta_0 = 2 \) in the same format as in Figures 1 – 3. Figures 7 – 9 present the results from the regressions with logistic regression function. The results are presented in the same manner as in the cases with the power regression function given in Figures 1 – 3. The scaling factors for the NLS estimators in this case are taken from the convergence rates given in (17) and (18).

The simulation results largely corroborate the limit theories derived in (16) – (18) earlier in this section. As expected from (16), the limit distributions of the NLS estimators from the power regression function for both parameters \( \alpha \) and \( \beta \) are dependent upon the presence of the endogeneity. When there is no endogeneity with \( \rho = 0, \) the limit distributions of both NLS estimators \( \hat{\alpha}_n \) and \( \hat{\beta}_n \) become mixed normal and their \( t \)-statistics have standard normal limit distribution. These can be seen from the panels in the first row of each figures. As \( \rho \) increases, however, the distributions of NLS estimators and \( t \)-statistics for both parameters become more nonstandard and nonnormal. This is true for both cases with the values of \( \beta_0 = 0.5, 2, \) although the density estimates appear to be more nonstandard and depart more from the normality in the cases with the smaller \( \beta_0. \) The dependence of the distributions on the degree of endogeneity \( \rho \) does not vanish as the sample size increases and indeed the nonstandardness and nonnormality of the distributions persist.

The finite sample performance of the NLS estimators from the logistic regression is also as predicted by the limit theories given in (17) and (18). The estimated densities for the NLS estimators \( \hat{\alpha}_n \) and \( \hat{\beta}_n \) of \( \alpha \) and \( \beta \) indicate the NLS estimator \( \hat{\alpha}_n \) converges faster than \( \hat{\beta}_n. \) They also show that the distributions of the \( \hat{\alpha}_n \) depends on the degree of endogeneity \( \rho \) present in the model, and that their dependence on \( \rho \) does not vanish even with larger sample size. The distributions of \( \hat{\beta}_n, \) however, do not dependent upon the magnitude of \( \rho. \)

---

4 The densities are estimated via usual kernel method using normal kernel function.
In the smaller samples, they seem to have some biases when \( \rho = 1 \), but they disappear as the sample size increases.

The distributions of the \( t \)-statistics also corroborate our limit theory. The estimated density of the \( t \)-ratio for \( \alpha \) is clearly nonnormal when \( \rho \neq 0 \), and the departure from the normality becomes much more noticeable as \( \rho \) increases. This continues to be the case even when the sample size gets large. On the other hand, the estimated densities of the \( t \)-ratios for \( \beta \) are quite close to the standard normal density which is also drawn in the same panel for easy comparison. And they approach closer to the limit standard normal density as the sample size increases.

7. Conclusion

To be completed.

8. Mathematical Proofs

8.1 Useful Lemmas

**Lemma A1** Let \( T \) be a transformation on \( \mathbb{R} \).

(a) If Assumptions 2.2 with \( \nu = 2 \), and 2.3 hold, and if \( T \) is locally Riemann-integrable and

\[
|T(x)| < c/(1 + |x|^{1+\epsilon})
\]

for some \( c > 0 \) and \( \epsilon > 0 \), then

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} T(x_t) \rightarrow_d L(1, 0) \int_{-\infty}^{\infty} T(x) \mu(dx)
\]

as \( n \rightarrow \infty \).

(b) If Assumptions 2.2 with \( \nu = 3 \), and 2.3 hold, and if \( T \) is bounded and

\[
\int_{-\infty}^{\infty} |x|^{1/2+\epsilon}|T(x)| \mu(dx) < \infty
\]

for some \( \epsilon > 0 \), then we have

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} T(x_t)v_t \rightarrow_d MN \left(0, L(1, 0) \int_{-\infty}^{\infty} T^2(x) \mu(dx)\right)
\]

as \( n \rightarrow \infty \).

**Proof of Lemma A1** For part (a), see Theorems 2.1, Chapter IV, in Borodin and Ibragimov (1995). To prove part (b), we let

\[
f(x, y) = T(y)(y - x)
\]
and apply Theorems 1.1 and 1.3 of Borodin (1986). Note that we have $f(x, x + y) = T(x + y)$. The integrability of $T$ follows immediately from the condition $\int_{-\infty}^{\infty} |x|^{1/2+\epsilon} |T(x)| \mu(dx) < \infty$. We also have the square integrability of $T$ with respect to both the Lebesgue and counting measures. Clearly, with respect to the counting measure, the square integrability of $T$ is implied by the integrability $T$. Moreover, being bounded, the square integrability $T$ with respect to Lebesgue measure also follows from the integrability of $T$. We therefore have $\int_{-\infty}^{\infty} T^2(x) \mu(dx) < \infty$ under the given conditions.

We first check if the conditions (1.1), (1.2), (1.7) and (1.8) in Borodin (1986) are met. We have

$$\int E f^2(x, x + v_t) \mu(dx) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f^2(x, x + y)p(y)\mu(dy)\mu(dx)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} T^2(x + y)y^2p(y)\mu(dy)\mu(dx)$$

$$= \left( \int_{-\infty}^{\infty} T^2(x) \mu(dx) \right) \left( \int_{-\infty}^{\infty} y^2p(y)\mu(dy) \right) < \infty$$

as required. Moreover,

$$\int_{-\infty}^{\infty} E |v_t|^{1/2+\epsilon} |f(x, x + v_t)| \mu(dx) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |f(x, x + y)||y|^{1/2+\epsilon} p(y)\mu(dy)\mu(dx)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |T(x + y)||y|^{3/2+\epsilon} p(y)\mu(dy)\mu(dx)$$

$$= \left( \int_{-\infty}^{\infty} |T(x)| \mu(dx) \right) \left( \int_{-\infty}^{\infty} |y|^{2/3+\epsilon} p(y) \mu(dy) \right) < \infty$$

and

$$\int_{-\infty}^{\infty} E |x|^{1/2+\epsilon} |f(x, x + v_t)| \mu(dx) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x|^{1/2+\epsilon} |f(x, x + y)|p(y)\mu(dy)\mu(dx)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x|^{1/2+\epsilon} |T(x + y)||y|p(y)\mu(dy)\mu(dx)$$

$$\leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (|x|^{1/2+\epsilon} + |y|^{1/2+\epsilon})|T(x)||y|p(y)\mu(dy)\mu(dx)$$

$$= \left( \int_{-\infty}^{\infty} |x|^{1/2+\epsilon} |T(x)| \mu(dx) \right) \left( \int_{-\infty}^{\infty} |y|^{3/2+\epsilon} p(y) \mu(dy) \right) < \infty$$

as was to be shown.

We now let $h, \nu, \rho$ and $b$ be defined as in Borodin (1986). Then it follows that

$$h(x) = E f(x, x + v_t)$$

$$= \int_{-\infty}^{\infty} f(x, x + y)p(y)\mu(dy)$$

$$= \int_{-\infty}^{\infty} T(x + y)\nu p(y)\mu(dy)$$
and therefore, \( \int_{-\infty}^{\infty} h(x) \mu(dx) = 0 \). We also have

\[
\nu(x) = \int_{-\infty}^{\infty} e^{-ixy} h(y) \mu(dy)
\]

\[
= \int_{-\infty}^{\infty} e^{-ixy} \int_{-\infty}^{\infty} T(y + z) p(z) \mu(dz) \mu(dy)
\]

\[
= \int_{-\infty}^{\infty} e^{-ix(y-z)} \int_{-\infty}^{\infty} T(y) p(z) \mu(dz) \mu(dy)
\]

\[
= \left( \int_{-\infty}^{\infty} e^{-ixy} T(y) \mu(dy) \right) \left( \int_{-\infty}^{\infty} e^{ixz} p(z) \mu(dz) \right)
\]

so that

\[
|\nu(x)| \leq \left( \int_{-\infty}^{\infty} |T(y)| \mu(dy) \right) \left( \int_{-\infty}^{\infty} |z| p(z) \mu(dz) \right)
\]

i.e., \( \nu \) is bounded by a constant, and

\[
\rho(x) = \int_{-\infty}^{\infty} e^{ixy} \mathbf{E} (e^{iyz} f(y, y + v_t)) \mu(dy)
\]

\[
= \int_{-\infty}^{\infty} e^{ixy} \int_{-\infty}^{\infty} e^{iyz} p(y, y + z) \mu(dy) \mu(dy)
\]

\[
= \int_{-\infty}^{\infty} e^{ixz} \int_{-\infty}^{\infty} e^{iyz} T(y + z) p(z) \mu(dz) \mu(dy)
\]

\[
= \left( \int_{-\infty}^{\infty} e^{ixz} T(y) \mu(dy) \right) \left( \int_{-\infty}^{\infty} z p(z) \mu(dz) \right) = 0
\]

Finally,

\[
b = \int_{-\infty}^{\infty} \mathbf{E} f^2(x, x + v_t) \mu(dx)
\]

\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f^2(x, x + y) p(y) \mu(dy) \mu(dx)
\]

\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} T^2(x + y) y^2 p(y) \mu(dy) \mu(dx)
\]

\[
= \int_{-\infty}^{\infty} T^2(x) \mu(dx)
\]

The result in part (b) now follows immediately from Borodin (1986).

**Lemma A2** If Assumptions 2.2 with \( \nu = 3 \) and 2.3 hold, then

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (1\{x_t \geq 0\} - 1\{x_{t-1} \geq 0\}) v_t \rightarrow_d -L(1, 0) \int_{-\infty}^{y} x p(x) \mu(dx) \mu(dy)
\]
\[ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (1\{x_t < 0\} - 1\{x_{t-1} < 0\})v_t \to_d \mathcal{L}(1,0) \int_{-\infty}^{y} x p(x) \mu(dx) \mu(dy) \]

as \( n \to \infty \).

**Proof of Lemma A2**  Let

\[ f(x,y) = (1\{y \geq 0\} - 1\{x \geq 0\})(y - x) \]

note that

\[ f(x, x + v_t) = (1\{x + v_t \geq 0\} - 1\{x \geq 0\})v_t \]
\[ = (1\{x + v_t \geq 0\}1\{x < 0\} - 1\{x + v_t < 0\}1\{x \geq 0\})v_t \]
\[ = (1\{v_t \geq -x > 0\} - 1\{v_t < -x \leq 0\})v_t \]

As in the proof of Lemma A1, we first check the conditions for Theorems 1.1 and 1.3 of Borodin (1986).

We have

\[ \int_{-\infty}^{\infty} \mathbb{E} f^2(x, x + v_t) \mu(dx) \]
\[ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (1\{y \geq -x > 0\} + 1\{y < -x \leq 0\})y^2 p(y) \mu(dy) \mu(dx) \]
\[ = \left( \int_{-\infty}^{0} \int_{-x}^{\infty} + \int_{0}^{\infty} \int_{-\infty}^{-x} \right) y^2 p(y) \mu(dy) \mu(dx) < \infty \]

Furthermore,

\[ \int_{-\infty}^{\infty} \mathbb{E} |v_t|^{1/2+\epsilon} |f(x, x + v_t)| \mu(dx) \]
\[ \leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (1\{y \geq -x > 0\} + 1\{y < -x \leq 0\})|y|^{3/2+\epsilon} p(y) \mu(dy) \mu(dx) \]
\[ = \left( \int_{-\infty}^{0} \int_{-x}^{\infty} + \int_{0}^{\infty} \int_{-\infty}^{-x} \right) |y|^{3/2+\epsilon} p(y) \mu(dy) \mu(dx) < \infty \]

and

\[ \int_{-\infty}^{\infty} \mathbb{E} |x|^{1/2+\epsilon} |f(x, x + v_t)| \mu(dx) \]
\[ \leq \int_{-\infty}^{\infty} |x|^{1/2+\epsilon} \int_{-\infty}^{\infty} (1\{y \geq -x > 0\} + 1\{y < -x \leq 0\})|y| p(y) \mu(dy) \mu(dx) \]
\[ = \left( \int_{-\infty}^{0} (-x)^{1/2+\epsilon} \int_{-x}^{\infty} + \int_{0}^{\infty} x^{1/2+\epsilon} \int_{-\infty}^{-x} \right) |y| p(y) \mu(dy) \mu(dx) < \infty \]

as was to be shown.
Now we let

\[ h(x) = Ef(x, x + v_t) \]

Then we have

\[
Ef(x, x + v_t) = -1\{x \geq 0\} \int_{-\infty}^{-x} yp(y)\mu(dy) + 1\{x < 0\} \int_{-x}^{\infty} yp(y)\mu(dy) = -\int_{-\infty}^{-x} yp(y)\mu(dy)
\]

since \( \int_{-\infty}^{\infty} yp(y)\mu(dy) = 0 \). Therefore, follows from Theorem 1.1 and 1.4 of Borodin (1986) that

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} f(x_{t-1}, x_t) \to_d L(1, 0) \int_{-\infty}^{\infty} h(x)\mu(dx) = -L(1, 0) \int_{-\infty}^{\infty} \int_{-\infty}^{y} xp(x)\mu(dx)\mu(dy)
\]

as \( n \to \infty \). We thus have shown the first part of the stated results. The proof for the second part of the stated results is completely analogous and omitted.

**Lemma A3** Let \( T \) be a transformation on \( \mathbb{R} \) such that

(a) \( |T(x)| \leq c|x|^a \) for all \( \mathbb{R}\{-0\} \) in a neighborhood of origin for some constants \( a > -1 \) and \( c > 0 \), and

(b) \( T(x) \) is differentiable with continuous derivative for all \( \mathbb{R}\{-0\} \) such that \( |T'(x)| \leq c|x|^b \) for some constants \( b \) and \( c > 0 \) in a neighborhood of the origin.

Moreover, let \( W_n \) be a stochastic process taking values in \( D[0, 1] \) such that \( W_n \to_p W \) as \( n \to \infty \), where \( W \) is a Brownian motion. Then we have

\[
\int_0^1 T(W_n(r)) dr \to_p \int_0^1 T(W(r)) dr
\]

as \( n \to \infty \).

**Proof of Lemma A3** The stated result follows directly from the proof of Theorem 3.2 in Park (2003).

**8.2 Proofs of Theorems**

**Proof of Lemma 3.2** For the proof of part (a), we write

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} T\left(\frac{x_t}{\sqrt{n}}\right) v_t = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} T\left(\frac{x_{t-1}}{\sqrt{n}}\right) v_t + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[ T\left(\frac{x_t}{\sqrt{n}}\right) - T\left(\frac{x_{t-1}}{\sqrt{n}}\right) \right] v_t \quad (19)
\]

It follows immediately from PP that

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} T\left(\frac{x_{t-1}}{\sqrt{n}}\right) v_t \to_d \int_0^1 T(V(r)) dV(r) \quad (20)
\]
as \( n \to \infty \). Next, we write

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[ T \left( \frac{x_t}{\sqrt{n}} \right) - T \left( \frac{x_{t-1}}{\sqrt{n}} \right) \right] v_t = A_n + B_n + C_n + D_n
\]  

(21)

where

\[
A_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[ T \left( \frac{x_t}{\sqrt{n}} \right) - T \left( \frac{x_{t-1}}{\sqrt{n}} \right) \right] \{ x_t \geq 0 \} \{ x_{t-1} \geq 0 \} v_t
\]

\[
B_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[ T \left( \frac{x_t}{\sqrt{n}} \right) - T \left( \frac{x_{t-1}}{\sqrt{n}} \right) \right] \{ x_t \geq 0 \} \{ x_{t-1} < 0 \} v_t
\]

\[
C_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[ T \left( \frac{x_t}{\sqrt{n}} \right) - T \left( \frac{x_{t-1}}{\sqrt{n}} \right) \right] \{ x_t < 0 \} \{ x_{t-1} \geq 0 \} v_t
\]

\[
D_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[ T \left( \frac{x_t}{\sqrt{n}} \right) - T \left( \frac{x_{t-1}}{\sqrt{n}} \right) \right] \{ x_t < 0 \} \{ x_{t-1} < 0 \} v_t
\]

Whenever \( x_{t-1} < 0 \) and \( x_t \geq 0 \), we have

\[ v_t > 0 \]

Therefore, if we define

\[
M_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{ x_t \geq 0 \} \{ x_{t-1} < 0 \} v_t
\]  

(22)

\[
N_n = \sup_{1 \leq t \leq n} \left| T \left( \frac{x_t}{\sqrt{n}} \right) - T \left( \frac{x_{t-1}}{\sqrt{n}} \right) \right| \{ x_t \geq 0 \} \{ x_{t-1} < 0 \}
\]  

(23)

then it follows that

\[
B_n \leq M_n N_n
\]  

(24)

Since \( T \) is assumed to be locally bounded, we have \( N_n = O_p(1) \). Moreover, it follows from the proof of Lemma A2 that \( M_n = O_p(1) \). We therefore have \( B_n = O_p(1) \). It is quite obvious that we may use the same argument and show that \( C_n = O_p(1) \).

Now we consider \( A_n \) and \( D_n \). When \( x_t > 0 \) and \( x_{t-1} > 0 \), we have

\[
T \left( \frac{x_t}{\sqrt{n}} \right) - T \left( \frac{x_{t-1}}{\sqrt{n}} \right) = \nabla T \left( \frac{x_{t-1}}{\sqrt{n}} + \frac{w_t}{\sqrt{n}} \right) \frac{v_t}{\sqrt{n}}
\]

where \( \{ w_t \} \) is a random sequence such that \( |w_t| \leq |v_t| \). It follows that

\[
A_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[ T \left( \frac{x_t}{\sqrt{n}} \right) - T \left( \frac{x_{t-1}}{\sqrt{n}} \right) \right] \{ x_t \geq 0 \} \{ x_{t-1} \geq 0 \} v_t
\]

\[
= \frac{1}{n} \sum_{t=1}^{n} \nabla T \left( \frac{x_{t-1}}{\sqrt{n}} + \frac{w_t}{\sqrt{n}} \right) \{ x_t \geq 0 \} \{ x_{t-1} \geq 0 \} v_t^2
\]
\[
\begin{align*}
= & \frac{1}{n} \sum_{t=1}^{n} \nabla T \left( \frac{x_{t-1}}{\sqrt{n}} + \frac{w_t}{\sqrt{n}} \right) 1\{x_t \geq 0\} 1\{x_{t-1} \geq 0\} \\
& + \frac{1}{n} \sum_{t=1}^{n} \nabla T \left( \frac{x_{t-1}}{\sqrt{n}} + \frac{w_t}{\sqrt{n}} \right) 1\{x_t \geq 0\} 1\{x_{t-1} \geq 0\} (v_t^2 - 1)
\end{align*}
\] (25)

Under the given conditions, we have
\[
\left\{ x_t \geq 0 \right\} \left\{ x_{t-1} \geq 0 \right\} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \nabla T \left( \frac{x_{t-1}}{\sqrt{n}} + \frac{w_t}{\sqrt{n}} \right) 1\{x_t \geq 0\} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( 1\{x_t \geq 0\} - 1\{x_{t-1} \geq 0\} \right) v_t \to_d \int_{0}^{1} \nabla T(V(r)) 1\{V(r) \geq 0\} dr
\]
due to Lemma A3. Moreover,
\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \nabla T \left( \frac{x_{t-1}}{\sqrt{n}} + \frac{w_t}{\sqrt{n}} \right) 1\{x_t \geq 0\} 1\{x_{t-1} \geq 0\} (v_t^2 - 1) = O_p(1)
\]
and we may easily show that
\[
\frac{1}{n} \sum_{t=1}^{n} \left( \nabla T \left( \frac{x_{t-1}}{\sqrt{n}} + \frac{w_t}{\sqrt{n}} \right) - \nabla T \left( \frac{x_{t-1}}{\sqrt{n}} \right) \right) 1\{x_t \geq 0\} 1\{x_{t-1} \geq 0\} (v_t^2 - 1) = o_p(1)
\]
using Cauchy-Schwarz and Lemma A3. Note that
\[
\max_{1 \leq t \leq n} \frac{|w_t|}{\sqrt{n}} \leq \max_{1 \leq t \leq n} \frac{|v_t|}{\sqrt{n}} \to_p 0
\]
We thus have shown that \( A_n = O_p(1) \). We may similarly show that \( D_n = O_p(1) \). The proof for part (a) is therefore complete.

For part (b), we first consider the case that \( T \) has vanishing derivative on \( \mathbb{R} \setminus \{0\} \). We may then assume without loss of generality that \( T(x) = a1\{x \geq 0\} + b1\{x < 0\} \) for some constants \( a \) and \( b \). Write
\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} 1\{x_t \geq 0\} v_t
\]
\[
= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} 1\{x_{t-1} \geq 0\} v_t + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( 1\{x_t \geq 0\} - 1\{x_{t-1} \geq 0\} \right) v_t
\]
Then we may deduce from PP that
\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} 1\{x_{t-1} \geq 0\} v_t \to_d \int_{0}^{1} 1\{V(r) \geq 0\} dV(r)
\]
as \( n \to \infty \), and the stated result follows from Lemma A2.

We now consider the case for which \( T \) is continuous and has nonvanishing derivative. We have (19) – (21) as in the proof of part (a). We first show that
\[
B_n = o_p(1)
\] (26)
Whenever $x_{t-1} < 0$ and $x_t \geq 0$, we have

$$|x_t|, |x_{t-1}| \leq |v_t|$$

as well as $v_t > 0$. If we let $N_n$ be defined as in (23), then we may therefore readily deduce that $N_n = o_p(1)$, since $T$ is continuous at the origin and

$$\max_{1 \leq t \leq n} \frac{|v_t|}{\sqrt{n}} \rightarrow_p 0$$

This establishes (26). It is quite obvious that we may deduce

$$C_n = o_p(1)$$

(27)

similarly as for $B_n$.

We may write $A_n$ as in (25) and obtain

$$A_n \rightarrow_d \int_0^1 \nabla T(V(r))1\{V(r) \geq 0\} \, dr$$

(28)

as $n \rightarrow \infty$. Similarly, we may derive

$$D_n \rightarrow_d \int_0^1 \nabla T(V(r))1\{V(r) < 0\} \, dr$$

(29)

as $n \rightarrow \infty$. The stated result now follows from (20) and (26) – (29). \[ \square \]

**Proof of Lemma 3.5** The stated result follows immediately from part (b) of Lemma A1. \[ \square \]

**Proof of Lemma 3.6** From Definition 3.4 we have $F(\lambda x, \pi) = \kappa(\lambda, \pi)H(x, \pi) + R(x, \lambda, \pi)$, and this gives

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

Note that $Q$ is regular and

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} Q\left(\frac{x_t}{\sqrt{n}}, \pi\right) v_t = O_p(1)$$

due to Lemma 3.2 (a), and $(\kappa^{-1}(\pi))(\lambda) \rightarrow 0$ as $\lambda \rightarrow \infty$. Note also that $\kappa(\cdot, \pi) = 1$ when $H(\cdot, \pi)$ has vanishing derivative. The stated results now follow directly from Lemma 3.2 (b). \[ \square \]
Proof of Theorem 4.1  Under the given conditions and our previous results in (8), (9) and Lemma 3.5, we may show that the NLS estimator $\hat{\theta}_n$ is asymptotically equivalent to the least squares estimator of $\theta$ in the linear regression

$$y_t = \hat{f}(x_t, \theta_0)'\theta + \varepsilon_t \tag{30}$$

The stated result thus follows immediately. The proof is essentially identical to that of Theorem 5.1 in PP, except for our covariance asymptotics under endogeneity given in Lemma 3.5.

Proof of Theorem 4.2  Once again, the proof is virtually the same as that of Theorem 5.2 in PP. We only need to replace their covariance asymptotics (11) by a mixture of our covariance asymptotics under endogeneity in Lemma 3.6 and theirs. In this case, we also have the asymptotic equivalence between the NLS estimator $\hat{\theta}_n$ and the least squares estimator of $\theta$ in the linear regression (30) defined in the proof of Theorem 4.1.

Proof of Theorem 4.3  The arguments in the proofs of Theorem 4.1 and 4.2 also apply to the proof of Theorem 4.3, which is analogous to that of Theorem 5.3 in PP. The details are therefore omitted.

References


Figure 1: Densities of NLS Estimators, Power Function, $\beta_0 = 0.5$, $n = 200$
Figure 2: Densities of NLS Estimators, Power Function, $\beta_0 = 0.5$, $n = 500$
Figure 3: Densities of NLS $t$-ratios, Power Function, $\beta_0 = 0.5$
Figure 4: Densities of NLS Estimators, Power Function, $\beta_0 = 2$, $n = 200$
Figure 5: Densities of NLS Estimators, Power Function, $\beta_0 = 2$, $n = 500$
Figure 6: Densities of NLS $t$-ratios, Power Function, $\beta_0 = 2$
Figure 7: Densities of NLS Estimators, Logistic Function, $n = 200$
Figure 8: Densities of NLS Estimators, Logistic Function, $n = 500$
Figure 9: Densities of NLS $t$-ratios, Logistic Function