A note on nonlinear models with integrated regressors and convergence order results*

Robert de Jong[†] Ling Hu[‡]

September 2, 2010

Abstract

This note concerns an asymptotic distribution result from the literature on nonlinear estimation in the presence of integrated variables. It points out a way of strengthening local asymptotic distribution results from the literature towards results that hold for the global minimizer of the criterion function. An asymptotic distribution result for the global minimizer that assumes that a convergence rate for the global minimizer is in evidence is stated, and applied to nonstationary nonlinear least squares.

Keywords and phrases: Nonlinear cointegration; Cointegration; Unit root

JEL classification code: C22

*We thank Professors Phillips, Wooldridge and Jeganathan for discussions and suggestions. This version: January 2009.

[†]444 Arps Hall, Department of Economics, Ohio State University, 1945 N. High Street, Columbus 43210, USA, email de-jong.8@osu.edu

[‡]Direct correspondence to: 410 Arps Hall, Department of Economics, Ohio State University, 1945 N. High Street, Columbus 43210, USA, email huling@gmail.com

1 Introduction

Let $Q_T(\cdot) : \mathbb{R}^p \to \mathbb{R}$ denote a random function and let $\hat{\theta}$ denote the (global) minimizer of this criterion function over a parameter space Θ. Let $S_T(\theta)$ and $H_T(\theta)$ denote $(\partial/\partial\theta)Q_T(\theta)$ and $(\partial^2/\partial\theta\partial\theta')Q_T(\theta)$ respectively, and let $|M|$ for any matrix M denote $(\mathrm{tr}(M'M))^{1/2}$.

In the literature on nonlinear minimization estimators (e.g. Park and Phillips (2001)), it has been argued that for the existence of a sequence of local minimizers θ_T^* that follow the standard asymptotic distribution results, the following two conditions suffice, in addition to standard requirements of compactness of the parameter space, measurability and continuity, and the true parameter not being located on the boundary of the parameter space:

1. $(D_T^{-1/2} H_T(\theta_0) D_T^{-1/2}, D_T^{-1/2} S_T(\theta_0)) \xrightarrow{d} (\mathcal{A}_0, \mathcal{J}_0)$ where $\mathcal{A}_0$ is positive definite with probability one;
2. $\max_{\mathcal{N}_T^0} |C_T^{-1/2}(H_T(\theta_0) - H_T(\theta)) C_T^{-1/2}| = o_p(1)$, where $\mathcal{N}_T^0 = \{\theta \in \Theta : |C_T^{1/2}(\theta - \theta_0)| \leq 1\}$.

This reasoning is based on Wooldridge (1994), Theorem 8.1 and 10.1. Here C_T and D_T are deterministic diagonal matrices of dimension $(p \times p)$ such that $C_{Tii} \to \infty$ and $D_{Tii} \to \infty$ for $i = 1, \ldots, p$, and $C_T D_T^{-1} \to 0$ as $T \to \infty$. In standard estimation theory for minimization estimators involving averages of i.i.d. random variables, we would typically choose $D_T = T I_p$, where I_p denotes the $(p \times p)$ identity matrix.

Asymptotic results for a sequence of local minimizers have a long history in statistics; Cramèr (1946) already formulated results for local minimizers. Such results leave the problem that in a practical situation for a given sample, one does not know which local minimizer is the one for which the asymptotic distribution behavior has been determined.

In this note, it is argued that in the absence of global convexity of the objective function, an assumption of the type $C_T^{1/2}(\hat{\theta} - \theta_0) = O_p(1)$ can be added to our list of key assumptions in order to additionally obtain limit theory for the global minimizer $\hat{\theta}$. Note that in the case of a probit or ordered probit model (see Park and Phillips (2000), Hu and Phillips (2004) and Phillips, Jin and Hu (2005)), the global concavity of the loglikelihood ensures that the minimizer is necessarily unique, implying that a local minizer necessarily equals the global minimizer. See Pratt (1981) for a proof of this concavity result.

An earlier version of this note had a detailed example in which the global minimizer of an objective function was $O_p(T^{-1/3})$, while a sequence of local minimizers θ_T^* satisfied $T^{1/2}(\theta_T^* - \theta_0) \xrightarrow{d} N(0, 1)$.

2 A limit theorem for minimization estimators

This section establishes the result alluded to in the Introduction:

Theorem 1 *Let* $\{Q_T : \mathcal{W} \times \Theta \to \mathbb{R}, T = 1, 2, \ldots\}$ *be a sequence of objective functions defined on the data space* $\mathcal{W}$ *and the parameter space* $\theta \subset \mathbb{R}^p$*. Assume that*

1. Θ *is compact and* $\theta_0 \in \operatorname{int}(\Theta)$*;*

2. $Q_T(\cdot,\cdot)$ *satisfies the standard measurability and second order differentiability conditions on* $\mathcal{W} \times \Theta$*,* $T = 1, 2\ldots$*;*

3. *For sequences of positive definite diagonal matrices* C_T *and* D_T *such that* $C_T D_T^{-1} \to 0$ *as* $T \to \infty$ *and* $C_{Tii} \to \infty$ *for all* $i = 1, \ldots, p$*,*

$$\sup_{\mathcal{N}_T^0} |C_T^{-1/2}(H_T(\theta_0) - H_T(\theta))C_T^{-1/2}| = O_p(1)$$

   where $\mathcal{N}_T^0 = \{\theta \in \Theta : |C_T^{1/2}(\theta - \theta_0)| \leq 1\}$*;*

4. $(D_T^{-1/2} H_T(\theta_0) D_T^{-1/2}, D_T^{-1/2} S_T(\theta_0)) \xrightarrow{d} (\mathcal{A}_0, \mathcal{J}_0)$ *where* $\mathcal{A}_0$ *is positive definite with probability one;*

5. $C_T^{1/2}(\hat{\theta} - \theta_0) = O_p(1)$*.*

Then

$$D_T^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{A}_0^{-1} \mathcal{J}_0.$$

The proofs of the theorems of this note are given in the Appendix. Theorem 1 has a simple proof and should not considered new, because it only uses well-known arguments and uses a simple reordering of the various objects using the C_T and D_T matrices.

The above result means that wherever the "local minimizer" reasoning (as outlined in the Introduction) was applied, addition of the regularity condition that $C_T^{1/2}(\hat{\theta} - \theta_0) = O_p(1)$ is sufficient in order to obtain asymptotic distribution results for the global minimum. However, this condition is a strengthened version of a consistency result, and is not necessarily easy to obtain.

3 Nonlinear nonstationary least squares

As in Park and Phillips (2001), in this section we consider the nonlinear regression model $y_t = g(x_t, \theta) + \varepsilon_t$ where x_t is an integrated process, and ε_t is a martingale difference sequence. We study the nonlinear least square estimator

$$\hat{\theta} = \operatorname{argmin}_{\theta \in \Theta} \sum_{t=1}^{T} (y_t - g(x_t, \theta))^2.$$

In order to derive convergence rates for nonlinear least squares estimators, we use the following result. This result is a simple adaptation of a classical result that is used to show consistency of nonlinear least squares estimators if optimization is not over a compact set, but over the entire real line; see Pötscher and Prucha (1997) for a discussion. Below, as in Park and Phillips (2001), $\dot{g}(\cdot,\cdot)$ denotes $(\partial/\partial\theta)g(x_t,\theta)$.

Theorem 2 *Define* $A_T = \sum_{t=1}^T \inf_{\theta\in\Theta} |\dot{g}(x_t,\theta)|^2$ *and* $B_T = \sup_{\theta\in\Theta} |\sum_{t=1}^T \varepsilon_t \dot{g}(x_t,\theta)|$. *If* m_T *is a strictly positive deterministic sequence such that* $m_T \to 0$ *and* $B_T/(m_T A_T) = O_p(1)$, *then* $m_T^{-1}(\hat{\theta}-\theta_0) = O_p(1)$.

The above result is potentially restrictive, but can be used to establish a convergence rate for the nonlinear least squares estimator. By combining Theorems 1 and 2, it now follows that the limit theory as established in Park and Phillips (2001) also holds for the global minimizer of the objective function.

3.1 The integrable function case

If $\dot{g}(x,\cdot)$ is integrable over x for any θ, we suggest the addition of the following assumptions to those imposed by Park and Phillips (2001):

Assumption 1 $\inf_{\theta\in\Theta}|\dot{g}(x,\theta)|^2$ *is a well-defined function of* x,

$$T^{-1/2}\sum_{t=1}^T \inf_{\theta\in\Theta}|\dot{g}(x_t,\theta)|^2 \xrightarrow{d} L(1,0)\int_{-\infty}^{\infty}\inf_{\theta\in\Theta}|\dot{g}(s,\theta)|^2 ds > 0,$$

and

$$\sup_{\theta\in\Theta}|\sum_{t=1}^T \varepsilon_t \dot{g}(x_t,\theta)| = O_p(T^{1/4}).$$

While the first assumption seems not very strict, it may be complicated to derive the second assumption. Given Assumption 1, an application of Theorem 2 now gives the convergence rate for the least squares estimator:

Corollary 1 *Under Assumption 1,* $T^{1/4}(\hat{\theta}-\theta_0) = O_p(1)$.

Corollaries 1 and 2 (below) are straightforward to prove from Theorem 2.

The above corollary establishes the correct convergence rate for the nonlinear least squares estimator for the integrable case. Since in Park and Phillips (2001) condition 2 of the Introduction has been verified for some C_T sequence that is $o(T^{1/4})$, it follows from Theorem 1 that the extra conditions of Assumption 1 suffice for showing that the limit theory of Park and Phillips (2001) holds for the global minimizer of the objective function.

3.2 The asymptotically homogeneous case

The treatment of Park and Phillips (2001) of the case where $\dot{g}(\cdot,\cdot)$ is asymptotically homogeneous rests on the notion that as $\lambda \to \infty$, $\dot{g}(\lambda x, \theta) \approx \dot{\nu}(\lambda)\dot{H}(x,\theta)$. The following assumption therefore appears natural in this setting:

Assumption 2 $\inf_{\theta\in\Theta} |\dot{g}(x,\theta)|^2$ *is a well-defined function of x, for some* $\lambda > 0$*,*

$$\dot{\nu}(T^{1/2})^{-2}T^{-1}\sum_{t=1}^{T}\inf_{\theta\in\Theta}|\dot{g}(x_t,\theta)|^2 \xrightarrow{d} \int_0^1 \inf_{\theta\in\Theta}|\dot{H}(\lambda W(r),\theta)|^2 dr > 0,$$

and

$$\sup_{\theta\in\Theta}|\sum_{t=1}^{T}\varepsilon_t \dot{g}(x_t,\theta)| = O_p(\dot{\nu}(T^{1/2})T^{1/2}).$$

Corollary 2 *Under Assumption 2,* $\dot{\nu}(T^{1/2})T^{1/2}(\hat{\theta}-\theta_0) = O_p(1)$*.*

The above result also establishes the correct convergence rate. Therefore, similarly to the integrable case, it follows that under Assumption 2, the results of Park and Philips (2001) for the asymptotically homogeneous case also hold for the global minimizer of the objective function.

References

Cramèr, H. (1946), *Mathematical Methods of Statistics.* Princeton University Press.

Hu, L. and P. C. B. Phillips (2004), Nonstationary discrete choice, *Journal of Econometrics* 120, 103-138.

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* 68, 1249-1280.

Park, J.Y. and P.C.B. Phillips (2001), Nonlinear regressions with integrated time series, *Econometrica* 69, 117-161.

Phillips, P.C.B., S. Jin and L. Hu (2005), Nonstationary Discrete Choice: A Corrigendum and Addendum, forthcoming in *Journal of Econometrics.*

Pollard, D. (1985), New ways to prove central limit theorems, *Econometric Theory* 1, 295-313.

Pötscher, B.M and I. Prucha (1997), *Dynamic Nonlinear Econometric Models: Asymptotic Theory.* Berlin: Springer-Verlag.

Pratt, J.W. (1981), Concavity of the log likelihood, *Journal of the American Statistical Association* 76, 103-106.

Wooldridge, J. (1994), *Estimation and inference for dependent processes.* In R.F. Engle and D.L. McFadden (eds.) *Handbook of Econometrics*, Vol. IV, pp. 2639-2738. Elsevier, Amsterdam.

Appendix: Mathematical Proofs

Proof of Theorem 1:

Because $C_{Tii} \to \infty$ as $T \to \infty$ for $i = 1, \ldots, p$, $\hat{\theta}$ will eventually be inside an open neighborhood of θ_0 that lies inside Θ with arbitrarily large probability by Assumption 1.5, implying that with arbitrarily large probability, for some mean value $\tilde{\theta}$, using the usual convention of ignoring the dependence of $\tilde{\theta}$ on its location in the vector,

$$0 = D_T^{-1/2} S_T(\hat{\theta}) = D_T^{-1/2} S_T(\theta_0) + (D_T^{-1/2} H_T(\theta_0) D_T^{-1/2}) C_T^{1/2}(\hat{\theta} - \theta_0)$$

$$+ (D_T^{-1/2} C_T^{1/2})(C_T^{-1/2}(H_T(\tilde{\theta}) - H_T(\theta_0)) C_T^{-1/2}) C_T^{1/2}(\hat{\theta} - \theta_0).$$

Because $C_T^{1/2}(\hat{\theta} - \theta_0) = O_p(1)$ by assumption and because $C_T D_T^{-1} \to 0$, it follows that if

$$C_T^{-1/2}(H_T(\tilde{\theta}) - H_T(\theta_0)) C_T^{-1/2} = O_p(1),$$

the theorem is complete. Because $C_T^{1/2}|\tilde{\theta} - \theta_0| \leq 1$ with probability approaching to one as $T \to \infty$, the result follows from our assumption on $H_T(\cdot)$. □

Proof of Theorem 2:

Note that

$$P(m_T^{-1}|\hat{\theta} - \theta_0| > K)$$

$$\leq P(\inf_{\theta\in\Theta:|\theta-\theta_0|>Km_T} Q_T(\theta) \leq Q_T(\theta_0))$$

$$= P(\inf_{\theta\in\Theta:|\theta-\theta_0|>Km_T} \{2\sum_{t=1}^{T} \varepsilon_t(g(x_t,\theta_0) - g(x_t,\theta)) + \sum_{t=1}^{T}(g(x_t,\theta_0) - g(x_t,\theta))^2\} \leq 0)$$

$$\leq P(\inf_{\theta\in\Theta:|\theta-\theta_0|>Km_T} \{-2|\theta-\theta_0| \sup_{\theta\in\Theta} |\sum_{t=1}^{T} \varepsilon_t \dot{g}(x_t,\theta)| + |\theta-\theta_0|^2 \sum_{t=1}^{T} \inf_{\theta\in\Theta} |\dot{g}(x_t,\theta)|^2\} \leq 0)$$

$$= P(\inf_{\theta\in\Theta:|\theta-\theta_0|>Km_T} \{-2|\theta-\theta_0|B_T + |\theta-\theta_0|^2 A_T\} \leq 0).$$

The latter expression is minimal for $|\theta - \theta_0| = Km_T$ if $Km_T \geq B_T/A_T$. However

$$\lim_{K\to\infty} \limsup_{T\to\infty} P(Km_T \leq B_T/A_T) = 0$$

because $B_T/(m_T A_T) = O_p(1)$, implying that for all $K > 0$

$$P(m_T^{-1}|\hat{\theta} - \theta_0| > K)$$

$$\leq P(Km_T \leq B_T/A_T) + P(m_T^{-1}|\hat{\theta} - \theta_0| > K \cap Km_T > B_T/A_T)$$

$$\leq P(Km_T \leq B_T/A_T) + P(-2B_T Km_T + (Km_T)^2 A_T \leq 0)$$

$$= P(Km_T \leq B_T/A_T) + P(-2B_T + Km_T A_T \leq 0)$$

$$\leq 2P(B_T/(m_T A_T) \geq K/2).$$

The result now follows by taking limits as $T \to \infty$ and $K \to \infty$ respectively. □