

# Nonlinear models with integrated regressors and convergence order results\*

Robert de Jong<sup>†</sup>      Ling Hu<sup>‡</sup>

August 4, 2006

## Abstract

This paper discusses complications that arise in the theory of nonlinear estimation in the presence of integrated regressors. A high-level asymptotic distribution theorem is established and convergence rates for nonlinear least squares estimators are derived.

## 1 Introduction

Let  $Q_T(\cdot) : \mathbb{R}^p \rightarrow \mathbb{R}$  denote a random function and let  $\hat{\theta}$  denote a minimizer of this criterion function over a parameter space  $\Theta$ . 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  $(\text{tr}(M'M))^{1/2}$ . In the literature on nonlinear minimization estimators, it has been argued that for obtaining the standard limit theory for  $\hat{\theta}$  based on the asymptotic behavior of  $S_T(\theta_0)$  and  $H_T(\theta_0)$ , 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;

---

\*We thank Professors Phillips, Wooldridge and Jeganathan for discussions and suggestions. This version: May 2006.

<sup>†</sup>429 Arps Hall, Department of Economics, Ohio State University, 1945 N. High Street, Columbus 43210, USA, email [dejong@econ.ohio-state.edu](mailto:dejong@econ.ohio-state.edu)

<sup>‡</sup>403 Arps Hall, Department of Economics, Ohio State University, 1945 N. High Street, Columbus 43210, USA, email [linghu@econ.ohio-state.edu](mailto:linghu@econ.ohio-state.edu)

$$2. \quad \max_{\mathcal{N}_T^0} |C_T^{-1/2}(H_T(\theta_0) - H_T(\theta))C_T^{-1/2}| = o_p(1),$$

where

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

Cf. Park and Phillips (2001); the 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  $C_{Tii} \rightarrow \infty$  and  $D_{Tii} \rightarrow \infty$  for  $i = 1, \dots, p$ , and  $C_T D_T^{-1} \rightarrow 0$  as  $T \rightarrow \infty$ . 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.

In this paper, it is argued that the above conditions by itself are not sufficient for establishing the limit behavior of the global minimizer  $\hat{\theta}$ , and it is argued that in the absence of global convexity of the objective function, an assumption of the type

$$C_T^{1/2}(\hat{\theta} - \theta_0) = o_p(1) \tag{1}$$

needs to be added to our list of key assumptions in order to obtain limit theory for  $\hat{\theta}$ . This paper will give an example that illustrates the issue, suggests a route for modifying the reasoning by using the assumption of Equation (1), and discusses the implications for several results that rely on the above reasoning.

The above conditions are in fact sufficient for establishing that there exists a solution  $\theta_T^*$  such that  $(\partial/\partial\theta)Q_T(\theta_T^*) = 0$  a.s. and  $\theta_T^*$  is asymptotically normally distributed, as established in Wooldridge's (1994) Theorems 8.1 and 10.1. However, in a practical situation we may not know where this asymptotically normal solution is located, unless the objective function is globally convex. Therefore, in the general case where we do not have a global convexity property, a global argument is needed to ensure that the estimator will eventually end up in a  $D_T^{-1/2}$ -neighborhood of the true parameter, and this global argument is not guaranteed by the above conditions.

## 2 An example

Consider the following situation. Assume that the (observed)  $y_t$  is a mean zero i.i.d. sequence of random variables such that  $E y_t^2 = \sigma^2 = 1$  and  $E y_t^4 < \infty$ . Let  $\hat{\sigma}_T^2 = T^{-1} \sum_{t=1}^T (y_t - \bar{y}_T)^2$ . The objective function to be minimized is

$$Q_T(\theta) = \begin{cases} T^{-1} \sum_{t=1}^T (y_t - \theta)^2 - \hat{\sigma}_T^2 & \text{for } \theta \notin [2T^{-1/3}, 3T^{-1/3}] \\ P_T(\theta) - \hat{\sigma}_T^2 & \text{for } \theta \in [2T^{-1/3}, 3T^{-1/3}] \end{cases}$$

where  $P_T(\theta)$  is a function that attains its unique minimum of  $\sigma^2 = 1$  for  $\theta = 2.5T^{-1/3}$  and is such that  $Q_T(\cdot)$  is twice continuously differentiable everywhere, including at  $\theta = 2T^{-1/3}$  and at  $\theta = 3T^{-1/3}$ . Note that  $\bar{y}_T \leq 2T^{-1/3}$  a.s. as  $T \rightarrow \infty$ . This  $Q_T(\cdot)$  satisfies

$$\sup_{\theta \in [-1, 1]} |Q_T(\theta) - \theta^2| \xrightarrow{as} 0,$$

implying that we can only sensibly define  $\theta_0 = 0$ . A picture of this situation is given in Appendix A.

For  $\theta \notin [2T^{-1/3}, 3T^{-1/3}]$ ,  $S_T(\cdot)$  and  $H_T(\cdot)$  equal

$$S_T(\theta) = -2 \sum_{t=1}^T (y_t - \theta)$$

and

$$H_T(\theta) = 2T,$$

implying that for  $D_T = T$ ,

$$D_T^{-1/2} S_T(\theta_0) = -2T^{-1/2} \sum_{t=1}^T y_t \xrightarrow{d} N(0, 4)$$

and

$$D_T^{-1} H_T(\theta_0) = 2.$$

Therefore for any sequence  $C_T^{-1/2} \leq 2T^{-1/3}$  such that  $C_T^{-1/2} = o(1)$ , the condition

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

where

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

will hold. Therefore, one might be tempted to conclude that in this case the minimization estimator  $\hat{\theta}$  satisfies

$$T^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, 1).$$

However,

$$P(\hat{\theta} = 2.5T^{-1/3}) = P(1 - T^{-1} \sum_{t=1}^T (y_t - \bar{y}_T)^2 < 0) \rightarrow 0.5$$

where the last result follows from a standard application of the central limit theorem. This illustrates that more structure is needed than outlined in conditions 1. and 2. of the Introduction before asymptotic normality can be obtained.

In the above situation, there does in fact exist an asymptotically normal solution to the equation  $(\partial/\partial\theta)Q_T(\theta) = 0$ , as asserted in Wooldridge's (1994) Theorems 8.1 and 10.1; this solution will be the sample average  $\bar{y}_T$ . However, this solution is not necessarily equal to the global minimum, and in a practical situation, we have no means of deciding which solution to choose, other than using the global minimum.

### 3 A limit theorem for minimization estimators

This section establishes the result alluded to in the Introduction:

**Theorem 1** *Let  $\{Q_T : \mathcal{W} \times \Theta \rightarrow \mathbb{R}, T = 1, 2, \dots\}$  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.  $\Theta$  is compact and convex and  $\theta_0 \in \text{int}(\Theta)$ ;
2.  $Q_T(\cdot, \cdot)$  satisfies the standard measurability and second order differentiability conditions on  $\mathcal{W} \times \Theta$ ,  $T = 1, 2, \dots$ ;
3. For sequences of positive definite diagonal matrices  $C_T$  and  $D_T$  such that  $C_T D_T^{-1} \rightarrow 0$  as  $T \rightarrow \infty$  and  $C_{Tii} \rightarrow \infty$  for all  $i = 1, \dots, p$ ,

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

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 paper are given in Appendix B.

The above result means that wherever the reasoning that we argued is incomplete 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 normality for the global minimum. However, this condition is a strengthened version of a consistency result, and is not necessarily easy to obtain. In an analysis such as Pollard's (1985), obtaining the correct rate for the minimization estimator is the bulk of the analytical problem. In view of the simplicity of the derivation of Theorem 1, one might argue that obtaining the rate result is essentially the key towards obtaining a limit distribution result for the minimization estimator.

The theorem proven above is (correctly) incapable of showing the limit distribution of the example of Section 2 of this paper. This is because for that example,  $C_T^{1/2}(\hat{\theta} - \theta_0) = o_p(1)$  only if  $C_T$  is chosen as  $T^{2/3-\delta}$  for some small  $\delta > 0$ . However, for such a choice of  $C_T$ , Assumption 1.3 will fail.

## 4 Nonlinear nonstationary least squares

As in Park and Phillips (2001), in this section we consider the nonlinear regression model

$$y_t = g(x_t, \theta) + \varepsilon_t \tag{2}$$

where  $x_t$  is an integrated process and  $\varepsilon_t$  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

$$B_T = \sup_{\theta \in \Theta} \left| \sum_{t=1}^T \varepsilon_t \dot{g}(x_t, \theta) \right|.$$

If  $m_T$  is a strictly positive deterministic sequence such that  $m_T \rightarrow 0$  and  $B_T/(m_T A_T) = O_p(1)$ , then

$$m_T^{-1}(\hat{\theta} - \theta_0) = O_p(1).$$

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

It was pointed out to us by Professor Phillips that Professor Jeganathan and himself are working on general results on convergence rates for nonlinear least squares estimators. It appears from the preprint that Professor Jeganathan kindly shared with us that the results obtained by these authors might be substantially more general and complicated than the simple approach outlined here. For example, the case where  $\dot{g}(x, \theta)$  is an integrable function that attains the value of 0 for some choice of  $\theta$  is ruled out by the assumptions of the section below, but appears to be allowed by the results of these authors. Also, objective functions that are more general than nonlinear least squares are dealt with in their work.

## 4.1 The integrable function case

If  $\dot{g}(x, \cdot)$  is integrable over  $x$  for any  $\theta$ , it is not very restrictive to assume that  $\inf_{\theta \in \Theta} |\dot{g}(x_t, \theta)|^2$  is integrable as well, given the compactness of  $\Theta$ .

Therefore, the following assumption does not add significantly to the set of conditions 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

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

Given this assumption only, 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.

It should be noted that while  $\inf_{\theta \in \Theta} |\dot{g}(x_t, \theta)|^2$  can in principle be zero in some cases, the condition above is not overly restrictive. For example, in the case  $g(x, \theta) = \Phi(\theta x_t)$ , which one may consider as a nonlinear least squares alternative to probit estimation, we get  $\dot{g}(x, \theta) = x_t \phi(\theta x_t)$ , implying that for  $\Theta = [-1, 1]$ , say, we have  $\inf_{\theta \in \Theta} |\dot{g}(x_t, \theta)|^2 = x_t^2 \phi(-|x_t|)^2$ . This last function satisfies the requirements of the above theorem.

## 4.2 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  $\lambda \rightarrow \infty$ ,

$$\dot{g}(\lambda x, \theta) \approx \dot{\nu}(\lambda) \dot{H}(x, \theta).$$

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

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

**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 Phillips (2001) for the asymptotically homogeneous case also hold for the global minimizer of the objective function.

## 5 Nonstationary probit and nonstationary ordered discrete choice

Nonstationary probit was considered in Park and Phillips (2000). Because the probit objective function is globally concave, a solution to the equation  $(\partial/\partial\theta)Q_T(\theta) = 0$  will also be the global maximum. Therefore, showing the asymptotic properties of an asymptotically normal solution to  $(\partial/\partial\theta)Q_T(\theta) = 0$  suffices for analyzing the global maximizer of the loglikelihood, and the issues discussed in Sections 1 and 2 of this paper do not occur.

Nonstationary discrete choice models were considered in Hu and Phillips (2004) and Phillips, Jin and Hu (2005). For the ordered probit model, the loglikelihood is concave in the index parameter  $(\beta', \mu)'$ ; see Pratt (1981) for a proof of this result. Therefore, it follows that the analysis of the above two papers is complete as it stands.

## References

- 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, Ams-

terdam.

## Appendix A

A picture of the situation in the example of Section 2 of this paper looks as follows.

$$Q_T(\theta)$$

$$\sigma^2 - \hat{\sigma}_T^2 \qquad \bar{y}_T \qquad 2T^{-1/3} \qquad 3T^{-1/3} \qquad \theta \rightarrow$$

## Appendix B: Mathematical Proofs

### Proof of Theorem 1:

Because  $C_{Tii} \rightarrow \infty$  as  $T \rightarrow \infty$  for  $i = 1, \dots, p$ ,  $\hat{\theta}$  will be in the interior of  $\Theta$  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,

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

Because  $C_T^{1/2}(\hat{\theta} - \theta_0) = o_p(1)$  by assumption and because  $C_T D_T^{-1} \rightarrow 0$ , 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  $T \rightarrow \infty$ , the result follows from our assumption on  $H_T(\cdot)$ .  $\square$

### Proof of Theorem 2:

Note that

$$P(m_T^{-1} |\hat{\theta} - \theta_0| > K) \\ \leq P\left(\inf_{\theta \in \Theta: |\theta - \theta_0| > Km_T} Q_T(\theta) \leq Q_T(\theta_0)\right) \\ = P\left(\inf_{\theta \in \Theta: |\theta - \theta_0| > Km_T} \left\{2 \sum_{t=1}^T \varepsilon_t (g(x_t, \theta_0) - g(x_t, \theta)) + \sum_{t=1}^T (g(x_t, \theta_0) - g(x_t, \theta))^2\right\} \leq 0\right) \\ \leq P\left(\inf_{\theta \in \Theta: |\theta - \theta_0| > Km_T} \left\{-2|\theta - \theta_0| \sup_{\theta \in \Theta} \left|\sum_{t=1}^T \varepsilon_t \dot{g}(x_t, \theta)\right| + |\theta - \theta_0|^2 \sum_{t=1}^T \inf_{\theta \in \Theta} |\dot{g}(x_t, \theta)|^2\right\} \leq 0\right) \\ = P\left(\inf_{\theta \in \Theta: |\theta - \theta_0| > Km_T} \{-2|\theta - \theta_0| B_T + |\theta - \theta_0|^2 A_T\} \leq 0\right).$$

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

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

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  $T \rightarrow \infty$  and  $K \rightarrow \infty$  respectively.  $\square$