Nonlinear regressions with nonstationary time series

Nigel Chan and Qiying Wang*

The University of Sydney

March 20, 2014

Abstract

This paper develops asymptotic theory for a nonlinear parametric cointegrating regression model. We establish a general framework for weak consistency that is easy to apply for various nonstationary time series, including partial sums of linear processes and Harris recurrent Markov chains. We provide limit distributions for nonlinear least squares estimators, extending the previous works. We also introduce endogeneity to the model by allowing the error to be serially dependent on and cross correlated with the regressors.

JEL Classification: C13, C22.

Key words and phrases: Cointegration, nonlinear regressions, consistency, limit distribution, nonstationarity, nonlinearity, endogeneity.

1 Introduction

The past few decades have witnessed significant developments in cointegration analysis. In particular, extensive researches have focused on cointegration models with linear structure. Although it gives data researchers convenience in implementation, the linear structure is often too restrictive. In particular, nonlinear responses with some unknown parameters often arise in the context of economics. For empirical examples, we refer to Granger and Teräsvirta (1993) as well as Teräsvirta et al. (2011). In this situation, it is expected that nonlinear cointegration captures the features of many long-run relationships in a more realistic manner.

*The author thanks to the Editor, Associate Editor and three referees for their very helpful comments on the original version. The authors also acknowledge research support from Australian Research Council.
Address correspondence to Qiying Wang: School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia. E-mail: qiying@maths.usyd.edu.au

A typical nonlinear parametric cointegrating regression model has the form

$$y_t = f(x_t, \theta_0) + u_t, \quad t = 1, ..., n, \tag{1}$$

where $f : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}$ is a known nonlinear function, x_t and u_t are regressor and regression errors, and θ_0 is an m-dimensional true parameter vector that lies in the parameter set Θ. With the observed nonstationary data $\{y_t, x_t\}_{t=1}^n$, this paper is concerned with the nonlinear least squares (NLS) estimation of the unknown parameters $\theta_0 \in \Theta$. In this regard, Park and Phillips (2001) (PP henceforth) considered x_t to be an integrated, $I(1)$, process. Based on PP framework, Chang et al. (2001) introduced an additional linear time trend term and stationary regressors into model (1). Chang and Park (2003) extended to nonlinear index models driven by integrated processes. More recently, Choi and Saikkonen (2010), Gao et al. (2009) and Wang and Phillips (2012) developed statistical tests for the existence of a nonlinear cointegrating relation. Park and Chang (2011) allowed the regressors x_t to be contemporaneously correlated with the regression errors u_t and Shi and Phillips (2012) extended the model (1) by incorporating a loading coefficient.

The present paper has a similar goal to the previously mentioned papers but offers more general results, which have some advantages for empirical studies. First of all, we establish a general framework for weak consistency of the NLS estimator $\hat{\theta}_n$, allowing for the x_t to be a wider class of nonstationary time series. The set of sufficient conditions is easy to apply for various nonstationary regressors, including partial sums of linear processes and recurrent Markov chains. Furthermore, we provide limit distributions for the NLS estimator $\hat{\theta}_n$. It deserves to mention that the routine employed in this paper to establish the limit distributions of $\hat{\theta}_n$ is different from those used in the previous works, e.g. Park and Phillips (2001). Roughly speaking, our routine is related to joint distributional convergence of a martingale under target and its conditional variance, rather than using classical martingale limit theorem which requires establishing the convergence in probability for the conditional variance. In nonlinear cointegrating regressions, there are some advantages for our methodology since it is usually difficult to establish the convergence in probability for the conditional variance, in particular, in the situation that the regressor x_t is a nonstationary time series. Second, in addition to the commonly used martingale innovation structure, our model allows for serial dependence in the equilibrium errors u_t and the innovations driving x_t. It is important as our model permits joint determination of x_t and y_t, and hence the system is a time series structural model. Under such situation, the weak consistency and limit distribution of the NLS estimator $\hat{\theta}_n$ are also established.

This paper is organized as follow. Section 2 presents our main results on weak consistency of the NLS estimator $\hat{\theta}_n$. Theorem 2.1 provides a general framework. Its applications to integrable and non-integrable f are given in Theorems 2.2–2.5, respectively. Section 3 investigates the limit distributions of $\hat{\theta}_n$ in which the model (1) has a martingale structure. Extension to endogeneity is presented in Section 4. As mentioned above, our routine establishing the limit distribution of $\hat{\theta}_n$ is different from previous works. Section 5 performs simulation, reporting and discussing the numerical values of means and standard errors which provide the evidence of accuracies of our NLS estimator. Section 6 presents an empirical example, providing a link between our theory and real applications. The model of interest is the carbon Kuznets curve relating the per capita CO_2 emissions and per capita GDP. Endogeneity occurs in this example due to potential misreporting of GDP, omitted variable bias and reverse causality. Section 7 concludes the paper. Section 8 provides partial technical proofs. Full details of the technical proofs can be found in the Supplemental Material of this paper, where we also provide a unit root test for our empirical example, and other details for simulation.

Throughout the paper, we denote constants by $C, C_1, C_2, ...$ which may be different at each appearance. For a vector $x = (x_1, ..., x_m)$, assume that $\|x\| = (x_1^2 + ... + x_m^2)^{1/2}$, and for a matrix A, the norm operator $\|\cdot\|$ is defined by $\|A\| = \sup_{x:\|x\|=1} \|xA\|$. Furthermore, the parameter set $\Theta \subset \mathbb{R}^m$ is assumed to be compact and convex, and the true parameter vector θ_0 is an interior point of Θ.

2 Weak consistency

This section considers the estimation of the unknown parameters θ_0 in model (1) by NLS. Let $Q_n(\theta) = \sum_{t=1}^n (y_t - f(x_t, \theta))^2$. The NLS estimator $\hat{\theta}_n$ of θ_0 is defined to be the minimizer of $Q_n(\theta)$ over $\theta \in \Theta$, that is,

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} Q_n(\theta), \tag{2}$$

and the error estimator is defined by $\hat{\sigma}_n^2 = n^{-1} \sum_{t=1}^n \hat{u}_t^2$, where $\hat{u}_t = y_t - f(x_t, \hat{\theta}_n)$. To investigate the weak consistency for the NLS estimator $\hat{\theta}_n$, this section assumes the regression model (1) to have a martingale structure. In this situation, our sufficient conditions are closely related to those of Wu (1981), Lai (1994) and Skouras (2000), intending to provide a general framework. In comparison to the papers mentioned, our assumptions are easy to apply, particularly in nonlinear cointegrating regression situation as stated in the two examples below. Extension to endogeneity between x_t and u_t is investigated in Section 4.

2.1 A framework

We make use of the following assumptions for the development of weak consistency.

Assumption 2.1. For each $\pi, \pi_0 \in \Theta$, there exists a real function $T : \mathbb{R} \to \mathbb{R}$ such that

$$|f(x, \pi) - f(x, \pi_0)| \le h(||\pi - \pi_0||)\, T(x), \tag{3}$$

where $h(x)$ is a bounded real function such that $h(x) \downarrow h(0) = 0$, as $x \downarrow 0$.

Assumption 2.2. (i) $\{u_t, \mathcal{F}_t, 1 \le t \le n\}$ is a martingale difference sequence satisfying $E(|u_t|^2|\mathcal{F}_{t-1}) = \sigma^2$ and $\sup_{1\le t\le n} E(|u_t|^{2q}|\mathcal{F}_{t-1}) < \infty$ a.s., where $q > 1$; and
(ii) x_t is adapted to $\mathcal{F}_{t-1}$, $t = 1, ..., n$.

Assumption 2.3. There exists an increasing sequence $0 < \kappa_n \to \infty$ such that

$$\kappa_n^{-2} \sum_{t=1}^{n} [T(x_t) + T^2(x_t)] = O_P(1), \tag{4}$$

and for any $0 < \eta < 1$ and $\theta \ne \theta_0$, where $\theta, \theta_0 \in \Theta$, there exist $n_0 > 0$ and $M_1 > 0$ such that

$$P\Big(\sum_{t=1}^{n} (f(x_t, \theta) - f(x_t, \theta_0))^2 \ge \kappa_n^2 / M_1\Big) \ge 1 - \eta, \tag{5}$$

for all $n > n_0$.

Theorem 2.1. *Under Assumptions 2.1–2.3, the NLS estimator $\hat\theta_n$ is a consistent estimator of θ_0, i.e. $\hat\theta_n \to_P \theta_0$. If in addition $\kappa_n^2 n^{-1} = O(1)$, then $\hat\sigma_n^2 \to_P \sigma^2$, as $n \to \infty$.*

Assumptions 2.1 and 2.2 are the same as those used in Skouras (2000), which are standard in the NLS estimation theory. Also see Wu (1981) and Lai (1994). Assumption 2.3 is used to replace (3.8), (3.9) and (3.11) in Skouras (2000), in which some uniform conditions are used. In comparison to Skouras (2000), our Assumption 2.3 is related to the conditions on the regressor x_t and is more natural and easy to apply. In particular, it is directly applicable in the situation that T is integrable and the regressor x_t is a nonstationary time series, as stated in the following sub-section.

2.2 Assumption 2.3: integrable functions

Due to Assumption 2.1, $f(x,\theta)-f(x,\theta_0)$ is integrable in x if T is an integrable function. This class of functions includes $f(x,\theta_1,\theta_2)=\theta_1|x|^{\theta_2}I(x\in[a,b])$, where a and b are finite constants, the Gaussian function $f(x,\theta)=e^{-\theta x^2}$, the Laplacian function $f(x,\theta)=e^{-\theta|x|}$, the logistic regression function $f(x,\theta)=e^{\theta|x|}/(1+e^{\theta|x|})$, etc. In this sub-section, two commonly used non-stationary regressors x_t are shown to satisfy Assumption 2.3 if T is integrable.

Example 1 (Partial sums of linear processes). Let $x_t=\sum_{j=1}^t \xi_j$, where $\{\xi_j, j\geq 1\}$ is a linear process defined by

$$\xi_j=\sum_{k=0}^{\infty}\phi_k\,\epsilon_{j-k}, \tag{6}$$

where $\{\epsilon_j, -\infty<j<\infty\}$ is a sequence of i.i.d. random variables with $E\epsilon_0=0$, $E\epsilon_0^2=1$ and the characteristic function $\varphi(t)$ of ϵ_0 satisfying $\int_{-\infty}^{\infty}|\varphi(t)|dt<\infty$. The coefficients ϕ_k are assumed to satisfy one of the following conditions:

C1. $\phi_k\sim k^{-\mu}\rho(k)$, where $1/2<\mu<1$ and $\rho(k)$ is a function slowly varying at ∞.

C2. $\sum_{k=0}^{\infty}|\phi_k|<\infty$ and $\phi\equiv\sum_{k=0}^{\infty}\phi_k\neq 0$.

Put $d_n^2=\mathbb{E}x_n^2$. As in Wang, Lin and Gulati (2003), we have

$$d_n^2=\mathbb{E}x_n^2\sim\begin{cases}c_\mu n^{3-2\mu}\rho^2(n), & \text{under } \mathbf{C1},\\ \phi^2 n, & \text{under } \mathbf{C2},\end{cases} \tag{7}$$

where $c_\mu=1/((1-\mu)(3-2\mu))\int_0^\infty x^{-\mu}(x+1)^{-\mu}dx$. We have the following result.

Theorem 2.2. *Suppose x_t is defined as in Example 1 and Assumption 2.1 holds. Assume:*

(i) T is bounded and integrable, and

(ii) $\int_{-\infty}^{\infty}(f(s,\theta)-f(s,\theta_0))^2ds>0$ for all $\theta\neq\theta_0$.

Then (4) and (5) hold with $\kappa_n^2=n/d_n$. Consequently, if in addition Assumption 2.2, then $\hat{\theta}_n\to_P\theta_0$.

Theorem 2.2 improves Theorem 4.1 of PP in two folds. Firstly, we allow for more general regressor. The result under **C1** is new, which allows x_t to be long memory process, including the fractionally integrated process as an example. PP only allow x_t to satisfy **C2** with additional conditions on ϕ_k, that is, they require x_t to be a partial sums of a short memory process. Secondly, we remove the part (b) required in the definition of an I-regular function given in their Definition 3.3. Furthermore, we allow for certain non-integrable f such as the logistic regression function $f(x,\theta) = e^{\theta|x|}/(1+e^{\theta|x|})$, $\theta \geq \theta_0$ for some $\theta_0 > 0$, although we assume the integrability of T.

Example 2 (Recurrent Markov Chain). Let $\{x_k\}_{k\geq 0}$ be a Harris recurrent Markov chain with state space $(E,\mathcal{E})$, transition probability $P(x,A)$ and invariant measure π. We denote P_μ for the Markovian probability with the initial distribution μ, E_μ for correspondent expectation and $P^k(x,A)$ for the k-step transition of $\{x_k\}_{k\geq 0}$. A subset D of E with $0 < \pi(D) < \infty$ is called D-set of $\{x_k\}_{k\geq 0}$ if for any $A \in \mathcal{E}^+$,

$$\sup_{x\in E} E_x\Big(\sum_{k=1}^{\tau_A} I_D(x_k)\Big) < \infty,$$

where $\mathcal{E}^+ = \{A \in \mathcal{E} : \pi(A) > 0\}$ and $\tau_A = \inf\{n \geq 1 :\ x_n \in A\}$. By Theorem 6.2 of Orey (1971), D-sets not only exist, but generate the entire sigma $\mathcal{E}$, and for any D-sets C, D and any probability measure ν, μ on $(E,\mathcal{E})$,

$$\lim_{n\to\infty} \sum_{k=1}^{n} \nu P^k(C) / \sum_{k=1}^{n} \mu P^k(D) \;=\; \frac{\pi(C)}{\pi(D)}, \tag{8}$$

where $\nu P^k(D) = \int_{-\infty}^{\infty} P^k(x,D)\nu(dx)$. See Nummelin (1984) for instance.

Let a D-set D and a probability measure ν on $(E,\mathcal{E})$ be fixed. Define

$$a(t) \;=\; \pi^{-1}(D) \sum_{k=1}^{[t]} \nu P^k(D), \quad t \geq 0.$$

By recurrence, $a(t) \to \infty$. Here and below, we set the state space to be the real space, that is $(E,\mathcal{E}) = (R,\mathcal{R})$. We have the following result.

Theorem 2.3. *Suppose x_t is defined as in Example 2 and Assumption 2.1 holds. Assume:*

(i) T is bounded and $\int_{-\infty}^{\infty} |T(x)|\pi(dx) < \infty$, and

(ii) $\int_{-\infty}^{\infty} (f(s,\theta) - f(s,\theta_0))^2 \pi(ds) > 0$ for all $\theta \neq \theta_0$.

Then (4) and (5) hold with $\kappa_n^2 = a(n)$. Consequently, if in addition Assumption 2.2, then $\hat{\theta}_n \to_P \theta_0$.

Theorem 2.3 seems to be new to the literature. By virtue of (8), the asymptotic order of $a(t)$ depends only on $\{x_k\}_{k\geq 0}$. It is interesting to notice that Theorem 2.3 does not impose the β-regular condition as commonly used in the literature. The Harris recurrent Markov chain $\{x_k\}_{k\geq 0}$ is called β-regular if

$$\lim_{\lambda\to\infty} a(\lambda t)/a(\lambda) \;=\; t^\beta, \quad \forall t > 0, \tag{9}$$

where $0 < \beta \leq 1$. See Chen (1999) for instance.

2.3 Assumption 2.3: beyond integrable functions

As noticed in Section 2.2, Assumption 2.3 holds for certain non-integrable f although we require the integrability of T. Assumption 2.3 is more involved for general non-integrable f when T is also non-integrable. In the latter situation, we require certain relationships between f and T.

Assumption 2.4. (i) There exists a regular function[1] $g(x, \theta, \theta_0)$ on Θ satisfying

$$\int_{|s|\le\delta} g^2(s, \theta, \theta_0)ds > 0$$

for all $\theta \neq \theta_0$ and $\delta > 0$, a real function $T : \mathbb{R} \to \mathbb{R}$ and a positive real function $v(\lambda)$ which is bounded away from zero as $\lambda \to \infty$, such that for any bounded x,

$$\sup_{\theta,\theta_0\in\Theta} |f(\lambda x, \theta) - f(\lambda x, \theta_0) - v(\lambda)\, g(x, \theta, \theta_0)|/T(\lambda x) \;=\; o(1), \qquad (10)$$

as $\lambda \to \infty$.

(ii) There exists a real function $T_1 : \mathbb{R} \to \mathbb{R}$ such that $T(\lambda x) \le v(\lambda)\, T_1(x)$ as $|\lambda x| \to \infty$ and $T_1 + T_1^2$ is locally integrable (i.e. integrable on any compact set).

Theorem 2.4. *Suppose Assumption 2.4 holds. Suppose that there exists a continuous Gaussian process $G(t)$ such that $x_{[nt],n} \Rightarrow G(t)$, on $D[0, 1]$, where $x_{i,n} = x_i/d_n$ and $0 < d_n \to \infty$ is a sequence of real numbers. Then (4) and (5) hold with $\kappa_n^2 = nv^2(d_n)$. Consequently, if in addition Assumptions 2.1–2.2 and assumption that the Ts in Assumptions 2.1 and 2.4 coincide, then $\hat{\theta}_n \to_P \theta_0$.*

Assumptions 2.1 and 2.4 are quite general, including many commonly used regression functions. Typical examples include $f(x, \theta) = (x + \theta)^2$, $\theta e^x/(1 + e^x)$, $\theta \log |x|$, $\theta|x|^\alpha$ (α is fixed) and $\theta_0 + \theta_1|x| + ... + \theta_k|x|^k$. The class of functions satisfying Assumptions 2.1 and 2.4 are similar to, but wider than those imposed in Theorem 4.2 of PP. For instance, Assumptions 2.1 and 2.4 (hence Theorem 2.4) are applicable for the function $f(x, \theta) = (x + \theta)^2$ [with $T(x) = T_1(x) = |x|$, $v(\lambda) = \lambda$ and $g(x, \theta, \theta_0) = 2(\theta - \theta_0)\, x$], but Theorem 4.2 of PP is not directly applicable for this function. See, e.g., Example 4.1 (c) of PP. We also mention that our condition on the regressor x_t is much more general than

[1]Function H is called regular on Θ if (a) for all $\theta \in \Theta$, there exist for each $\epsilon > 0$ continuous functions $\underline{H}_\epsilon$, $\overline{H}_\epsilon$, and a constant $\delta_\epsilon > 0$ such that $\underline{H}_\epsilon(x, \theta) \le H(y, \theta) \le \overline{H}_\epsilon(x, \theta)$ for all $|x - y| < \delta_\epsilon$ on K, a compact set of R, and such that $\int_K (\overline{H}_\epsilon - \underline{H}_\epsilon)(x, \theta)dx \to 0$ as $\epsilon \to 0$, and (b) for all $x \in R$, $H(x, \cdot)$ is equicontinuous in a neighborhood of x.

that of PP, as we only require that $x_{[nt]}/d_n$ converges weakly to a continuous Gaussian process. This kind of weak convergence condition is very likely close to be necessary.

We next consider a class of asymptotically homogeneous functions. In this regard, we follow PP. Let $f : \mathbb{R} \times \Theta \to \mathbb{R}$ have the structure:

$$f(\lambda x, \theta) = v(\lambda, \theta) h(x, \theta) + b(\lambda, \theta)\, A(x, \theta)\, B(\lambda x, \theta), \tag{11}$$

where $\sup_{\theta\in\Theta} |b(\lambda, \theta)\, v^{-1}(\lambda, \theta)| \to 0$, as $\lambda \to \infty$; $\sup_{\theta\in\Theta} |A(x, \theta)|$ is locally bounded, that is, bounded on bounded intervals; $\sup_{\theta\in\Theta} |B(\lambda x, \theta)|$ is bounded on R; $h(x, \theta)$ is regular on Θ satisfying $\int_{|s|\le\delta} h^2(s, \theta) ds > 0$ for all $\theta \ne \theta_0$ and $\delta > 0$, and $v(\lambda, \theta)$ satisfy: there exist $\epsilon > 0$ and a neighborhood N of $\bar\theta$ such that as $\lambda \to \infty$

$$\inf_{\substack{|p-\bar p|<\epsilon \\ |q-\bar q|<\epsilon}} \inf_{\theta\in N} |pv(\lambda, \theta) - qv(\lambda, \theta_0)| \to \infty, \tag{12}$$

for any $\bar\theta \ne \theta_0$ and $\bar p, \bar q > 0$.

Theorem 2.5. *Suppose that f in model (1) has the structure (11), and in addition to Assumption 2.2, there exists a continuous Gaussian process $G(t)$ such that $x_{[nt],n} \Rightarrow G(t)$ on $D[0, 1]$, where $x_{i,n} = x_i/d_n$ and $0 < d_n \to \infty$ is a sequence of real numbers. Then, the NLS estimator $\hat\theta_n$ defined by (2) is a consistent estimator of θ_0, i.e. $\hat\theta_n \to_P \theta_0$.*

The conditions on f given in Theorem 2.5 are the same as those used in Theorem 4.3 of PP, which are satisfied by the functions such as the Box-Cox transformation $(|x|^\theta - 1)/\theta$. The difference between current Theorem 2.5 and Theorem 4.3 of PP is that we only require that $x_{[nt]}/d_n$ converges weakly to a continuous Gaussian process, which is close to be necessary.

To investigate the weak consistency of $\hat\theta_n$, this section makes use of three different sets of conditions on f: Assumption 2.1 with T being integrable (Theorems 2.2–2.3); Assumptions 2.1 and 2.4 (Theorem 2.4) and the condition (11) (Theorem 2.5). We remark that these condition sets are mutually exclusive. There are examples of the function f such that one of these assumptions holds, but not other two. For instance, the function $f(x, \theta) = \theta/(1 + x^2)$ satisfies Assumption 2.1 with $h(x) = x$ and $T(x) = 1/(1 + x^2)$, but Assumption 2.4 and the condition (11) fail; the function $f(x, \theta) = (x + \theta)^2$ satisfies Assumptions 2.1 and 2.4, but it does not satisfy condition (11).

3 Limit distribution

This section considers limit distribution of $\hat{\theta}_n$. In what follows, let $\dot{Q}_n$ and $\ddot{Q}_n$ be the first and second derivatives of $Q_n(\theta)$ in the usual way, that is, $\dot{Q}_n = \partial Q_n/\partial\theta$ and $\ddot{Q}_n = \partial^2 Q_n/\partial\theta\partial\theta'$. Similarly we define $\dot{f}$ and $\ddot{f}$. We assume these quantities exist whenever they are introduced. The following result comes from an application of Lemma 1 in Andrews and Sun (2004), which provides a framework and plays a key part in our development on limit distribution of $\hat{\theta}_n$.

Theorem 3.1. *There exists a sequence of $m\times m$ nonrandom nonsingular matrices D_n with $\| D_n^{-1} \| \to 0$, as $n\to\infty$, such that*

(i) $\sup_{\theta:\|D_n(\theta-\theta_0)\|\le k_n} \ \| (D_n^{-1})' \sum_{t=1}^n \big[\dot{f}(x_t,\theta)\dot{f}(x_t,\theta)' - \dot{f}(x_t,\theta_0)\dot{f}(x_t,\theta_0)'\big] D_n^{-1} \| = o_P(1)$,

(ii) $\sup_{\theta:\|D_n(\theta-\theta_0)\|\le k_n} \ \| (D_n^{-1})' \sum_{t=1}^n \ddot{f}(x_t,\theta) \big[f(x_t,\theta) - f(x_t,\theta_0)\big] D_n^{-1} \| = o_P(1)$,

(iii) $\sup_{\theta:\|D_n(\theta-\theta_0)\|\le k_n} \ \| (D_n^{-1})' \sum_{t=1}^n \ddot{f}(x_t,\theta)\, u_t\, D_n^{-1} \| = o_P(1)$,

(iv) $Y_n := (D_n^{-1})' \sum_{t=1}^n \dot{f}(x_t,\theta_0)\dot{f}(x_t,\theta_0)'\, D_n^{-1} \to_D M$, *where $M>0$, a.s., and*

$$Z_n := (D_n^{-1})' \sum_{t=1}^n \dot{f}(x_t,\theta_0)\, u_t \ = O_P(1),$$

for some sequence of constants $\{k_n, n\ge 1\}$ for which $k_n\to\infty$, as $n\to\infty$. Then, there exists a sequence of estimators $\{\hat{\theta}_n, n\ge 1\}$ satisfying $\dot{Q}_n(\hat{\theta}_n)=0$ with probability that goes to one and

$$D_n(\hat{\theta}_n - \theta_0) = Y_n^{-1} Z_n + o_P(1). \tag{13}$$

If we replace (iv) by the following (iv)', then $D_n(\hat{\theta}_n-\theta_0)\to_D M^{-1}Z$.

(iv)' for any $\alpha_i' = (\alpha_{i1},...,\alpha_{im})\in R^m, i=1,2,3,$

$$(\alpha_1' Y_n \alpha_2, \quad \alpha_3' Z_n) \ \to_D \ (\alpha_1' M \alpha_2, \ \alpha_3' Z),$$

where $M>0$, a.s. and $P(Z<\infty)=1$.

Our routine in establishing the limit distribution of $\hat{\theta}_n$ is essentially different from that of PP, and is particularly convenient in the situation that f, $\dot{f}$ and $\ddot{f}$ all are integrable in which $D_n = \kappa_n \mathbf{I}$ for some $0<\kappa_n\to\infty$, where $\mathbf{I}$ is an identity matrix. See next section for examples. The condition AD3 of PP requires to show $Y_n \to_P M$ instead of (iv). It is equivalent to say that, under the PP's routine, one requires to show (at least under an enlarged probability space)

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n g(x_t) \to_P \int_{-\infty}^{\infty} g(s)ds\, L_W(1,0), \tag{14}$$

where x_t is an integrated process, g is a real integrable function and $L_W(t,s)$ is the local time of the standard Brownian Motion $W(t)$[2]. The convergence in probability is usually hard or impossible to establish without enlarging the probability space. Our routine essentially reduces the convergence in probability to less restrictive convergence in distribution. Explicitly, in comparison to (14), we only need to show that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g(x_t) \to_D \int_{-\infty}^{\infty} g(s)ds\, L_W(1,0). \tag{15}$$

This allows to extend the nonlinear regressions to a much wider class of nonstationary regressor data series, and enables our methodology and proofs straightforward and neat.

3.1 Limit distribution: integrable functions

We now establish our results on convergence in distribution for the $\hat{\theta}_n$, which mainly involves settling the conditions (i)–(iii) and (iv)' in Theorem 3.1. First consider the situation that f is an integrable function, together with some additional conditions on x_t and u_t.

Assumption 3.1. (i) x_t is defined as in Example 1, that is, $x_t = \sum_{j=1}^{t}\xi_j$, where ξ_j satisfies (6);

(ii) $\mathcal{F}_k$ is a sequence of increasing σ-fields such that $\epsilon_k \in \mathcal{F}_k$ and ϵ_{k+1} is independent of $\mathcal{F}_k$ for all $k \geq 1$, and $\epsilon_k \in \mathcal{F}_1$ for all $k \leq 0$;

(iii) $\{u_k, \mathcal{F}_k\}_{k\geq 1}$ forms a martingale difference satisfying $\max_{k\geq m} |E(u_{k+1}^2 \mid \mathcal{F}_k) - 1| \to 0$ a.s., and for some $\delta > 0$, $\max_{k\geq 1} E(|u_{k+1}|^{2+\delta} \mid \mathcal{F}_k) < \infty$.

Assumption 3.2. Let $p(x,\theta)$ be one of f, $\dot{f}_i$ and $\ddot{f}_{ij}$, $1 \leq i,j \leq m$.

(i) $p(x,\theta_0)$ is a bounded and integrable real function;

(ii) $\Sigma = \int_{-\infty}^{\infty} \dot{f}(s,\theta_0)\dot{f}(s,\theta_0)'ds > 0$ and $\int_{-\infty}^{\infty}(f(s,\theta) - f(s,\theta_0))^2 ds > 0$ for all $\theta \neq \theta_0$;

(iii) There exists a bounded and integrable function $T_p : \mathbb{R} \to \mathbb{R}$ such that $|p(x,\theta) - p(x,\theta_0)| \leq h_p(||\theta - \theta_0||)\, T_p(x)$, for each $\theta, \theta_0 \in \Theta$, where $h_p(x)$ is a bounded real function such that $h_p(x) \downarrow h_p(0) = 0$, as $x \downarrow 0$.

[2]Here and below, the process $\{L_\zeta(t,s), t \geq 0, s \in R\}$ is said to be the local time of a measurable process $\{\zeta(t), t \geq 0\}$ if, for any locally integrable function $T(x)$,

$$\int_0^t T[\zeta(s)]ds = \int_{-\infty}^{\infty} T(s)L_\zeta(t,s)ds, \quad \text{all } t \in R,$$

with probability one.

Theorem 3.2. *Under Assumptions 3.1 and 3.2, we have*

$$\sqrt{n/d_n}(\hat{\theta}_n - \theta_0) \to_D \Sigma^{-1/2}\, \boldsymbol{N} L_G^{-1/2}(1,0), \tag{16}$$

where $\boldsymbol{N}$ *is a standard normal random vector, which is independent of* $G(t)$ *defined by*

$$G(t) \quad = \quad \begin{cases} W_{3/2-\mu}(t), & \textit{under } \boldsymbol{C1}, \\ W(t), & \textit{under } \boldsymbol{C2}. \end{cases} \tag{17}$$

Here and below, $W_\beta(t)$ denotes the fractional Brownian motion with $0 < \beta < 1$ on $D[0,1]$, defined as follows:

$$W_\beta(t) = \frac{1}{A(\beta)}\Big[\int_{-\infty}^{0} \Big[(t-s)^{\beta-1/2} - (-s)^{\beta-1/2}\Big] dW(s) + \int_0^t (t-s)^{\beta-1/2} dW(s)\Big],$$

where $W(s)$ is a standard Brownian motion and

$$A(\beta) = \Big(\frac{1}{2\beta} + \int_0^\infty \Big[(1+s)^{\beta-1/2} - s^{\beta-1/2}\Big]^2 ds\Big)^{1/2}.$$

Remark 3.1. Theorem 3.2 improves Theorem 5.1 of PP by allowing x_t to be a long memory process, which includes fractional integrated processes as an example.

Using the same routine and slight modifications of assumptions, we have the following limit distribution when x_t is a β-regular Harris recurrent Markov chain.

Assumption 3.1*. (i) x_t is defined as in Example 2 satisfying (9), that is, x_t is a β-regular Harris recurrent Markov chain;

(ii) $E(u_k \mid \mathcal{F}_{nk}) = 0$ for any $1 \le k \le n, n \ge 1$, $\max_{k \ge m} |E(u_{k+1}^2 \mid \mathcal{F}_{nk}) - 1| \to 0$ a.s. and for some $\delta > 0$, $\max_{k\ge 1} E(|u_{k+1}|^{2+\delta} \mid \mathcal{F}_{nk}) < \infty$, where $\mathcal{F}_{nk} = \sigma(\mathcal{F}_k, x_1, ..., x_n)$.

Assumption 3.2*. Let $p(x,\theta)$ be one of f, $\dot{f}_i$ and $\ddot{f}_{ij}$, $1 \le i, j \le m$.

(i) $p(x,\theta_0)$ is a bounded real function with $\int_\infty^\infty |p(x,\theta_0)|\pi(dx) < \infty$;

(ii) $\Sigma_\pi = \int_{-\infty}^\infty \dot{f}(s,\theta_0)\dot{f}(s,\theta_0)'\pi(ds) > 0$ and $\int_{-\infty}^\infty (f(s,\theta) - f(s,\theta_0))^2\pi(ds) > 0$ for all $\theta \ne \theta_0$;

(iii) There exists a bounded function $T_p : \mathbb{R} \to \mathbb{R}$ with $\int_{-\infty}^\infty T_p(x)\pi(dx) < \infty$ such that $|p(x,\theta) - p(x,\theta_0)| \le h_p(||\theta - \theta_0||)\, T_p(x)$, for each $\theta, \theta_0 \in \Theta$, where $h_p(x)$ is a bounded real function such that $h_p(x) \downarrow h_p(0) = 0$, as $x \downarrow 0$.

Theorem 3.3. *Under Assumptions 3.1* and 3.2*, we have*

$$\sqrt{a(n)}(\hat{\theta}_n - \theta_0) \to_D \Sigma_\pi^{-1/2}\, \boldsymbol{N} \Pi_\beta^{-1/2}, \tag{18}$$

where $\boldsymbol{N}$ *is a standard normal random vector, which is independent of* Π_β*, and for* $\beta = 1$, $\Pi_\beta = 1$*, and for* $0 < \beta < 1$, $\Pi_\beta^{-\beta}$ *is a stable random variable with Laplace transform*

$$E \exp\{-t\, \Pi_\beta^{-\beta}\} = \exp\Big\{ - \frac{t^\beta}{\Gamma(\beta+1)} \Big\}, \quad t \geq 0. \tag{19}$$

Remark 3.2. Theorem 3.3 is new, even for the stationary case ($\beta = 1$) in which $a(n) = n$ and the NLS estimator $\hat{\theta}_n$ converges to a normal variate. The random variable Π_β in the limit distribution, after scaled by a factor of $\Gamma(\beta+1)^{-1}$, is a Mittag-Leffler random variable with parameter β, which is closely related to a stable random variable. For details regarding the properties of this distribution, see page 453 of Feller (1971) or Theorem 3.2 of Karlsen and Tjøstheim (2001).

Remark 3.3. Assumption 3.1* (ii) imposes a strong orthogonal property between the regressor x_t and the error sequence u_t. It is not clear at the moment whether such condition can be relaxed to a less restrictive one where x_t is adapted to $\mathcal{F}_{nt}$ and x_{t+1} is independent of $\mathcal{F}_{nt}$ for all $1 \leq t \leq n$. We leave this for future research.

Remark 3.4. Unlike linear cointegration, integrable regression functions do not proportionately transfer the effect of outliers of the nonstationary regressor to the response. This is useful when practitioners want to lessen the impact of extreme values of x_t on the response. In addition, such functions arise in macro-economics when the dependent variable has an uneven response to the regressor. Empirical examples include central banks' market intervention and the currency exchange target zones, described in Phillips (2001).

Remark 3.5. The convergence rates in (16) and (18) are reduced by the nonstationarity property of the regressor x_t when f is an integrable function. For the partial sums of linear processes in Theorem 3.2, comparing to the stationary situation where the convergence rate is $n^{1/2}$, the rate is reduced to $n^{3/4-\mu/2}\rho^{1/2}(n)$ for the long memory case, and is further reduced to $n^{1/4}$ for the short memory case. For the β-regular Harris recurrent Markov chain, note that by (9), the asymptotic order $a(n)$ is regularly varying at infinity, i.e., there exists a slowly varying function $\rho(n)$ such that $a(n) \sim n^\beta \rho(n)$. Therefore, the convergence rate of the NLS estimator decreases as the regularity index β of the chain decreases from one (stationary situation) to zero.

3.2 Limit distribution: beyond integrable functions

We next consider the limit distribution of $\hat{\theta}_n$ when the regression function f is non-integrable. Unlike PP and Chang et al. (2001), we use a more general regressor x_t in the development of our asymptotic distribution.

Assumption 3.3. (i) $\{u_t, \mathcal{F}_t, 1 \le t \le n\}$ is a martingale difference sequence satisfying $E(|u_t|^2|\mathcal{F}_{t-1}) = \sigma^2$ and $\sup_{1\le t\le n} E(|u_t|^{2q}|\mathcal{F}_{t-1}) < \infty$ a.s., where $q > 1$;

(ii) x_t is adapted to $\mathcal{F}_{t-1}$, $t = 1, ..., n$, $\max_{1\le t\le n} |x_t|/d_n = O_P(1)$, where $d_n^2 = var(x_n) \to \infty$; and

(iii) There exists a vector of continuous Gaussian processes (G, U) such that

$$(x_{[nt],n}, n^{-1/2} \sum_{i=1}^{[nt]} u_i) \to_D (G(t), U(t)), \tag{20}$$

on $D_{R^2}[0, 1]$, where $x_{t,n} = x_t/d_n$.

Assumption 3.4. Let $p(x, \theta)$ be one of f, $\dot{f}_i$ and $\ddot{f}_{ij}$, $1 \le i, j \le m$. There exists a real function $T_p : \mathbb{R} \to \mathbb{R}$ such that

(i) $|p(x, \theta) - p(x, \theta_0)| \le A_p(||\theta - \theta_0||)\, T_p(x)$, for each $\theta, \theta_0 \in \Theta$, where $A_p(x)$ is a bounded real function satisfying $A_p(x) \downarrow A_p(0) = 0$, as $x \downarrow 0$;

(ii) For any bounded x, $\sup_{\theta\in\Theta} |p(\lambda x, \theta) - v_p(\lambda)\, h_p(x, \theta)|/T_p(\lambda x) = o(1)$, as $\lambda \to \infty$, where $h_p(x, \theta)$ on Θ is a regular function and $v_p(\lambda)$ is a positive real function which is bounded away from zero as $\lambda \to \infty$; and

(iii) $T_p(\lambda x) \le v_p(\lambda)\, T_{1p}(x)$ as $|\lambda x| \to \infty$, where $T_{1p} : \mathbb{R} \to \mathbb{R}$ is a real function such that $T_{1p} + T_{1p}^2$ is locally integrable (i.e. integrable on any compact set).

For notation convenience, let $v(t) = v_f(t)$, $h(x, \theta) = h_f(x, \theta)$, $\dot{v}_i(t) = v_{\dot{f}_i}(t)$ and $\dot{v}(t) = (\dot{v}_1(t), ..., \dot{v}_m(t))$. Similarly, we define $\ddot{v}$(t), $\dot{h}(x, \theta)$ and $\ddot{h}(x, \theta)$. We have the following main result.

Theorem 3.4. *Suppose Assumptions 3.3 and 3.4 hold. Further assume that* $\sup_{1\le i,j\le m} |\frac{v(d_n)\, \ddot{v}_{ij}(d_n)}{\dot{v}_i(d_n)\, \dot{v}_j(d_n)}| < \infty$ *and* $\int_{|s|\le\delta} \dot{h}(s, \theta_0)\dot{h}(s, \theta_0)' ds > 0$ *for all* $\delta > 0$. *Then we have*

$$D_n\, (\hat{\theta}_n - \theta_0) \to_D \Big(\int_0^1 \Psi(t)\Psi(t)' dt\Big)^{-1} \int_0^1 \Psi(t)\, dU(t), \tag{21}$$

on $D[0, 1]$, *as* $n \to \infty$, *where* $\Psi(t) = \dot{h}(G(t), \theta_0)$ *and* $D_n = diag(\sqrt{n}\dot{v}_1(d_n), ..., \sqrt{n}\dot{v}_m(d_n))$.

Remark 3.6. Except the joint convergence, other conditions in establishing Theorem 3.4 are similar to Theorem 5.2 of PP. PP made use of the concept of H_0-regular. Our conditions on f, $\dot{f}_i$ and $\ddot{f}_{ij}$, $1 \le i, j \le m$ are more straightforward and easy to identify. Furthermore, the joint convergence assumption under present paper is quite natural when $f, \dot{f}$ and $\ddot{f}$ all satisfy Assumption 3.4 (ii). Indeed, under $\dot{f}_i(\lambda x, \theta) \sim \dot{v}_i(\lambda)\dot{h}_i(x, \theta)$, one can easily obtain the following asymptotics under our joint convergence.

$$D_n^{-1} \sum_{t=1}^{n} \dot{f}(x_t, \theta_0) u_t \to_D \int_0^1 \dot{h}(G(t), \theta_0) dU(t), \tag{22}$$

on $D[0,1]$, which is required in the proof of (21). See, e.g., Kurtz and Protter (1991) and Hansen (1992).

Remark 3.7. If we further assume that $U(t)$ and $G(t)$ are asymptotic independent, that is, the long run relationship between the regressor sequence x_t and the innovative sequence u_t vanishes asymptotically, the limiting distribution in (21) will become mixed normal. In particular, when $U(t)$ is a standard Wiener process, we have

$$D_n\,(\hat{\theta}_n - \theta_0) \to_D \Big(\int_0^1 \Psi(t)\Psi(t)' dt \Big)^{-1/2} \mathbf{N}, \tag{23}$$

where $\mathbf{N}$ is a standard normal random vector.

Remark 3.8. Nonlinear cointegrating regressions with structure described in Theorem 3.4 are useful to modelling money demand functions. In such cases, y_t is the logarithm of the real money balance, x_t is the nominal interest rate, and f can either be $f(x, \alpha, \beta) = \alpha + \beta \log |x|$ or $f(x, \alpha, \beta) = \alpha + \beta \log(\frac{1+|x|}{|x|})$. See Bae and de Jong (2007) and Bae et al. (2006) for empirical studies investigating the estimation of money demand functions in USA and Japan respectively. Also, see Bae et al. (2004) for the derivation of these functional forms from the underlying money demand theories studied in macro-economics.

Remark 3.9. Another example is the Michaelis–Menton model, which was considered by Lai (1994). In such case, the regression function is $f(x, \theta_1, \theta_2) = \theta_1/(\theta_2 + x)I\{x \ge 0\}$. Bates and Watts (1988) employs such model to investigate the dependence of rate of enzymatic reactions on the concentration of substrate.

Remark 3.10. Unlike in the situation where f is integrable, there is super convergence when f has the structure given in Assumption 3.4 and x_t is nonstationary. The convergence rate is given by D_n, whose elements $\sqrt{n}\dot{v}_i(d_n)$ are all faster than the standard rate $\sqrt{n}$ in the stationary situation.

We finally remark that, if f is an asymptotically homogeneous function having the structure (11), it is also possible to extend PP's result from a short memory linear process to general nonstationary time series as described in Assumption 3.3. The idea of proof for such result can be easily generalized from Theorem 5.3 of PP. Also see Chang et al. (2001). As it only involves slight modification of notation, we omit the details.

4 Extension to endogeneity

Assumption 2.2 ensures the model (1) having a martingale structure. The result in this regard is now well known. However, there is little work on allowing for contemporaneous correlation between the regressors and regression errors. Using a nonparametric approach, Wang and Phillips (2009b) considered kernel estimate of f and allowed the equation error u_t to be serially dependent and cross correlated with x_s for $|t-s| < m_0$, thereby inducing endogeneity in the regressor, i.e., $cov(u_t, x_t) \neq 0$. In relation to present paper de Jong (2002) considered model (1) without exogeneity, and assuming certain mixing conditions. Chang and Park (2011) considered a simple prototypical model, where the regressor and regression error are driven by i.i.d. innovations. This section provides some extensions to these works. Our main results show that there is an essential effect of endogeneity on the asymptotics, introducing the bias in the related limit distribution. Full investigation in this regard requires new asymptotic results, which will be left for future work.

4.1 Asymptotics: integrable functions

We first consider the situation that f satisfies Assumption 2.1 with T being integrable. Let $\eta_i \equiv (\epsilon_i, \nu_i)$, $i \in Z$ be a sequence of i.i.d. random vectors satisfying $E\eta_0 = 0$ and $E\|\eta_0\|^{2q} < \infty$, where $q \geq 1$. Let the characteristic function $\varphi(t)$ of ϵ_0 satisfy $\int_{-\infty}^{\infty} |\varphi(t)|^2 dt < \infty$ and $\int_{-\infty}^{\infty} |t|^3 |\varphi(t)|^m dt < \infty$ for some $m > 0$. As noticed in Remark 4 of Jeganathan (2008), the conditions on the characteristics function $\varphi(t)$ are not very restrictive.

Assumption 4.1. (i) $x_t = \sum_{j=1}^{t} \xi_j$, where $\xi_j = \sum_{k=0}^{\infty} \phi_k \epsilon_{j-k}$ and the coefficients ϕ_k satisfies (i) $\sum_{k=0}^{\infty} k|\phi_k| < \infty$ and $\sum_{k=0}^{\infty} \phi_k \neq 0$ or (ii) $\phi_k \sim k^{-\mu}\rho(k)$, where $1/2 < \mu < 1$ and $\rho(k)$ is a function slowly varying at ∞.

(ii) $u_t = \sum_{k=0}^{\infty} \psi_k \nu_{t-k}$, where the coefficients ψ_k satisfies that $\sum_{k=0}^{\infty} k^2 |\psi_k|^2 < \infty$, $\psi \equiv \sum_{k=0}^{\infty} \psi_k \neq 0$ and $\sum_{k=0}^{\infty} |\psi_k| \max\{1, |\tilde{\phi}_k|\} < \infty$ where $\tilde{\phi}_k = \sum_{i=0}^{k} \phi_i$.

As we do not impose the independence between ϵ_k and ν_k, Assumption 4.1 provides the endogeneity in the model (1), and it is much general than the conditions set given in Chang and Park (2011). We have the following result.

Theorem 4.1. *Suppose Assumptions 2.1 and 4.1 hold. Further assume:*

(i) T is bounded and integrable, and
(ii) $\int_{-\infty}^{\infty}(f(s,\theta)-f(s,\theta_0))^2ds>0$ for all $\theta\neq\theta_0$.

Then the NLS estimator $\hat\theta_n$ defined by (2) is a consistent estimator of θ_0, i.e. $\hat\theta_n\to_P\theta_0$.

If in addition Assumption 3.2, then, as $n\to\infty$,

$$\sqrt{n/d_n}(\hat\theta_n-\theta_0)\to_D \Sigma^{-1}\,\Lambda^{1/2}\,\boldsymbol{N}L_G^{-1/2}(1,0), \tag{24}$$

where $\widehat{\dot f}(\mu)=\int e^{i\mu x}\dot f(x,\theta_0)dx$,

$$\Lambda\ =(2\pi)^{-1}\int \widehat{\dot f}(\mu)\widehat{\dot f}(\mu)'[Eu_0^2+2\textstyle\sum_{r=1}^{\infty}E(u_0u_re^{-i\mu x_r})]\,d\mu, \tag{25}$$

and other notations are given as in Theorem 3.1.

4.2 Asymptotics: beyond integrable functions

This section considers the situation that f satisfies Assumption 2.1 with T being non-integrable. Let again $\eta_i\equiv(\epsilon_i,\nu_i), i\in Z$ be a sequence of i.i.d. random vectors with $E\eta_0=0$ and $E\|\eta_0\|^{2q}<\infty$, where $q>2$. We make use of the following assumption in related to the model (1).

Assumption 4.2. (i) $x_t=\sum_{j=1}^t\xi_j$, where $\xi_j=\sum_{k=0}^\infty\phi_k\epsilon_{j-k}$ with $\phi=\sum_{k=0}^\infty\phi_k\neq0$ and $\sum_{k=0}^\infty|\phi_k|<\infty$;

(ii) $u_t=\sum_{k=0}^\infty\psi_k\,\nu_{t-k}$, where the coefficients ψ_k are assumed to satisfy $\sum_{k=0}^\infty k|\psi_k|<\infty$, $\psi\equiv\sum_{k=0}^\infty\psi_k\neq0$ and $\sum_{k=0}^\infty k|\psi_k||\phi_k|<\infty$.

Theorem 4.2. *Suppose that Assumption 4.2 holds and in addition to Assumption 2.4, for any $\theta\neq\theta_0$, $g(x,\theta,\theta_0)$ is twice differentiable with that $g'(x,\theta,\theta_0)$ is locally bounded (i.e., bounded on any compact set). Then the NLS estimator $\hat\theta_n$ defined by (2) is a consistent estimator of θ_0, i.e. $\hat\theta_n\to_P\theta_0$.*

Define $\dot h^x(x,\theta)=\frac{\partial\dot h(x,\theta)}{\partial x}$.

Theorem 4.3. *In addition to the conditions of Theorem 4.2, Assumptions 3.4 holds. Further assume that:*

(i) $\sup_{1\le i,j\le m}|\frac{v(\sqrt{n})\,\ddot{v}_{ij}(\sqrt{n})}{\dot{v}_i(\sqrt{n})\,\dot{v}_j(\sqrt{n})}| < \infty$ *and* $\int_{|s|\le\delta}\dot{h}(s,\theta_0)\dot{h}(s,\theta_0)'ds > 0$ *for all* $\delta > 0$*, and*

(ii) $\dot{h}^x(x,\theta)$ *is locally bounded (i.e., bounded on any compact set).*

Then the limit distribution of $\hat{\theta}_n$ *is given by*

$$D_n(\hat{\theta}_n - \theta_0) \to_D \Big(\int_0^1 \Psi(t)\Psi(t)'dt\Big)^{-1}\big[\sigma_{\xi u}\int_0^1 \dot{\Psi}(t)\,dt + \int_0^1 \Psi(t)\,dU(t)\big], \tag{26}$$

as $n\to\infty$*, where* $\Psi(t) = \dot{h}(W(t),\theta_0)$*,* $\dot{\Psi}(t) = \dot{h}^x(W(t),\theta_0)$*,* $\sigma_{\xi u} = \sum_{j=0}^{\infty}E(\xi_0 u_j)$*,* $D_n =$ *diag*$(\sqrt{n}\dot{v}_1(\sqrt{n}),...,\sqrt{n}\dot{v}_m(\sqrt{n}))$ *and* $(W(t),U(t))$ *is a bivariate Brownian motion with covariance matrix*

$$\Delta = \begin{pmatrix}\phi^2E\epsilon_0^2 & \phi\psi E\epsilon_0\nu_0\\ \phi\psi E\epsilon_0\nu_0 & \psi^2E\nu_0^2\end{pmatrix}.$$

5 Simulation

In this section, we investigate the finite sample performance of the NLS estimator $\hat{\theta}_n$ of nonlinear regressions with endogeneity. Chang and Park (2011) performed simulation of similar model, but only considered the error structure u_t to be i.i.d. innovation. We intend to investigate the sampling behavior of $\hat{\theta}_n$ under different degree of serially dependence of u_t on itself. To this end, we generate our data in the following way:

$$x_t = x_{t-1} + \epsilon_t,$$
$$v_t = \sqrt{1-\rho^2}\,w_t + \rho\epsilon_t,$$

$$f_1(x,\theta) = \exp\{-\theta|x|\} \quad \text{and} \quad f_2(x,\alpha,\beta_1,\beta_2) = \alpha + \beta_1 x + \beta_2 x^2, \tag{27}$$

where $\{w_t\}$ and $\{\epsilon_t\}$ are i.i.d. $N(0,3^2)$ variables, and ρ is the correlation coefficient that controls the degree of endogeneity. The true value of θ is set as 0.1 and that of (α,β_1,β_2) is set as $(-20,10,0.1)$. The error structure is generated according to the following three scenarios:

S1: $u_t = v_t$,

S2: $u_t = \sum_{j=1}^{\infty} j^{-10}\,v_{t-j+1}$, and

S3: $u_t = \sum_{j=1}^{\infty} j^{-4}\,v_{t-j+1}$.

Scenario **S1** is considered by Chang and Park (2011), which eliminates the case that u_t is serially correlated. Scenarios **S2** and **S3** introduce self-dependence to the error sequence. The decay rate of **S2** is faster than that of **S3**, which implies that the level of self dependence of u_t increases from **S1** to **S3**.

It is obvious that f_1 is an integrable function and f_2 satisfies the conditions in Theorem 4.3, with

$$v(\lambda, \alpha, \beta_1, \beta_2) = \lambda^2 \quad \text{and} \quad h(x, \alpha, \beta_1, \beta_2) = \beta_2 x^2.$$

In addition, we have

$$\dot{h}(x, \alpha, \beta_1, \beta_2) = (1, x, x^2)' \quad \text{and} \quad \dot{h}^x(x, \alpha, \beta_1, \beta_2) = (0, 1, 2x)'.$$

It would be convenient for later discussion to write out the limit distributions of the estimators for the f_2 case. By Theorem 4.3, we have

$$\begin{pmatrix} n^{1/2}(\hat{\alpha}_n - \alpha_0) \\ n(\hat{\beta}_{1n} - \beta_{10}) \\ n^{3/2}(\hat{\beta}_{2n} - \beta_{20}) \end{pmatrix} \to_D \Sigma^{-1} \begin{pmatrix} U(1) \\ \sigma_{\xi u} + \int_0^1 W(t)dU(t) \\ 2\sigma_{\xi u} \int_0^1 W(t)dt + \int_0^1 W^2(t)dU(t) \end{pmatrix}, \tag{28}$$

where $\sigma_{\xi u} = \sum_{j=0}^{\infty} E(\epsilon_0 u_j)$ and

$$\Sigma = \begin{pmatrix} 1 & \int_0^1 W(t)dt & \int_0^1 W^2(t)dt \\ \int_0^1 W(t)dt & \int_0^1 W^2(t)dt & \int_0^1 W^3(t)dt \\ \int_0^1 W^2(t)dt & \int_0^1 W^3(t)dt & \int_0^1 W^4(t)dt \end{pmatrix}.$$

In our simulations, we draw samples of sizes $n = 200, 500$ to estimate the NLS estimators and their t-ratios. Each simulation is run for 10,000 times, and their densities are estimated using kernel method with normal kernel function. It is expected that as ρ increases, the degree of endogeneity increases, and hence the variance of the distribution will increase due to (24) and (26). It is also expected that the more serial correlation of u_t, the higher the variance of the limit distribution due to the cross terms appeared in (25) and (26). Our simulation results largely corroborate with our theoretical results in Section 4.

Firstly, the means of the estimators are close to the true values, and the deviations from true values decrease as the sample sizes increase. The reductions of standard errors from $n = 200$ to 500 are close to the theoretical values. For example, the ratio of standard errors for $\hat{\alpha}_n$ between the sample size 200 and 500 is 1.564 in scenario **S1**, which is close to theoretical value $\sqrt{\frac{500}{200}} = 1.581$. This confirms that our estimators are accurate in all scenarios. We present all these numerical values in Tables 2 and 3, which are given in the Supplemental Materials of this paper (Section 12).

Secondly, comparing the **S1**–**S3** curves in each plot, we can see that the **S3** curves have fat tails and lower peak than those of **S1**–**S2**. This verifies that high dependence of u_t on its own past will increase the variance of limit distribution. On the other hand, comparing the shapes of the curves for different values of ρ, the curves with $\rho = 0$ have highest peaks, and the peakedness decrease as ρ increases. This matches with our expectation that, if the cross dependence between u_t and x_t increases, the variance of limit distribution will increase. The following Figure 1 provides the density estimates of $\hat{\theta}_n$ with $n = 500$. Additional figures for density estimates of $\hat{\theta}_n$ and $(\hat{\alpha}_n, \hat{\beta}_{1n}, \hat{\beta}_{2n})$ are given in the Supplemental Materials of this paper (Figures 6-9).

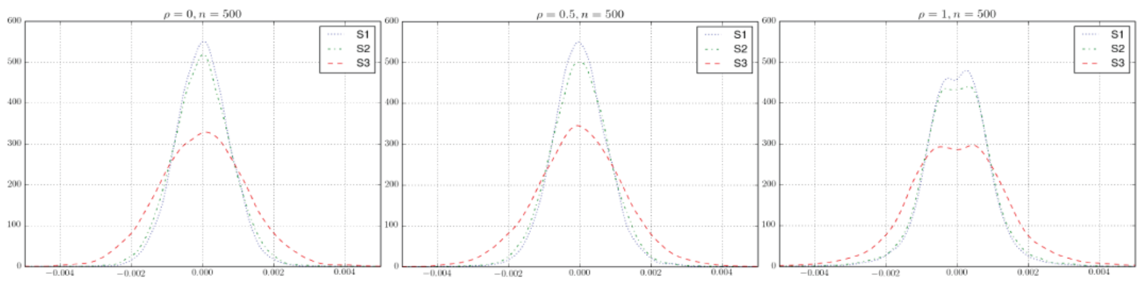

Figure 1: Density estimates of $\hat{\theta}_n$.

Finally, the sampling results for t-ratios are also much expected from our limit theories. In particular, we have interesting results for the t-ratios of $\hat{\beta}_{1n}$ in Figure 2. In scenario **S1**,

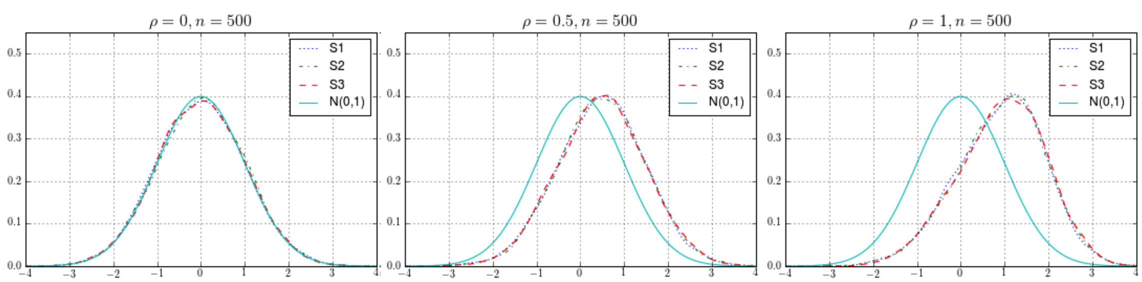

Figure 2: Density of $\hat{\beta}_{1n}$ t-ratios.

there is no endogeneity and the limiting t-ratios overlap with the normal curve. However, as we introduce endogeneity to the model in scenarios **S2** and **S3**, the t-ratios is shifted away from the standard normal. Such behavior can be explained by looking at the limit distribution of $\hat{\beta}_{1n}$ in (28). When $\rho \neq 0$, there exists an extra term of $\Sigma^{-1}\sigma_{\xi u}$, and such term will introduce bias to the limit distribution. More figures for t-ratios are given in the Supplemental Materials of this paper (Figures 10-13).

6 An empirical example: Carbon Kuznets Curve

In this section, we consider a simple example of the model with endogeneity. The equation of interest is the Carbon Kuznets Curve (CKC) relating the per capita CO_2 emission of a country to its per capita GDP. Basically, the CKC hypothesis states that there is an inverted-U shaped relationship between economic activities and per capita CO_2 emissions. As it is well accepted that the GDP variable displays some wandering nonstationary behaviors over time, such problem is relevant to nonlinear transformations of nonstationary regressors. Refer to Wagner (2008) and Müller-Fürstenberger and Wagner (2007) for detailed expositions of how such problem relates to the works of PP and Chang et al. (2001).

The formula that we currently consider has a quadratic formulation and is given by

$$\ln(e_t) = \alpha + \beta_1 \ln(x_t) + \beta_2(\ln(x_t))^2 + u_t, \quad 1 \le t \le n, \tag{29}$$

where e_t and x_t denote the per capital emissions of CO_2 and GDP in period t, and u_t is a stochastic error term. Figure 3 presents the plots of logarithm of CO_2 versus logarithm of GDP for Belgium, Denmark and France. The quadratic polynomial specification of CO_2 in GDP has been extensively analysed in the literature. See the introduction of Piaggio and Padilla (2012) for an overiew. Roughly speaking, the upward slope of the curve can be explained by the increase in natural resources depletion as the economic activities of grow. As the country continues to develop, technological advance and stricter regulatory policies will start contributing to a reduction in the emission of air pollutants, hence, result in an inverted-U shape.

In the usual CKC formulation, the logarithm of per capita CO_2 emissions and GDP are assumed to be integrated processes, see Müller-Fürstenberger and Wagner (2007). Also, it is clear that the nonlinear link function $f(x, \alpha, \beta_1, \beta_2) = \alpha + \beta_1 x + \beta_2 x^2$ satisfies the conditions in Theorem 4.3, and is considered in our simulations. There are studies which employ more complicated models, by including time trend, stochastic terms and additional explanatory variables such as the energy structure. To avoid complicating our discussion, we only consider GDP as the only explanatory variable.

In the literature most studies assumed that the parameters (β_1, β_2) are homogeneous across different countries. Several works investigated the appropriateness of restricting all countries adhering to the same values of (β_1, β_2). For example, List and Gallet (1999), Dijkgraaf et al. (2005), Piaggio and Padilla (2012). Particularly, Piaggio and Padilla (2012) tested this assumption by checking whether the confidence intervals of (β_1, β_2) across different countries overlap. They rejected the homogeneity assumption of parameters across different countries, based on the result that they could not find any possible group with more than 4 countries whose confidence intervals overlap with each other.

As an example of employing our NLS estimators $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_{1n}, \hat{\beta}_{2n})$ and calculating their confidence intervals, we carry out a similar study and examine whether the 95% confidence intervals of the parameters $(\hat{\beta}_{1n}, \hat{\beta}_{2n})$ across different countries overlap with each other. We extend previous studies by incorporating the endogeneity in the model. The assumption that x_t is an exogenous variable does not usually hold in practise as there are many potential sources of endogeneity. Firstly, endogeneity occurs due to measurement error of the explanatory variables x_t, including the misreporting of GDP of each country. Secondly, there might be omitted variable bias, which arises if there exists a relevant explanatory variable that is correlated with the regressor x_t (per capita GDP), but is excluded from the model. Finally, potential reverse causality will give rise to endogeneity. That is, the CO_2 emi ssions y_t might at the same time has a feedback effect on the GDP x_t (e.g., due to stricter government regulation).

We consider 14 countries, which are shown in Piaggio and Padilla (2012) that their response variables have a quadratic relationship described in (29) with the regressors. We use annual data in this example and for each country, there are 58 observations of the CO_2 emission data from 1951–2008 published by the Carbon Dioxide Information Analysis Center (Boden et al. (2009)). For per capita GDP, we adopt the GDP data from Maddison (2003), which is transformed to 1990 Geary-Khamis dollars. We performed a unit root test to ensure the data are nonstationary, and we observed that all Dickey-Fuller test statistics were greater than the critical value. Therefore, we do not reject the hypothesis that there is a unit root. The results are presented in the Supplementary Materials of

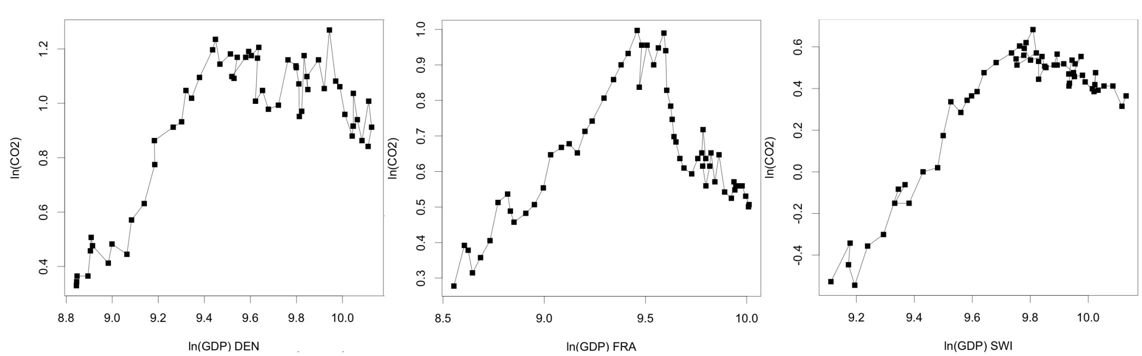

Figure 3: Plots of $\log(CO_2)$ against $\log(GDP)$ for Belgium, Denmark and France.

this paper (Section 11).

We estimate the parameters by minimizing the error function given in (2) using the `nls()` function in the statistical software R. Next, we calculate the confidence interval by finding the critical values of the limit distribution of $(\hat{\alpha}_n, \hat{\beta}_{1n}, \hat{\beta}_{2n})$ given in (29). Note that most of the quantities in (29) involve only the standard Brownian motion or integral of Brownian motions, which can be easily obtained by direct simulations. However, there exist two nuisance parameters, including the covariance matrix Δ of Brownian motions (W, U) and the autocovariances $\sigma_{\xi u}$ of linear processes $\{\xi_t, u_t\}$. Thus we have to replace the unknown parameters with their consistent estimates and adopt the empirical critical values instead of the real ones. Using the linear process estimation procedures described in Chang et al. (2001), we have $\hat{\phi}$, $\hat{\psi}$ as the estimators of ϕ, ψ, and $\hat{\epsilon}_t$ and $\hat{\nu}_t$ as the artificial error sequences. The parameters Δ and $\sigma_{\xi u}$ can then be estimated by

$$\hat{\Delta} = \begin{pmatrix} n^{-1}\hat{\phi}^2 \sum_{t=1}^n \hat{\epsilon}_t^2 & n^{-1}\hat{\phi}\hat{\psi} \sum_{t=1}^n \hat{\epsilon}_t\hat{\nu}_t \\ n^{-1}\hat{\phi}\hat{\psi} \sum_{t=1}^n \hat{\epsilon}_t\hat{\nu}_t & n^{-1}\hat{\psi}^2 \sum_{t=1}^n \hat{\nu}_t^2 \end{pmatrix}, \quad \text{and} \quad \hat{\sigma}_{\xi u} = \sum_{j=1}^{l} n^{-1} \sum_{t=1}^{n-j} \xi_t \hat{u}_{t+j},$$

for any $l = o(n^{1/4})$. See Ibragimov and Phillips (2008), Phillips and Solo (1992) and Phillips and Perron (1988).

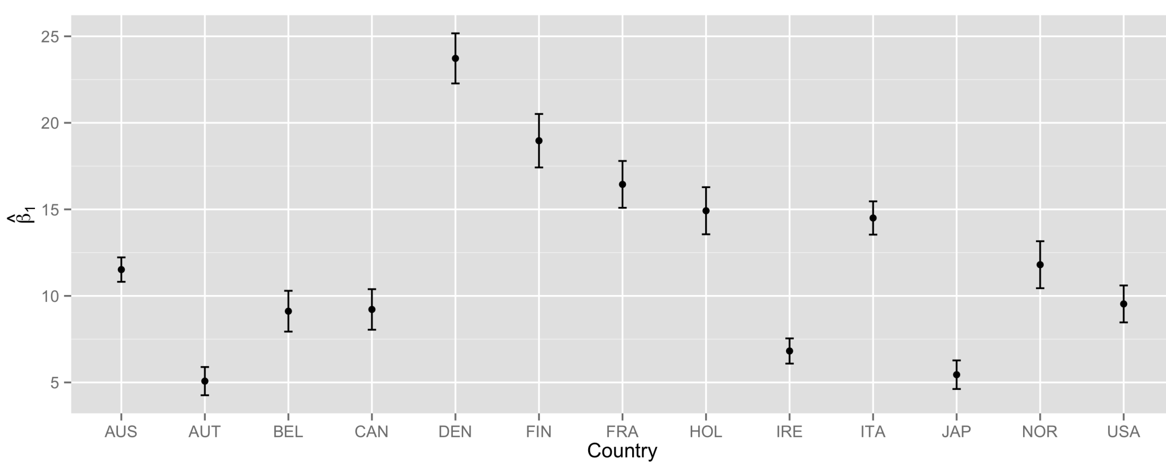

Figure 4: Estimates and 95% Confidence Intervals of $\hat{\beta}_{1n}$.

Tables 5 and 6 in the Supplementary Materials of this paper report the estimates and 95% confidence intervals of parameters β_1 and β_2, and Figures 4 and 5 here present the plots of the results. For the parameter β_1, we can not find any group with more than four groups of countries whose confidence intervals overlap. Such result is consistent with that of Piaggio and Padilla (2012). However, for the parameter $\hat{\beta}_{2n}$, among the 13 countries, we can find group of 11 countries whose confidence intervals overlap. Hence, the homogeneity of parameters can not be rejected as in Piaggio and Padilla (2012). There are two main reasons for the differences between our results and theirs. Firstly, they assume the innovation is i.i.d. normally distributed while we allow the error to be serially correlated with itself. Secondly, we incorporate the effect of endogeneity in our limit distribution of the estimators. Based on (29), both reasons will make the critical values and confidence intervals significantly larger than that of Piaggio and Padilla (2012). Similarly, we may construct the estimate and confidence interval for α. See Table 4 and Figure 14 in Supplementary Materials of this paper for this kind of result.

7 Conclusion and discussion

In this paper, we establish an asymptotic theory for a nonlinear parametric cointegrating regression model. A general framework is developed for establishing the weak consistency of the NLS estimator $\hat{\theta}_n$. The framework can easily be applied to a wide class of nonstationary regressors, including partial sums of linear processes and Harris recurrent Markov chains. Limit distribution of $\hat{\theta}_n$ is also established, which extends previous works. Furthermore, we introduce endogeneity to our model by allowing the error term to be serially dependent on itself, and cross-dependent on the regressor. We show that the limit distribution of $\hat{\theta}_n$ under the endogeneity situation is different from that with martingale error structure. This result is of interest in the applied econometric research area.

Asymptotics in this paper are limited to univariate regressors and nonlinear parametric cointegrating regression models. Specification of the nonlinear regression function f depends usually on the underlying theories of the subject. Illustration can be found in Bae et al. (2004), where the concept of the money in the utility function with the constant elasticity of substitution (MUFCES) is used to relate the money balance with nominal interest rate by the link function $f(x, \alpha, \beta) = \alpha + \beta \log(\frac{x}{1+x})$. More currently, nonparametric method has been developed to test the specification of f. See, e.g., Wang and Phillips (2012), for instance. Extension to multivariate regressors require substantial different techniques. This is because local time theory can not in general be extended to multivariate data, in which the techniques play a key rule in the asymptotics of the NLS estimator with nonstationary regressor when f is an integrable function. Possible extension to multivariate regressors is the index model discussed in Chang and Park (2003). This again requires some new techniques, and hence we leave it for future work.

8 Proofs of main results

This section provides proofs of the main results. We start with some preliminaries, which list the limit theorems that are commonly used in the proofs of the main results.

8.1 Preliminaries

Denote $\mathcal{N}_\delta(\theta_0) = \{\theta : \|\theta - \theta_0\| < \delta\}$, where $\theta_0 \in \Theta$ is fixed.

Lemma 8.1. *Let Assumption 2.1 hold. Then, for any x_t satisfying (4) with T being given as in Assumption 2.1,*

$$\sup_{\theta \in \mathcal{N}_\delta(\theta_0)} \kappa_n^{-2} \sum_{t=1}^n \big(|f(x_t,\theta) - f(x_t,\theta_0)| + |f(x_t,\theta) - f(x_t,\theta_0)|^2\big) \to_P 0, \tag{30}$$

as $n \to \infty$ first and then $\delta \to 0$. If in addition Assumption 2.2, then

$$\kappa_n^{-2} \sum_{t=1}^n [f(x_t,\theta_0) - f(x_t,\pi_0)]\, u_t \to_P 0, \tag{31}$$

for any $\theta_0, \pi_0 \in \Theta$, and

$$\sup_{\theta \in \mathcal{N}_\delta(\theta_0)} \kappa_n^{-2} \sum_{t=1}^n |f(x_t,\theta) - f(x_t,\theta_0)|\, |u_t| \to_P 0, \tag{32}$$

as $n \to \infty$ first and then $\delta \to 0$.

Proof. (30) is simple. As $\kappa_n^{-2}\sum_{t=1}^n [f(x_t,\theta_0) - f(x_t,\pi_0)]^2 \le C\kappa_n^{-2}\sum_{t=1}^n T^2(x_t) = O_P(1)$, (31) follows from Lemma 2 of Lai and Wei (1982). As for (32), the result follows from

$$\sum_{t=1}^n |f(x_t,\theta) - f(x_t,\theta_0)|\, |u_t| \le h(\|\theta - \theta_0\|) \sum_{t=1}^n T(x_t)\, |u_t|,$$

and $\sum_{t=1}^n T(x_t)\,|u_t| \le C\sum_{t=1}^n T(x_t) + \sum_{t=1}^n T(x_t)\,\big[|u_t| - E(|u_t| \mid \mathcal{F}_{t-1})\big] = O_P(\kappa_n^2)$. □

Lemma 8.2. *Let Assumption 3.1 (i) hold. For any bounded $g(x)$ satisfying $\int_{-\infty}^{\infty} |g(x)|dx < \infty$, we have*

$$\sum_{t=1}^n g(x_t) \;=\; O_P(n/d_n). \tag{33}$$

If in addition Assumption 3.1 (ii)–(iii), then, for any bounded $g_i(x), i = 1, 2$, satisfying $\int_{-\infty}^{\infty} |g_i(x)|dx < \infty$ and $\int_{-\infty}^{\infty} g_i(x)dx \neq 0$,

$$\Big\{\big(\frac{d_n}{n}\big)^{1/2} \sum_{t=1}^n g_1(x_t)u_t,\, \frac{d_n}{n}\sum_{t=1}^n g_2(x_t)\Big\} \;\to_D\; \Big\{\tau_1\, N\, L_G^{1/2}(1,0), \tau_2\, L_G(1,0)\Big\}, \tag{34}$$

where $\tau_1^2 = \int_{-\infty}^{\infty} g_1^2(s)ds$, $\tau_2 = \int_{-\infty}^{\infty} g_2(s)ds$ and N is a standard normal variate independent of $G(t)$.

Proof. The result (33) sees Lemma 3.2 of Wang and Phillips (2009b). Theorem 2.2 of Wang (2013) provides (34) with $g_2(x) = g_1^2(x)$. It is not difficult to see that (34) still holds for general $g_2(x)$. We omit the details. □

Lemma 8.3. *Let x_t be defined as in Example 2. For any g such that $\int_{-\infty}^{\infty} |g(x)|\pi(dx) < \infty$ and $\int_{-\infty}^{\infty} g(x)\pi(dx) \neq 0$, we have*

$$\sum_{t=1}^{n} g(x_t) = O_P[a(n)], \tag{35}$$

$$\Big[\sum_{t=1}^{n} g(x_t)\Big]^{-1} = O_P[a(n)^{-1}]. \tag{36}$$

If Assumption 3.1 holds, then, for any bounded $g_i(x), i = 1, 2$, satisfying $\int_{-\infty}^{\infty} |g_i(x)|\pi(dx) < \infty$ and $\int_{-\infty}^{\infty} g_i(x)\pi(dx) \neq 0$,*

$$\Big\{a(n)^{-1/2}\sum_{t=1}^{n} g_1(x_t)u_t,\, a(n)^{-1}\sum_{t=1}^{n} g_2(x_t)\Big\} \to_D \Big\{\tau_5\, N\,\Pi_\beta^{1/2}, \tau_6\,\Pi_\beta\Big\}, \tag{37}$$

where $\tau_5^2 = \int_{-\infty}^{\infty} g_1^2(s)\pi(ds)$, $\tau_6 = \int_{-\infty}^{\infty} g_2(s)\pi(ds)$, Π_β is defined as in Theorem 3.3 and N is a standard normal variate independent of Π_β.

Proof. See the Supplemental Material of this paper. □

Lemma 8.4. *Under Assumption 4.1, for any bounded $g_i(x), i = 1, 2$, satisfying $\int_{-\infty}^{\infty} |g_i(x)|dx < \infty$ and $\int_{-\infty}^{\infty} g_i(x)dx \neq 0$, we have*

$$\sum_{t=1}^{n} g_1(x_t)|u_t| = O_P(n/d_n), \tag{38}$$

and

$$\begin{aligned}\Big\{(n/d_n)^{-1/2}\textstyle\sum_{t=1}^{n} g_1(x_t)u_t,\, (n/d_n)^{-1}\sum_{t=1}^{n} g_2(x_t)\Big\}&\\ \to_D \Big\{\tau_3\, N\, L_G^{1/2}(1,0), \tau_4 L_G(1,0)\Big\},&\end{aligned} \tag{39}$$

where $\tau_3^2 = (2\pi)^{-1}\int_{-\infty}^{\infty} \hat{g}_1(\mu)^2[Eu_0^2 + 2\sum_{r=1}^{\infty} E(u_0u_re^{-i\mu x_r})]\,d\mu$, $\hat{g}_1(\mu) = \int_{-\infty}^{\infty} e^{i\mu x}g_1(x)dx$ and $\tau_4 = \int_{-\infty}^{\infty} g_2(s)ds$.

Proof. See Theorem 3 and 5 of Jeganathan (2008). □

Lemma 8.5. *Under Assumption 4.2, for any twice continuous differentiable function $g_i(x)$, $i = 1, 2$, with that $|g_i'(x)|$ is locally bounded (i.e., bounded on any compact set), we have*

$$\begin{aligned} &\Big\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_1\Big(\frac{x_t}{\sqrt{n}}\Big)u_t, \frac{1}{n}\sum_{t=1}^{n} g_2\Big(\frac{x_t}{\sqrt{n}}\Big)\Big\} \\ &\to_D \Big\{\sigma_{\xi u}\int_0^1 g_1'(W(t))dt + \int_0^1 g_1(W(t))dU(t), \int_0^1 g_2(W(t))dt\Big\}, \end{aligned} \tag{40}$$

as $n \to \infty$, where $\sigma_{\xi u} = \sum_{j=0}^{\infty} E(\xi_0 u_j)$, and $(W(t), U(t))$ is a bivariate Brownian motion with covariance matrix

$$\Delta = \begin{pmatrix} \phi^2 E\epsilon_0^2 & \phi\psi E\epsilon_0\nu_0 \\ \phi\psi E\epsilon_0\nu_0 & \psi^2 E\nu_0^2 \end{pmatrix}.$$

Consequently, under the conditions of Theorem 4.2, we have

$$\sum_{t=1}^{n}(f(x_t, \pi_0) - f(x_t, \theta_0))u_t = o_P\big[n\, v(\sqrt{n})\big]. \tag{41}$$

Proof. The result (40) sees Theorem 4.3 of Ibragimov and Phillips (2008) with minor improvements. To see (41), recall $\max_{1\le t\le n}|x_t|/\sqrt{n} = O_P(1)$. Without loss of generality, we assume $\max_{1\le t\le n}|x_t|/\sqrt{n} \le K_0$ for some $K_0 > 0$. It follows from Assumption 2.4 that

$$\begin{aligned} &\sum_{t=1}^{n}(f(x_t, \pi_0) - f(x_t, \theta_0))u_t \\ =\;& v(\sqrt{n})\sum_{t=1}^{n} g(x_t/\sqrt{n}, \pi_0, \theta_0)u_t + o_P(v(\sqrt{n}))\sum_{t=1}^{n} T_1(x_t/\sqrt{n})|u_t| \\ =\;& o_P[nv(\sqrt{n})], \end{aligned}$$

as required. □

Lemma 8.6. *Under Assumptions 3.1 (i) and 3.2, we have*

$$\frac{d_n}{n}\sup_{\theta\in\mathcal{N}_\delta(\theta_0)}\sum_{t=1}^{n}\big|\dot f_i(x_t,\theta)\,\dot f_j(x_t,\theta) - \dot f_i(x_t,\theta_0)\,\dot f_j(x_t,\theta_0)\big| \;\to_P\; 0, \tag{42}$$

$$\frac{d_n}{n}\sup_{\theta\in\mathcal{N}_\delta(\theta_0)}\sum_{t=1}^{n}\big|\ddot f_{ij}(x_t,\theta)\,[f(x_t,\theta) - f(x_t,\theta_0)]\big| \;\to_P\; 0, \tag{43}$$

as $n \to \infty$ first and then $\delta \to 0$, for any $1 \le i, j \le m$. If in addition Assumption 3.1 (ii)–(iii), then

$$\frac{d_n}{n}\sup_{\theta\in\mathcal{N}_\delta(\theta_0)}\big|\sum_{t=1}^{n}\ddot f_{ij}(x_t,\theta)\,u_t\big| \to_P 0, \tag{44}$$

as $n \to \infty$ first and then $\delta \to 0$, for any $1 \le i, j \le m$.

Similarly, under Assumptions 3.1 (i) and 3.2*, (42) and (43) are true if we replace d_n/n by $a(n)^{-1}$. If in addition Assumption 3.1* (ii), then (44) holds if we replace d_n/n by $a(n)^{-1}$.*

Proof. See the Supplemental Material of this paper. □

Lemma 8.7. *Under Assumptions 3.3 and 3.4, we have*

$$\frac{1}{n\dot{v}_i(d_n)\dot{v}_j(d_n)} \sup_{\theta \in \mathcal{N}_\delta(\theta_0)} \sum_{t=1}^{n} \big| \dot{f}_i(x_t, \theta)\, \dot{f}_j(x_t, \theta) - \dot{f}_i(x_t, \theta_0)\, \dot{f}_j(x_t, \theta_0) \big| \to_P 0, \tag{45}$$

$$\frac{1}{nv(d_n)\ddot{v}_{ij}(d_n)} \sup_{\theta \in \mathcal{N}_\delta(\theta_0)} \sum_{t=1}^{n} \big| \ddot{f}_{ij}(x_t, \theta)\, [f(x_t, \theta) - f(x_t, \theta_0)] \big| \to_P 0, \tag{46}$$

$$\frac{1}{n\ddot{v}_{ij}(d_n)} \sup_{\theta \in \mathcal{N}_\delta(\theta_0)} \big| \sum_{t=1}^{n} \ddot{f}_{ij}(x_t, \theta)\, u_t \big| \to_P 0, \tag{47}$$

as $n \to \infty$ first and then $\delta \to 0$, for any $1 \le i, j \le m$.

Proof. See the Supplemental Material of this paper. □

Lemma 8.8. *Let Assumption 3.3 hold. Then, for any regular functions $g(x, \theta)$ and $g_1(x, \theta)$ on Θ, we have*

$$\Big\{ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} g\Big(\frac{x_t}{d_n}, \theta\Big)\, u_t,\ \frac{1}{n} \sum_{t=1}^{n} g_1\Big(\frac{x_t}{d_n}, \theta\Big) \Big\} \to_D \Big\{ \int_0^1 g\big(G(t), \theta\big) dU(t),\ \int_0^1 g_1\big(G(t), \theta\big) dt \Big\}. \tag{48}$$

Proof. If $g(x, \theta)$ and $g_1(x, \theta)$ are continuous functions, the result follows from (20) and the continuous mapping theorem. See, Kurtz and Prötter (1991) for instance. The extension from continuous function to regular function is standard in literature. The details can be found in PP or Park and Phillips (1999). □

Lemma 8.9. *Let $D_n(\theta, \theta_0) = Q_n(\theta) - Q_n(\theta_0)$. Suppose that, for any $\delta > 0$,*

$$\liminf_{n \to \infty} \inf_{|\theta - \theta_0| \ge \delta} D_n(\theta, \theta_0) > 0 \quad in \quad probability, \tag{49}$$

then $\hat{\theta}_n \to_P \theta_0$.

Proof. See Lemma 1 of Wu (1981). □

8.2 Proofs of theorems

Proof of Theorem 2.1. Let $\mathcal{N}$ be any open subset of Θ containing θ_0. Since $\hat{\theta}_n$ is the minimizer of $Q_n(\theta)$ over $\theta \in \Theta$, by Lemma 8.9, proving consistency is equivalent to showing that, for any $0 < \eta < 1/3$ and $\theta \neq \theta_0$, where $\theta, \theta_0 \in \Theta$, there exist $n_0 > 0$ and $M_1 > 0$ such that

$$P\Big(\inf_{\theta \in \Theta \cap \mathcal{N}^c} D_n(\theta, \theta_0) \geq \kappa_n^2/M_1\Big) \geq 1 - 3\eta, \tag{50}$$

for all $n > n_0$.

Denote $\mathcal{N}_\delta(\pi_0) = \{\theta : \|\theta - \pi_0\| < \delta\}$. Since $\Theta \cap \mathcal{N}^c$ is compact, by the finite covering property of compact set, (50) will follow if we prove that, for any fixed $\pi_0 \in \Theta \cap \mathcal{N}^c$,

$$I_n(\delta, \pi_0) := \sup_{\theta \in \mathcal{N}_\delta(\pi_0)} \kappa_n^{-2} \Big| D_n(\theta, \theta_0) - D_n(\pi_0, \theta_0) \Big| \to_P 0, \tag{51}$$

as $n \to \infty$ first and then $\delta \to 0$, and for $\forall \eta > 0$, there exist $M_0 > 0$ and $n_0 > 0$ such that for all $n \geq n_0$ and $M \geq M_0$,

$$P\Big(D_n(\pi_0, \theta_0) \geq \kappa_n^2/M\Big) \geq 1 - 2\eta. \tag{52}$$

Indeed, due to (51), for any $0 < \eta < 1/3$ and $M_1 > 0$, there exist $n_0 > 0$ and $\delta_0 > 0$ such that

$$P(\max_{1 \leq j \leq m_0} I_n(\delta_0, \pi_j) \geq 1/2M_1) \leq \eta,$$

where m_0 and $\pi_j, 1 \leq j \leq m_0$, are chosen so that $\Theta \cap \mathcal{N}^c \subset \bigcup_{j=1}^{m_0} \mathcal{N}_{\delta_0}(\pi_j)$. Consequently, by taking $M_1 \geq M_0/(2m_0)$, it follows from (52) that

$$\begin{aligned} & P\Big(\inf_{\theta \in \Theta \cap \mathcal{N}^c} D_n(\theta, \theta_0) \geq \kappa_n^2/M_1\Big) \\ \geq\ & P\Big(\inf_{1 \leq j \leq m_0} D_n(\pi_j, \theta_0) \geq \kappa_n^2/(2M_1)\Big) - \eta \\ \geq\ & \inf_{1 \leq j \leq m_0} P\Big(D_n(\pi_j, \theta_0) \geq \kappa_n^2/(2m_0 M_1)\Big) - \eta \geq 1 - 3\eta, \end{aligned}$$

which yields (50).

We next prove (51) and (52). The result (51) is simple, which follows immediately from Lemma 8.1 and the fact that, for each fixed $\pi_0 \in \Theta \cap \mathcal{N}^c$,

$$\begin{aligned} I_n(\delta, \pi_0) \leq & \sup_{\theta \in \mathcal{N}_\delta(\pi_0)} \kappa_n^{-2} \sum_{t=1}^n (f(x_t, \theta) - f(x_t, \pi_0))^2 \\ & + \sup_{\theta \in \mathcal{N}_\delta(\pi_0)} \kappa_n^{-2} \sum_{t=1}^n |f(x_t, \theta) - f(x_t, \pi_0)||u_t|. \end{aligned}$$

To prove (52), we recall

$$D_n(\pi_0,\theta_0)=\sum_{t=1}^n(f(x_t,\pi_0)-f(x_t,\theta_0))^2-\sum_{t=1}^n(f(x_t,\pi_0)-f(x_t,\theta_0))u_t.$$

This, together with (5) and (31), implies that, for any $\eta>0$, there exists $n_0>0$ and $M_0>0$ such that for all $n>n_0$ and $M\geq M_0$,

$$\begin{aligned}P\Big(D_n(\pi_0,\theta_0)\geq \kappa_n^2M^{-1}\Big)&\geq P\Big(\sum_{t=1}^n(f(x_t,\pi_0)-f(x_t,\theta_0))^2\geq \kappa_n^2\,M^{-1}/2\Big)-\eta\\&\geq 1-2\eta,\end{aligned}$$

as required.

Finally, by noting $Q_n(\theta_0)=n^{-1}\sum_{t=1}^n u_t^2\to_P\sigma^2$ due to Assumption 2.2 and strong law of large number, it follows from the consistency of $\hat\theta_n$ and (51) that

$$|\hat\sigma_n^2-\sigma^2|\leq C\kappa_n^{-2}|Q_n(\hat\theta_n)-Q_n(\theta_0)|+o_P(1)=o_P(1).$$

□

Proof of Theorem 2.2. (4) follows from (33) of Lemma 8.2. (5) follows from (34) of Lemma 8.2 with $g_2(x)=(f(x,\theta)-f(x,\theta_0))^2$ and the facts that $P(L_G(1,0)>0)=1$ and $\int_{-\infty}^\infty(f(s,\theta)-f(s,\theta_0))^2ds>0$, for any $\theta\neq\theta_0$. □

Proof of Theorem 2.3. (4) follows from (35) of Lemma 8.3 with $g(x)=T(x)+T^2(x)$. (5) follows from (36) of Lemma 8.3 with $g(x)=(f(x,\theta)-f(x,\theta_0))^2$ and the facts that $\int_{-\infty}^\infty(f(s,\theta)-f(s,\theta_0))^2\pi(ds)>0$, for any $\theta\neq\theta_0$. □

Proof of Theorem 2.4. Recalling $T(\lambda x)\leq v(\lambda)\,T_1(x)$, we have

$$\begin{aligned}\frac{1}{nv^2(d_n)}\sum_{t=1}^n[T(x_t)+T^2(x_t)]\quad&\leq\quad\frac1n\sum_{t=1}^n\Big[T_1\Big(\frac{x_t}{d_n}\Big)+T_1^2\Big(\frac{x_t}{d_n}\Big)\Big]\\&\to_D\quad\int_0^1[T_1(G(t))+T_1^2(G(t))]dt,\end{aligned}\tag{53}$$

where the convergence in distribution comes from Berkes and Horváth (2006). This proves (4) with $\kappa_n=\sqrt{n}v(d_n)$ due to the local integrability of $T_1+T_1^2$. To prove (5), by letting $G(\lambda,x):=f(\lambda x,\theta)-f(\lambda x,\theta_0)-v(\lambda)g(x,\theta,\theta_0)$, we have

$$\begin{aligned}&\frac{1}{nv^2(d_n)}\sum_{t=1}^n(f(x_t,\theta)-f(x_t,\theta_0))^2\\&=\frac{1}{nv^2(d_n)}\sum_{t=1}^n\Big[G\Big(d_n,\frac{x_t}{d_n}\Big)+v(d_n)g\Big(\frac{x_t}{d_n},\theta,\theta_0\Big)\Big]^2\\&\geq\frac1n\sum_{t=1}^n g^2\Big(\frac{x_t}{d_n},\theta,\theta_0\Big)-|\Lambda_n|,\end{aligned}\tag{54}$$

where $\Lambda_n = \frac{2}{nv(d_n)} \sum_{t=1}^n G\Big(d_n, \frac{x_t}{d_n}\Big)\, g\Big(\frac{x_t}{d_n}, \theta, \theta_0\Big)$. The similar arguments as in the proof of (53) yield that

$$\frac{1}{n}\sum_{t=1}^n g^2\Big(\frac{x_t}{d_n}, \theta, \theta_0\Big) \quad \to_D \quad Z := \int_0^1 g^2(G(s), \theta, \theta_0) ds, \tag{55}$$

where $P(0 < Z < \infty) = 1$ due to Assumption 2.4. On the other hand, it follows from (10) that

$$\frac{1}{nv^2(d_n)}\sum_{t=1}^n G^2\Big(d_n, \frac{x_t}{d_n}\Big) \quad \le \quad \frac{o(1)}{nv^2(d_n)}\sum_{t=1}^n T^2(x_t) \le \frac{o(1)}{n}\sum_{t=1}^n T_1^2(x_t/d_n) = o_P(1).$$

Now (5) follows from (54), (55) and the fact that

$$\Lambda_n \le 2\Big|\frac{1}{nv^2(d_n)}\sum_{t=1}^n G^2\Big(d_n, \frac{x_t}{d_n}\Big)\Big|^{1/2} \Big|\frac{1}{n}\sum_{t=1}^n g^2\Big(\frac{x_t}{d_n}, \theta, \theta_0\Big)\Big|^{1/2} \to_P 0.$$

□

Proof of Theorem 2.5. See the Supplemental Material of this paper. □

Proof of Theorem 3.1. It is readily seen that

$$\begin{aligned}
\dot{Q}_n(\theta_0) &= -\sum_{t=1}^n \dot{f}(x_t, \theta_0)(y_t - f(x_t, \theta_0)) = -\sum_{t=1}^n \dot{f}(x_t, \theta_0) u_t, \\
\ddot{Q}_n(\theta) &= -\sum_{t=1}^n \dot{f}(x_t, \theta)\dot{f}(x_t, \theta)' - \sum_{t=1}^n \ddot{f}(x_t, \theta)(y_t - f(x_t, \theta)),
\end{aligned}$$

for any $\theta \in \Theta$. It follows from the conditions (i)–(iii) that

$$\sup_{\theta: \|D_n(\theta - \theta_0)\| \le k_n} \| (D_n^{-1})' \left[\ddot{Q}_n(\theta) - \ddot{Q}_n(\theta_0)\right] D_n^{-1} \| = o_P(1).$$

By using (iii) and (iv), we have $(D_n^{-1})' \dot{Q}_n(\theta_0) = O_P(1)$ and $(D_n^{-1})' \ddot{Q}_n(\theta_0)\, D_n^{-1} \to_D M$. As $M > 0$, a.s., $\lambda_{min}[(D_n^{-1})' \ddot{Q}_n(\theta_0)\, D_n^{-1}] \ge \eta_n$ with probability that goes to one for some $\eta_n > 0$, where $\lambda_{min}(A)$ denotes the smallest eigenvalue of A and $\eta_n \to 0$ can be chosen as slowly as required. Due to these facts, a modification of Lemma 1 in Andrews and Sun (2004) implies (13). Note that (iv)' implies (iv). The result $D_n(\hat{\theta}_n - \theta_0) \to_D M^{-1} Z$ follows from (13) and the continuing mapping theorem. □

Proof of Theorem 3.2. It suffices to verify the conditions (i)–(iii) and (iv)' of Theorem 3.1 with

$$D_n = \sqrt{n/d_n}\, \mathbf{I}, \quad Z = \Sigma^{1/2}\, \mathbf{N} L_G^{1/2}(1, 0), \quad M = \Sigma\, L_G(1, 0),$$

where $\mathbf{I}$ is an identity matrix, under Assumptions 3.1 and 3.2. In fact, as $n/d_n \to \infty$, we may take $k_n \to \infty$ such that $\theta : ||D_n(\theta-\theta_0)|| \le k_n$ falls in $\mathcal{N}_\delta(\theta_0) = \{\theta : ||\theta - \theta_0|| < \delta\}$. This, together with Lemma 8.6, yields (i)–(iii).

On the other hand, it follows from Lemma 8.2 with $g_1(x) = \alpha_3' \dot f(x, \theta_0)$ and $g_2(x) = \alpha_1' \dot f(x,\theta_0)\, \dot f(x,\theta_0)'\alpha_2$ that, for any $\alpha_i = (\alpha_{i1}, ..., \alpha_{im}) \in R^m, i = 1, 2, 3,$

$$
\begin{aligned}
&\Big\{\alpha_3'\, Z_n,\ \alpha_1' Y_n\, \alpha_2\Big\}\\
&= \Big\{ \Big(\frac{d_n}{n}\Big)^{1/2} \sum_{t=1}^n \alpha_3'\, \dot f(x_t,\theta_0) u_t,\ \frac{d_n}{n}\sum_{t=1}^n \alpha_1' \dot f(x_t,\theta_0)\, \dot f(x_t,\theta_0)'\alpha_2\Big\}\\
&\to_D \Big\{\tau\, N\, L_G^{1/2}(1,0), \alpha_1'\, M\, \alpha_2\Big\} =_D \Big\{\alpha_3'\, Z,\ \alpha_1'\, M\, \alpha_2\Big\},
\end{aligned}
\tag{56}
$$

where $\tau^2 = \int_{-\infty}^{\infty} [\alpha_3' \dot f(s,\theta_0)]^2 ds$, N is a standard normal random variable independent of $G(t)$ and we have used the fact that

$$
\begin{aligned}
\tau\, N &= \Big(\int_{-\infty}^{\infty} [\alpha_3' \dot f(s,\theta_0)]^2 ds\Big)^{1/2} N\\
&=_D \alpha_3' \Big(\int_{-\infty}^{\infty} \dot f(s,\theta_0) \dot f(s,\theta_0)' ds\Big)^{1/2} \mathbf{N} = \alpha_3'\, \Sigma^{1/2}\, \mathbf{N}.
\end{aligned}
$$

This proves (iv)'. □

Proof of Theorem 3.3. As in the proof of Theorem 3.2, it suffices to verify the conditions (i)–(iii) and (iv)' of Theorem 3.1 with

$$
D_n = a(n)\mathbf{I}, \quad Z = \Sigma_\pi^{1/2}\, \mathbf{N}\Pi_\beta, \quad M = \Sigma_\pi\, \Pi_\beta,
$$

under Assumptions 3.1* and 3.2*. The details are similar to that of Theorem 3.2, with Lemma 8.2 replaced by Lemma 8.3, and hence are omitted. □

Proof of Theorem 3.4. It suffices to verify the conditions (i)–(iii) and (iv)' of Theorem 3.1 with $D_n = \text{diag}(\sqrt{n}\dot v_1(d_n), ..., \sqrt{n}\dot v_m(d_n))$,

$$
Z = \int_0^1 \dot h(G(t), \theta_0) dU(t), \quad M = \int_0^1 \Psi(t)\Psi(t)' dt,
$$

under Assumptions 3.3 and 3.4. Recalling $\sup_{1\le i,j\le m} |\frac{v(d_n)\, \ddot v_{ij}(d_n)}{\dot v_i(d_n)\, \dot v_j(d_n)}| < \infty$, it follows from (46) that

$$
\begin{aligned}
&\frac{1}{n\dot v_i(d_n)\dot v_j(d_n)} \sup_{\theta\in\mathcal{N}_\delta(\theta_0)} \sum_{t=1}^n \big|\ddot f_{ij}(x_t,\theta)[f(x_t,\theta) - f(x_t,\theta_0)]\big|\\
&\le \frac{C}{n\ddot v_{ij}(d_n) v(d_n)} \sup_{\theta\in\mathcal{N}_\delta(\theta_0)} \sum_{t=1}^n \big|\ddot f_{ij}(x_t,\theta)[f(x_t,\theta) - f(x_t,\theta_0)]\big| = o_P(1),
\end{aligned}
$$

for any $1 \leq i, j \leq m$. On the other hand, as $\sqrt{n}\dot{v}_i(d_n) \to \infty$ for all $1 \leq i \leq m$, we may take $k_n \to \infty$ such that $\theta : ||D_n(\theta - \theta_0)|| \leq k_n$ falls in $\mathcal{N}_\delta(\theta_0) = \{\theta : ||\theta - \theta_0|| < \delta\}$. These facts imply that

$$\sup_{\theta:\|D_n(\theta-\theta_0)\|\leq k_n} \| (D_n^{-1})' \sum_{t=1}^{n} \ddot{f}(x_t,\theta) \left[f(x_t,\theta) - f(x_t,\theta_0)\right] D_n^{-1} \| = o_P(1).$$

This proves the required (ii). The proofs of (i) and (iii) are similar, we omit the details. Finally (iv)' follows from Lemma 8.8 with

$$g(x,\theta_0) = \alpha_3'\dot{h}(x,\theta_0) \quad \text{and} \quad g_1(x,\theta_0) = \alpha_1'\dot{h}(x,\theta_0)\dot{h}(x,\theta_0)'\alpha_2.$$

□

Proofs of Theorems 4.1–4.3. See the Supplemental Material of this paper. □

REFERENCES

Andrews, D. W. K. and Sun, Y., 2004. Adaptive local polynomial Whittle estimation of long-range dependence. Econometrica, 72, 569-614.

Boden, T. A., Marland, G. and Andres, R. J., 2009. Global, Regional, and National Fossil-Fuel CO_2 Emissions. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A..

Bae, Y. and de Jong, R., 2007. Money demand function estimation by nonlinear cointegration. Journal of Applied Econometrics, 22(4), 767–793.

Bae, Y. Kakkar, V. and Ogaki, M., 2004. Money demand in Japan and the liquidity trap. ohio State University Department of Economics Working Paper #04–06.

Bae, Y., Kakkar, V., and Ogaki, M., 2006. Money demand in Japan and nonlinear cointegration. Journal of Money, Credit, and Banking, 38(6), 1659–1667.

Bates, D. M. and Watts, D. G., 1988. Nonlinear Regression Analysis and Its Applications. analysis, 60(4), 856865.

Berkes, I. and Horváth, L., 2006. Convergence of integral functionals of stochastic processes. Econometric Theory, 22(2), 304–322.

Chang, Y. and Park, J. Y., 2011. Endogeneity in nonlinear regressions with integrated time series, Econometric Review, 30, 51–87.

Chang, Y. and Park, J. Y., 2003. Index models with integrated time series. Journal of Econometrics, 114, 73–106.

Chang, Y., Park, J. Y. and Phillips, P. C. B., 2001. Nonlinear econometric models with cointegrated and deterministically trending regressors. Econometric Journal, 4, 1–36.

Chen, X., 1999. How Often Does a Harris Recurrent Markov Chain Recur? The Annals of Probability, 27, 1324–1346.

Choi, I. and Saikkonen, P., 2010. Tests for nonlinear cointegration. Econometric Theory, 26, 682–709.

Dijkgraaf, E. and Vollebergh, H. R., 2005. A test for parameter homogeneity in CO_2 panel EKC estimations. Environmental and Resource Economics, 32(2), 229–239.

de Jong, R., 2002. Nonlinear Regression with Integrated Regressors but Without Exogeneity, Mimeograph, Department of Economics, Michigan State University.

Feller, W., 1971. Introduction to Probability Theory and Its Applications, vol. II, 2nd ed. Wiley.

Gao, J. K., Maxwell, Lu, Z. and Tjøstheim, D., 2009. Specification testing in nonlinear and nonstationary time series autoregression. The Annals of Statistics, 37, 3893–3928.

Granger, C. W. J. and Teräsvirta, T., 1993. Modelling Nonlinear Economic Relationships. Oxford University Press, New York.

Hall, P. and Heyde, C. C., 1980. Martingale limit theory and its application. Probability and Mathematical Statistics. Academic Press, Inc.

Hansen, B. E., 1992. Covergence to stochastic integrals for dependent heterogeneous processes. Econometric Theory, 8, 489–500.

Ibragimov, R. and Phillips, P. C. B., 2008. Regression asymptotics using martingale convergence methods Econometric Theory, 24, 888–947.

Jeganathan, P., 2008. Limit theorems for functional sums that converge to fractional Brownian and stable motions. Cowles Foundation Discussion Paper No. 1649, Cowles Foundation for Research in Economics, Yale University.

Karlsen, H. A. and Tjøstheim, D., 2001. Nonparametric estimation in null recurrent time series. The Annals of Statistics, 29, 372–416.

Kurtz, T. G. and Prötter, P., 1991. Weak limit theorems for stochastic integrals and stochastic differential equations. The Annals of Probability, 19, 1035–1070.

Lai, T. L., 1994. Asymptotic theory of nonlinear least squares estimations. The Annals of Statistics, 22, 1917–1930.

Lai, T. L. and Wei, C. Z., 1982. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems, The Annals of

Statistics, 10, 154–166.

List, J. and Gallet, C., 1999. The environmental Kuznets curve: does one size fits all? Ecological Economics 31, 409–423.

Maddison, A., 2003. The World Economy: Historical Statistics, OECD.

Müller-Fürstenberger, G. and Wagner, M., 2007. Exploring the environmental Kuznets hypothesis: Theoretical and econometric problems. Ecological Economics, 62(3), 648–660.

Nummelin, E., 1984. *General Irreducible Markov Chains and Non-negative Operators.* Cambridge University Press.

Orey, S., 1971. Limit Theorems for Markov Chain Transition Probabilities. Van Nostrand Reinhold, London.

Piaggio, M. and Padilla, E., 2012. CO_2 emissions and economic activity: Heterogeneity across countries and non-stationary series. Energy policy, 46, 370–381.

Park, J. Y. and Phillips, P. C. B., 1999. Asymptotics for nonlinear transformation of integrated time series. Econometric Theory, 15, 269–298.

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

Phillips, P. C. B., 2001. Descriptive econometrics for non-stationary time series with empirical illustrations. Cowles Foundation discuss paper No. 1023.

Phillips, P. C. B. and Perron, P., 1988. Testing for a unit root in time series regression. Biometrika, 75(2), 335–346.

Phillips, P. C. B. and Solo, V., 1992. Asymptotics for linear processes. The Annals of Statistics, 971–1001.

Teräsvirta, T., Tjøstheim, D. and Granger, C. W. J., 2011. Modelling Nonlinear Economic Time Series, Advanced Texts in Econometrics.

Shi, X. and Phillips, P. C. B., 2012. Nonlinear cointegrating regression under weak identification. Econometric Theory, 28, 509–547.

Skouras, K., 2000. Strong consistency in nonlinear stochastic regression models. The Annals of Statistics, 28, 871–879.

Wagner, M., 2008. The carbon Kuznets curve: A cloudy picture emitted by bad econometrics? Resource and Energy Economics, 30(3), 388–408.

Wang, Q., 2013. Martingale limit theorems revisited and non-linear cointegrating regression. *Econometric Theory*, forthcoming.

Wang, Q., Lin, Y. X. and Gulati, C. M., 2003. Asymptotics for general fractionally integrated processes with applications to unit root tests. Econometric Theory, 19, 143–164.

Wang, Q. and Phillips, P. C. B., 2009. Asymptotic theory for local time density estimation and nonparametric cointegrating regression. Econometric Theory, 25, 710–738.

Wang, Q. and Phillips, P. C. B., 2009. Structural nonparametric cointegrating regression. Econometrica, 77, 1901–1948.

Wang, Q. and Phillips, P. C. B., 2012. A specification test for nonlinear nonstationary models. The Annals of Statistics, 40, 727-758.

Wu, C. F., 1981. Asymptotic theory of nonlinear least squares estimations. The Annals of Statistics, 9, 501–513.

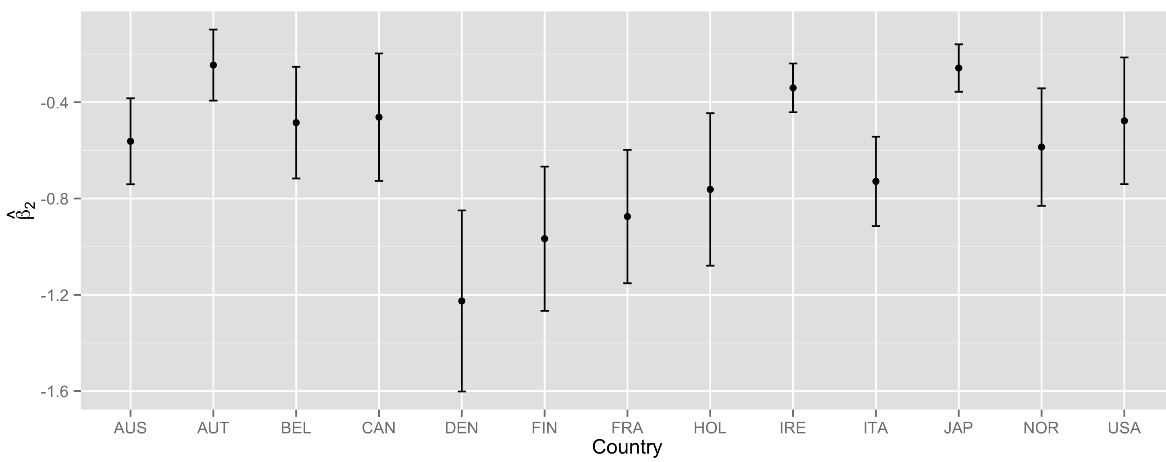

Figure 5: Estimates and 95% Confidence Intervals of $\hat{\beta}_{2n}$.