Nonlinear cointegrating regressions with nonstationary time series

Nigel Chan* and Qiying Wang

The University of Sydney

May 16, 2012

Abstract

This paper develops an asymptotic theory for a non-linear parametric co-integrating regression model. We establish a general framework for weak consistency that is easy to apply for various non-stationary time series, including partial sum of linear process and Harris recurrent Markov chain. We provide a limit distribution for the nonlinear least square estimator which significantly extends the previous work. We also introduce endogeneity in the model by allowing the error to be serially dependent and cross correlated with the regressor.

JEL Classification: C13, C22.

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

1   Introduction

The past few decades has witnessed significant developments in cointegration analysis. In particular, extensive researches have been focused on cointegration models with linear structure. Whilst it is convenient for practical implementation, nonetheless, it is restrictive, especially in the context of economics, often suggesting nonlinear responses with some unknown parameters. 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.

---

*Corresponding author. School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia. Tel.: +61-4-33010132. Fax.: +61-2-93514534. E-mail address: chanhiungai@gmail.com

A typical non-linear 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 data $\{y_t, x_t\}_{t=1}^n$, which may include non-stationary components, this paper is concerned with the nonlinear least square (NLS) estimation of the unknown parameters $\theta \in \Theta$. In this regard, Park and Phillips (1999, 2001), PP henceforth, considers x_t to be an integrated $I(1)$ process. Based on PP framework, Chang et. al (2001) introduced additional linear time trend term and stationary regressors into model (1). Chang and Park (2003) extends to nonlinear index models driven by integrated process. More recently, Choi and Saikkonen (2010), Gao, et al. (2009) and Wang and Phillips (2012) developed statistical tests for the existence of nonlinear cointegrating relation. Park and Chang (2011) allows the regressors x_t to be contemporaneous correlated with the regression errors u_t. Shi and Phillips (2012) extends the model (1) and incorporates a loading coefficient.

The present paper has a similar goal to the previous mentioned papers but offers more general results that we hope has some advantage. 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 non-stationary time series. The set of sufficient conditions are easy to apply to various non-stationary regressors, including partial sum of linear process and recurrent Markov chain. Furthermore, we provide a limit distribution for the NLS estimator $\hat{\theta}_n$. It deserves to mention that the routine in this paper to establish the limit distribution of $\hat{\theta}_n$ is different from previous work, e.g. Park and Phillips (2001). Roughly speaking, our routine is related to joint distributional convergence of a martingale under target and its conditional variances, rather than using classical martingale limit theorem by establishing the convergence in probability for the conditional variance. In nonlinear co-integrating regression, 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 non-stationary 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

2

distribution of the NLS estimator $\hat{\theta}_n$ are also established.

This paper is organized as follow. Section 2 presents our main results of weak consistency. The results on limit distribution of the NLS estimator are given in Section 3. Extension to endogeneity is presented in Section 4. Technical proofs are postponed to Section 6. Throughout the paper, we denote constants by $C, C_1, C_2, ...$ which may be different at each appearance, and assume that $\|x\| = (x_1^2 + ... + x_m^2)^{1/2}$ whenever $x = (x_1, ..., x_m)$. 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 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 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 estimate 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) having a martingale structure. In this situation, our sufficient conditions are closely related to those of Wu (1981), Lai (1994) and Skouras (2000), which provides a general framework. In comparison to the papers mentioned, our assumptions are easy to apply, particularly in non-linear cointegrating regression situation, as stated in two examples below. Extension to endogeneity between x_t and u_t is investigated in Section 4.

2.1 Main results

We make use of the following assumptions for the development of the 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*

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

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

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

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 $\theta \neq \theta_0$, where $\theta, \theta_0 \in \Theta$, there exist $n_0 > 0$ and $M_1 > 0$ such that

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

for all $n > n_0$.

THEOREM 2.1. *Under Assumption 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 $\kappa_n^2 n^{-1} = O_P(1)$, then $\hat{\sigma}_n^2 \to_P \sigma^2$, as $n \to \infty$.

Assumptions 2.1 and 2.2 are the same as 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), Assumption 2.3, which is related to the conditions on the regressor x_t, is more natural and easy to apply, particularly in the situation that $T(x)$ is integrable and the regressor x_t is an non-stationary time series, as stated in the following examples.

Example 1 (Partial sum of linear process). 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 iid random variables with $E\epsilon_0 = 0$, $E\epsilon_0^2 = 1$ and the characteristic function $\varphi(t)$ of ϵ_0 satisfies $\int_{-\infty}^{\infty} |\varphi(t)| dt < \infty$. 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

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

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. *If $T(x)$ is bounded and integrable and $\int_{-\infty}^{\infty}(f(s,\theta) - f(s,\theta_0))^2 ds > 0$ for all $\theta \neq \theta_0$, then (4) and (5) hold with $\kappa_n^2 = n/d_n$.*

Theorem 2.2 improves Theorem 4.1 of PP in two folds. First 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 allows x_t to satisfy **C2** with additional conditions on ϕ_k, that is, requires x_t to be a partial sum of short memory process. Secondly we remove the part (b) required in the definition of I-regular function given in their Definition 3.3. PP requires $f(x,\theta)$ to be piecewise smooth and we only require the $[f(s,\theta) - f(s,\theta_0)]^2$ to be integrable.

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 $A \in \mathcal{E}^+$,

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

where $\mathcal{E}^+ = \{A \in \mathcal{E} : \pi(A) > 0\}$ and $\tau_A = \inf\{n \geq 1 : x_n \in A\}$. As is well-known, 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. *If $T(x)$ is bounded and $\int_{-\infty}^{\infty} |T(x)|\pi(dx) < \infty$ and $\int_{-\infty}^{\infty}(f(s,\theta) - f(s,\theta_0))^2 \pi(ds) > 0$ for all $\theta \neq \theta_0$, then (4) and (5) hold with $\kappa_n^2 = a(n)$.*

Theorem 2.3 seems to be new to 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 common used in literature. The Harris recurrent Markov chain $\{x_k\}_{k\geq0}$ 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.

Remark 2.1. Although we assume the integrability, Theorem 2.1 allows for certain non-integrable $f(x,\theta)$. For instance, if $f(x,\theta) = \theta\, x^{\alpha}$, where $\alpha > 0$ is given, (3) holds with $h(x) = |x|$ and $T(x) = |x^{\alpha}|$. Under $x_t/d_n \Rightarrow_D G(t)$ on $D[0,1]$, where $G(t)$ is some continuous stochastic process, it follows from the continuous mapping theorem that

$$\frac{1}{n}\sum_{t=1}^{n}(x_t/d_n)^{2\alpha} \to_D \int_0^1 G(t)^{2\alpha}dt \tag{10}$$

on $D[0,1]$, which implies the required (4) and (5) with $\kappa_n^2 = n\, d_n^{2\alpha}$.

However, some modifications to our Theorem 2.1 are necessary for establishing the weak consistency of the $\hat{\theta}_n$, under more general non-integrable $f(x,\theta)$. In this regard, we follow PP to consider the asymptotic homogeneous function.

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 (see Appendix for definition of regularity) and $v(\lambda,\theta)$ belongs to one of the following cases:

Case 1 (i) $v(\lambda,\theta) = v(\lambda)$ and $v(\lambda)$ is bounded away from zero as $\lambda \to \infty$; (ii) for all $\theta \neq \theta_0$ and $\delta > 0$, $\int_{|s|\leq\delta}(h(s,\theta) - h(s,\theta_0))^2 ds > 0$.

Case 2 (i) for any $\bar{\theta} \neq \theta_0$ and $\bar{p},\bar{q} > 0$, there exist $\epsilon > 0$ and a neighborhood N of $\bar{\theta}$ such that as $\lambda \to \infty$

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

(ii) for all $\theta \neq \theta_0$ and $\delta > 0$, $\int_{|s|\leq\delta} h^2(s,\theta)ds > 0$.

THEOREM 2.4. *Suppose that $f(x,\theta)$ in model (1) has the structure (11), and in addition to Assumption 2.2, there exists a continuous Gaussian process $G(t)$ such that*

6

$x_{[nt],n} \Rightarrow_D G(t)$, 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. $\hat{\theta}_n \to_P \theta_0$.

Remark 2.2. The conditions on $f(x, \theta)$ given in Theorem 2.4 are the same as in PP, which is quite general including many commonly used regression functions. For example, $f(x, \theta) = \theta e^x/(1 + e^x), \theta \log|x|, \theta|x|^\alpha$ and $(|x|^\theta - 1)/\theta$. Our condition on the regressor x_t is much general than that of PP, where we only requires x_t/d_n converges to a continuous Gaussian process. This kind of weak convergence condition is quite weak and very likely close to necessary.

3 Limit Distribution

This section considers the limit distribution of $\hat{\theta}_n$. To 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. Under these notation, the first order Taylor expansion of the term $\dot{Q}_n$ yields that

$$\dot{Q}_n(\hat{\theta}_n) = \dot{Q}_n(\theta_0) + \ddot{Q}_n(\theta_n)(\hat{\theta}_n - \theta_0), \tag{13}$$

where θ_n lies in the line segment joining $\hat{\theta}_n$ and θ_0.

We start with some heuristic arguments in establishing the limit distribution of $\hat{\theta}_n$. To this end, assume that $\hat{\theta}_n$ satisfies (2), is an interior point of Θ and a consistency estimator of θ. Under these assumptions, we have $\dot{Q}_n(\hat{\theta}_n) = 0$. On the other hand, 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 $\theta \in \Theta$. Suppose there exists a sequence of constants $0 < \kappa_n \to \infty$ such that

$$\frac{1}{\kappa_n^2} \| \sum_{t=1}^{n} \left[\dot{f}(x_t, \theta_n) \dot{f}(x_t, \theta_n)' - \dot{f}(x_t, \theta_0) \dot{f}(x_t, \theta_0)' \right] \| \quad = \quad o_P(1); \tag{14}$$

$$\frac{1}{\kappa_n^2} \| \sum_{t=1}^{n} \ddot{f}(x_t, \theta_n) \left[f(x_t, \theta_n) - f(x_t, \theta_0) \right] \| \quad = \quad o_P(1); \tag{15}$$

$$\frac{1}{\kappa_n^2} \| \sum_{t=1}^{n} \ddot{f}(x_t, \theta_n) \, u_t \| \quad = \quad o_P(1); \tag{16}$$

$$\left\{ \kappa_n^{-1} \sum_{t=1}^{n} \alpha' \, \dot{f}(x_t, \theta_0) u_t, \; \kappa_n^{-2} \sum_{t=1}^{n} f(x_t, \theta_0) \right\} \; \to_D \; \left\{ \alpha' Y, \, Z \right\} \tag{17}$$

for any $\alpha' = (\alpha_1, ..., \alpha_m) \in R^m$; and

$$\left\{ \kappa_n^{-2} \sum_{t=1}^{n} \dot{f}_i(x_t, \theta_0) \dot{f}_j(x_t, \theta_0), \; \kappa_n^{-2} \sum_{t=1}^{n} f(x_t, \theta_0) \right\} \; \to_D \; \left\{ c_{i,j} \, Z, Z \right\}, \tag{18}$$

for all $1 \leq i, j \leq m$, where $\dot{f}_i = \partial f / \partial \theta_j$, $c_{i,j}$ are constants, Y is a random vector and Z is a positive random variable. We then have

$$\kappa_n^2 (\hat{\theta}_n - \theta_0) \; \to_D \; A^{-1}(\theta_0) Y / Z, \tag{19}$$

where $A^{-1}(\theta_0)$ is an inverse matrices of $A(\theta_0)$ defined by $A(\theta_0) = (c_{i,j})_{1 \leq i,j \leq m}$.

Indeed, it follows from (14)-(17) that

$$\frac{1}{\kappa_n^2} \| \ddot{Q}_n(\theta_n) - A_n(\theta_0)) \| \quad = \quad o_P(1),$$

where $A_n(\theta_0) = \sum_{t=1}^{n} \dot{f}(x_t, \theta_0) \dot{f}(x_t, \theta_0)'$. Furthermore, by noting

$$\left| \frac{\sum_{t=1}^{n} \dot{f}_i(x_t, \theta_0) \dot{f}_j(x_t, \theta_0)}{\sum_{t=1}^{n} f(x_t, \theta_0)} - c_{i,j} \right| = o_P(1),$$

for all $1 \leq i, j \leq m$ due to (18), we have

$$\| \frac{A_n(\theta_0)}{\sum_{t=1}^{n} f(x_t, \theta_0)} - A(\theta_0) \| \quad = \quad o_P(1).$$

These facts, together with (13) and $\|\hat{\theta}_n - \theta_0\| = o_P(1)$, implies that

$$\frac{\ddot{Q}_n(\theta_n)(\hat{\theta}_n - \theta_0)}{\sum_{t=1}^{n} f(x_t, \theta_0)} \quad = \quad A(\theta_0) \, (\hat{\theta}_n - \theta_0) + o_P(1) = - \frac{\sum_{t=1}^{n} \dot{f}(x_t, \theta_0) u_t}{\sum_{t=1}^{n} f(x_t, \theta_0)}.$$

This, together with (16), implies that,

$$\kappa_n (\hat{\theta}_n - \theta_0) \; \to_D \; A^{-1}(\theta_0) Y / Z, \tag{20}$$

8

which yields (19).

We remark that the routine in this paper to establish the limit distribution of $\hat{\theta}_n$ is different from PP. In comparison to the paper mentioned, their condition AD3 requires proving convergence in probability of (18). It is equivalent to say that, under the PP's routine, one requires to show (at least under an enlarged probability space)

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} g(x_t) \rightarrow_P \int_{-\infty}^{\infty} g(s)ds\, L_W(1,0) \tag{21}$$

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 $W(t)$[1]. 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 (21), we only need to show that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} g(x_t) \rightarrow_D \int_{-\infty}^{\infty} g(s)ds\, L_W(1,0) \tag{22}$$

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

We now establish our results on convergence in distribution for the $\hat{\theta}_n$, mainly settling the conditions to ensure (14)–(18) hold. First consider the situation that $f(x,\theta)$ 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. *(i) $f(x,\theta_0), \dot{f}_i(x,\theta_0)$ and $\ddot{f}_{ij}(x,\theta_0)$, $1 \leq i,j \leq m$, are bounded and integrable real functions; (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) in addition to Assumption 2.1, $\dot{f}_i$ and $\ddot{f}_{ij}$, $1 \leq i,j \leq m$, satisfy (3) with $T(x)$ being bounded and integrable.*

---

[1]Here and below, the local time process $L_G(t,s)$ of a Gaussian process $G(t)$ is defined by

$$L_G(t,s) = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_0^t I\{|G(r) - s| \leq \epsilon\}dr.$$

9

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

$$\sqrt{n/d_n}(\hat{\theta}_n - \theta_0) \to_D \Sigma^{-1/2} \boldsymbol{N} L_G^{-1/2}(1,0), \tag{23}$$

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

$$G(t) = \begin{cases} W_{\mu-3/2}(t), & under \ C1, \\ W(t), & under \ C2. \end{cases} \tag{24}$$

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

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

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

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

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

Using the same routine and slight modification 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, (ii) x_t satisfies (9), that is, x_t is a β-regular Harris recurrent Markov chain. (iii) $\{u_k, \mathcal{F}_k\}_{k \geq 1}$ forms a martingale difference (iv) $\max_{k \geq m} |E(u_{k+1}^2 \mid \mathcal{F}_{nk}) - 1| \to 0, a.s.$ and for some $\delta > 0$, $\max_{k \geq 1} E(|u_{k+1}|^{2+\delta} \mid \mathcal{F}_{nk}) < \infty$. and $E(u_k \mid \mathcal{F}_{nk}) = 0$ for any $1 \leq k \leq n$, $n \geq 1$, where $\mathcal{F}_{nk} = \sigma(\mathcal{F}_k, x_1, ..., x_n)$.*

ASSUMPTION 3.2*. *(i) $f(x,\theta_0), \dot{f}_i(x,\theta_0)$ and $\ddot{f}_{ij}(x,\theta_0)$, $1 \leq i,j \leq m$, are bounded and integrable (w.r.t. invariant measure π) real functions; (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 $\theta \neq \theta_0$; (iii) in addition to Assumption 2.1, $\dot{f}_i$ and $\ddot{f}_{ij}$, $1 \leq i,j \leq m$, satisfy (3) with $T(x)$ being bounded and $\int_{-\infty}^\infty |T(x)|\pi(dx) < \infty$.*

THEOREM 3.2. *Under Assumptions 3.1* and 3.2* we have*

$$\sqrt{a(n)}(\hat{\theta}_n - \theta_0) \to_D \Sigma_\pi^{-1/2} \boldsymbol{N} \Pi_\beta^{-1/2} \tag{25}$$

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

$$E \exp\{-t \Pi_\beta^{-\beta}\} = \exp\left\{-\frac{t^\beta}{\Gamma(\beta+1)}\right\}, \quad t \geq 0. \tag{26}$$

10

Remark 3.2. Theorem 3.2 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 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* (iv) 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 $1 \le t \le n$. We leave this for future research.

We next consider the limit distribution of $\hat{\theta}_n$ when regression function is asymptotically homogeneous function. In this regard, the result is essentially the same as in PP, except we use more general regressor x_t. We list here for convenience of reading. The technical proofs are very similar to those of PP and hence the details are omitted.

THEOREM 3.3. *Suppose that $f, \dot{f}$ and $\ddot{f}$ all have the structure (11) under Case 1, and in addition to Assumptions 2.2, there exists a vector of continuous Gaussian process (G, U) such that $(x_{[nt],n}, n^{-1} \sum_{i=1}^{[nt]} u_i) \Rightarrow_D (G, U)$, on $D[0,1]^2$, where $x_{t,n} = x_t/d_n$ with $d_n^2 = var(x_n) \to \infty$. Further assume that $\|(\dot{v} \otimes \dot{v})^{-1} v \ddot{v}\| < \infty$ and $\int_{|s| \le \delta} \dot{h}(s, \theta_0) \dot{h}(s, \theta_0)' ds > 0$ for all $\delta > 0$. Then we have*

$$\sqrt{n} \dot{v}(d_n)'(\hat{\theta}_n - \theta_0) \to_D \left(\int_0^1 \Psi(t)\Psi(t)' dt \right)^{-1} \int_0^1 \Psi(t)\, dU(t) \tag{27}$$

on $D[0,1]$, as $n \to \infty$, where $\Psi(t) = \dot{h}(G(t), \theta_0)$.

Remark 3.4. Except the joint convergence, the other conditions in establishing Theorem 3.3 are the same as in PP. The joint convergence under present paper is quite natural as $f, \dot{f}$ and $\ddot{f}$ all are asymptotically homogeneous function. Indeed, under $\dot{f}(\lambda x, \theta) \sim \dot{v}(\lambda) \dot{h}(x, \theta)$, one can easily obtain the following asymptotics under our joint convergence.

$$\frac{1}{\sqrt{n} \dot{v}(d_n)} \sum_{t=1}^n \dot{f}(x_t, \theta_0) u_t \to_D \int_0^1 \dot{h}(G(t), \theta_0) dU(t) \tag{28}$$

on $D[0,1]$, which is required in the proof of (27). See, e.g., Kurtz and Protter (1991) and Hansen (1992).

Remark 3.5. If we further assume that the $U(t)$ and $G(t)$ is 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 (27) will become mixed normal, specifically, we will have,

$$\sqrt{n}\dot{v}(d_n)'(\hat{\theta}_n - \theta_0) \to_D \left(\int_0^1 \Psi(t)\Psi(t)'dt \right)^{-1/2} \mathbf{N} \tag{29}$$

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

As in PP, the independence between the asymptotic order $v(\lambda, \theta)$ and the parameters θ can be removed if we impose extra conditions on the function $f(x, \theta)$.

Case 3 For any $\bar{s} > 0$ given, there exists $\epsilon > 0$ such that as $\lambda \to \infty$,

1. $\left\| (\dot{v} \otimes \dot{v})(\lambda, \theta_0)^{-1} \left(\sup_{|s| \le \bar{s}} |\ddot{f}(\lambda s, \theta_0)| \right) \right\| \to 0$,

2. $\lambda^{-1+\epsilon} \left\| (\dot{v} \otimes \dot{v})(\lambda, \theta_0)^{-1} \left(\sup_{|s| \le \bar{s}} \sup_{\theta \in \mathcal{N}(\epsilon, \lambda)} |\ddot{f}(\lambda s, \theta)| \right) \right\| \to 0$

3. $\lambda^{-1+\epsilon} \left\| (\dot{v} \otimes \dot{v} \otimes \dot{v})(\lambda, \theta_0)^{-1} \left(\sup_{|s| \le \bar{s}} \sup_{\theta \in \mathcal{N}(\epsilon, \lambda)} |\dddot{f} f(\lambda s, \theta)| \right) \right\| \to 0$

where $\mathcal{N}(\epsilon, \lambda) = \{\theta : \|\dot{v}(\lambda, \theta_0)'(\theta - \theta_0)\| \le \lambda^{-1+\epsilon}\}$.

THEOREM 3.4. *Suppose that $f, \dot{f}$ and $\ddot{f}$ all have the structure (11) under Case 3, and in addition to Assumptions 2.2, there exists a vector of continuous Gaussian process (G, U) such that $(x_{[nt],n}, n^{-1}\sum_{i=1}^{[nt]} u_i) \Rightarrow_D (G, U)$, on $D[0,1]^2$, where $x_{t,n} = x_t/d_n$ with $d_n^2 = var(x_n) \to \infty$. Further assume that $\int_{|s| \le c} \dot{h}(s, \theta_0)\dot{h}(s, \theta_0)'ds > 0$ for all $c > 0$. Then we have*

$$\sqrt{n}\dot{v}(d_n, \theta_0)'(\hat{\theta}_n - \theta_0) \to_D \tau^{-1} \int_0^1 \dot{h}(G(t), \theta_0)dU(t) \tag{30}$$

on $D[0,1]$, as $n \to \infty$, where $\tau = \int_0^1 \dot{h}(G(s), \theta_0)\dot{h}(G(s), \theta_0)'ds$.

Remark 3.6. Again except the joint convergence, the conditions on $f(x_t, \theta)$ in Theorem 3.4 are the same as in PP. While they are hard to check in general, they allow for popular regression such as $f(x, \alpha, \beta) = \alpha|x|^\beta$. The idea of proof of Theorem 3.4 can be easily generalized from the Brownian Motion case in PP to general non-stationary time series. 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 iid innovations. In this section we provide some significant extension to these works.

ASSUMPTION 4.1. *(i) $\eta_i \equiv (\epsilon_i, \nu_i), i \in Z$ be a sequence of iid random vectors satisfying $E\eta_0 = 0$ and $E\|\eta_0\|^{2q} < \infty$, where $q \geq 1$; the characteristic function $\varphi(t)$ of ϵ_0 satisfies $\int_{-\infty}^{\infty} |\varphi(t)|^2 dt < \infty$ and $\int_{-\infty}^{\infty} |t|^3 |\varphi(t)|^m dt < \infty$ for some $m > 0$;*

*(ii) $x_t = \sum_{j=1}^{t} \xi_j$ where ξ_j is defined as in (6), that is, $\xi_j = \sum_{k=0}^{\infty} \phi_k \epsilon_{j-k}$ with the coefficient ϕ_k satisfying **C1** or **C2**;*

(iii) $u_t = \sum_{k=0}^{\infty} \psi_k \eta'_{t-k}$, where the coefficient $\psi_k = (\psi_{k1}, \psi_{k2})$ are assumed to satisfy $\sum_{k=0}^{\infty} k^2 \|\psi_k\|^2 < \infty$, $\sum_{k=0}^{\infty} \psi_k \neq 0$ 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), which is much general than that given in Chang and Park (2011). The conditions on the characteristics function $\varphi(t)$ is not very restrictive, see Remark 4 of Jeganathan (2008) for more details.

We have the following result.

THEOREM 4.1. *Let $f(x, \theta)$ in model (1) satisfy Assumption 2.1 with $T(x)$ being bounded and integrable. Then, under Assumption 4.1, 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, the limit distribution of $\hat{\theta}_n$ is given by

$$\sqrt{n/d_n}(\hat{\theta}_n - \theta_0) \to_D \Sigma^{-1} \Lambda^{1/2} \mathbf{N} L_G^{-1/2}(1, 0), \tag{31}$$

as $n \to \infty$, where

$$\Lambda = (2\pi)^{-1} \int \widehat{\dot{f}}(\mu) \widehat{\dot{f}}(\mu)' [Eu_0^2 + 2 \sum_{r=1}^{\infty} E(u_0 u_r e^{-i\mu x_r})] \, d\mu \tag{32}$$

and $\widehat{\dot{f}}(\mu) = \int e^{i\mu x} \dot{f}(x, \theta_0) dx$ and $\mathbf{N}$ is a standard normal random vector.

13

Remark 4.1. Theorem 4.1 is applicable for $f(x_t, \theta)$ being an integrable function. For asymptotically homogeneous function, it requires a comprehensive analysis of asymptotic theory of the functional $n^{-1} \sum_{t=1}^{n} h(x_t/d_n)u_t$ converging to a stochastic integral, i.e.

$$\frac{1}{\sqrt{n}} \sum_{t=2}^{n} h(x_t/\sqrt{n}, \theta_0)u_t \to_D \sigma_{uv} \int_0^1 \dot{h}(W(s), \theta_0)ds + \int_0^1 h(W(s), \theta_0)dU(s)$$

under certain condition on the derivative of $f(x, \theta)$. See, e.g. Theorem 4.3 of Ibragimov and Phillips (2008). Such development under various cases of asymptotic homogeneous functions $f(x, \theta)$ and regressor x_t is lengthy and we leave it for future research work.

Finally, we give the consistency result for the error estimator.

COROLLARY 4.1. *Under the assumptions in Theorem 4.1, we have $\hat{\sigma}_n^2 \to_P Eu_0^2$, as $n \to \infty$.*

5   Simulation

In this section, we investigate the finite sample performance of the NLS estimator $\hat{\theta}_n$ of nonlinear regression with endogeneity. Chang and Park (2011) performed simulation of similar model, but only consider 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:

$$x_t = x_{t-1} + \epsilon_t,$$
$$v_t = \sqrt{1 - \rho^2}\, w_t + \rho\epsilon_t$$
$$f(x, \theta) = \exp\{-\theta|x|\} \tag{33}$$

where $\{w_t\}$ and $\{\epsilon_t\}$ are i.i.d standard normal variable, and ρ is the correlation coefficient that controls the degree of endogeneity. The true value of θ_0 is set as $\theta_0 = 0.1$. The error structure is generated according to the following three scenarios

    S1: $u_t = v_t$
    S2: $u_t = \sum_{j=1}^{t} j^{-10} v_{t-j+1}$
    S3: $u_t = \sum_{j=1}^{t} j^{-4} v_{t-j+1}$

Scenario **S1** is considered by Chang and Park (2011), which eliminates the case which 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**.

In our simulation, we drawn samples of size $n = 200, 500$ to estimate NLS estimator and its t-ratios. Each simulation is run for 10,000 times, and the density of the NLS estimator is estimated using kernel method with normal kernel function. Our results are presented in three tables. Table 1 is the estimated density of centered distribution of NLS estimator. Table 2 is the estimated density of centered distribution scaled by the convergence rate of $n^{1/4}$. Table 3 is the t-ratios of the NLS estimator. In all tables, the left columns are simulation results base on 200 samples, while the right columns base on 500 samples. We expect that as ρ increases, the degree of endogeneity increases, and therefore the variance of the distribution will increase due to (31). We also expect that the more serial correlation of u_t, the higher the variance of the limit distribution due to the cross term appeared in (32).

Our simulation results largely corroborate with our theoretical results in Section 4. Firstly, comparing the S1–S3 curves in each plots, we can see that as the shape S3 curves have fat tails and lower peak than that of S1-S2. This verifies that high dependence of u_t on its own past will increase the variance of limit distribution. Secondly, comparing the shape of the curves for different values of ρ, we can see that the curves with $\rho = 0$ have highest peak, 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.

Finally, the sampling results for t-ratios are also much expected from our limit theories, that the NLS estimator asymptotically converges to mixed normal under all scenarios and different values of ρ. As can be seen in first row ($\rho = 0$) of Table 3, the curves overlap standard normal curve for both cases $n = 200, 500$, this indicates that the t-ratios converges to a standard normal distribution rapidly when u_t is independent of x_t. However, for extreme case of complete endogeneity $\rho = 1$, the distribution is little away from standard normal when the sample size is small. Even when n increases, the limit distribution curve still does not completely overlap the standard normal curve. The deviation from normal is the most serious for S3 curve. From this we can see that the serial correlation of u_t and its cross dependence with the regressor x_t have significant impact on the convergence rate of the NLS estimator, and in particular, the finite sampling behavior when n is small.

15

Table 1: Density estimate of $\hat{\theta}_n$ (unscaled).
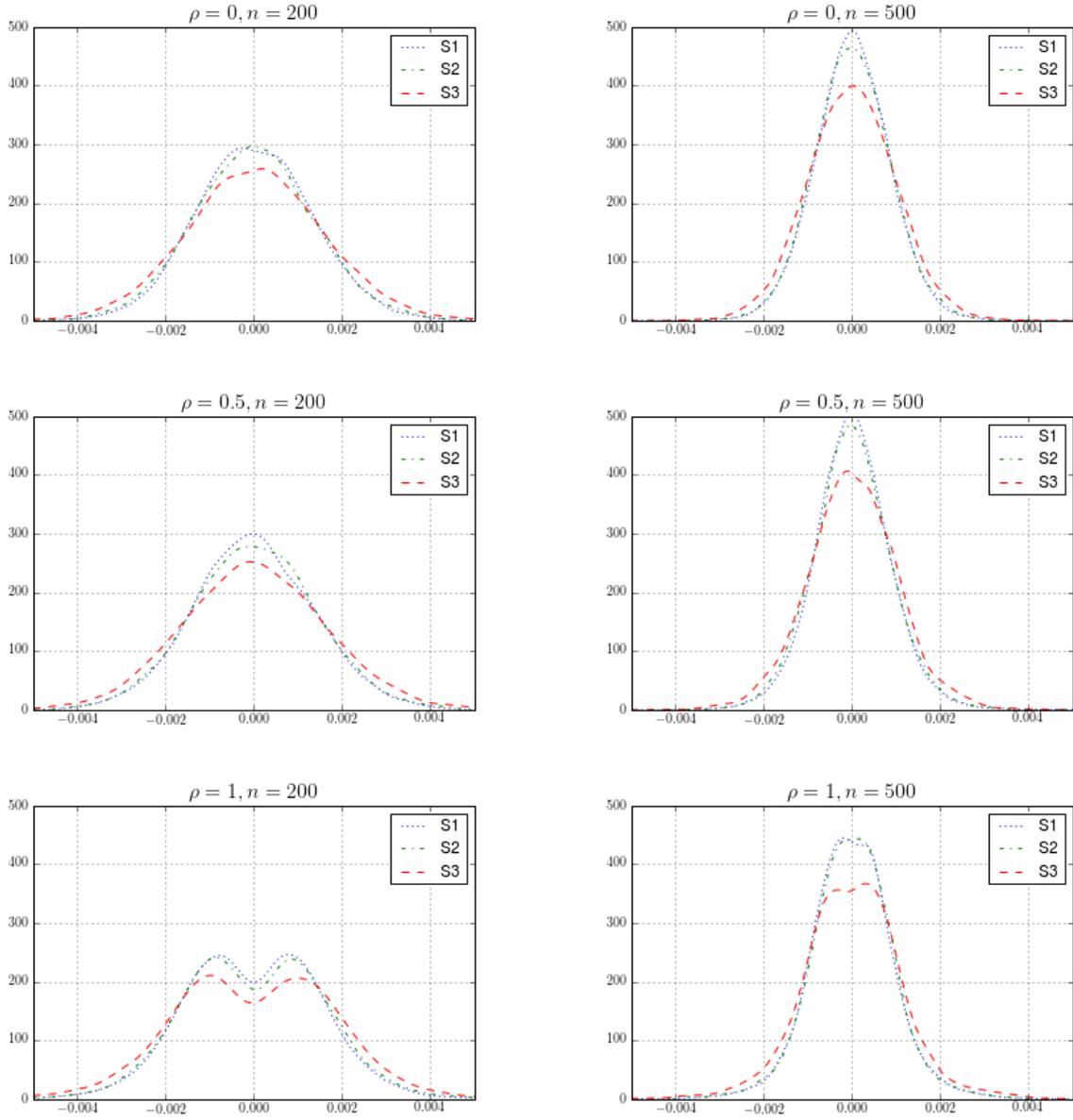
6 Conclusion

In this paper, we establish an asymptotic theory for a nonlinear parametric cointegrating 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 the partial sum of linear process and Harris recurrent Markov chain. Limit distribution of $\hat{\theta}_n$ is also established, and thus significantly extend previous works.

Table 2: Density estimate of $\hat{\theta}_n$ (scaled by $n^{1/4}$).
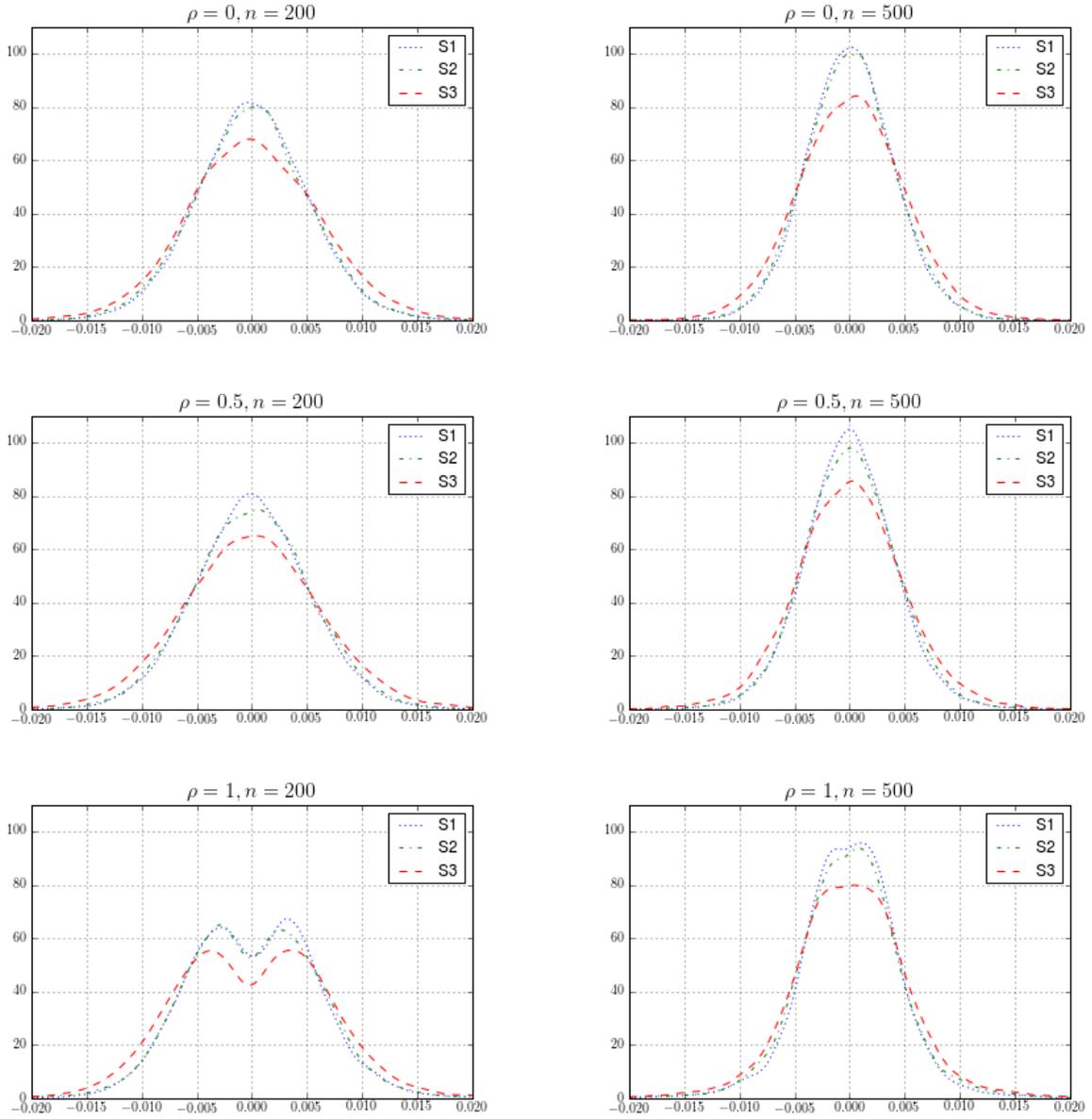
Furthermore, we introduce endogeneity to our model by allowing the error term serially dependent on itself, and cross-dependent on the regressor. We show that the limit distribution of $\hat{\theta}_n$ under the endogeneity situation is different from that with martingale error structure. This result is of interest in the applied econometric research area.

Table 3: Density of $\hat{\theta}_n$ t-ratios.
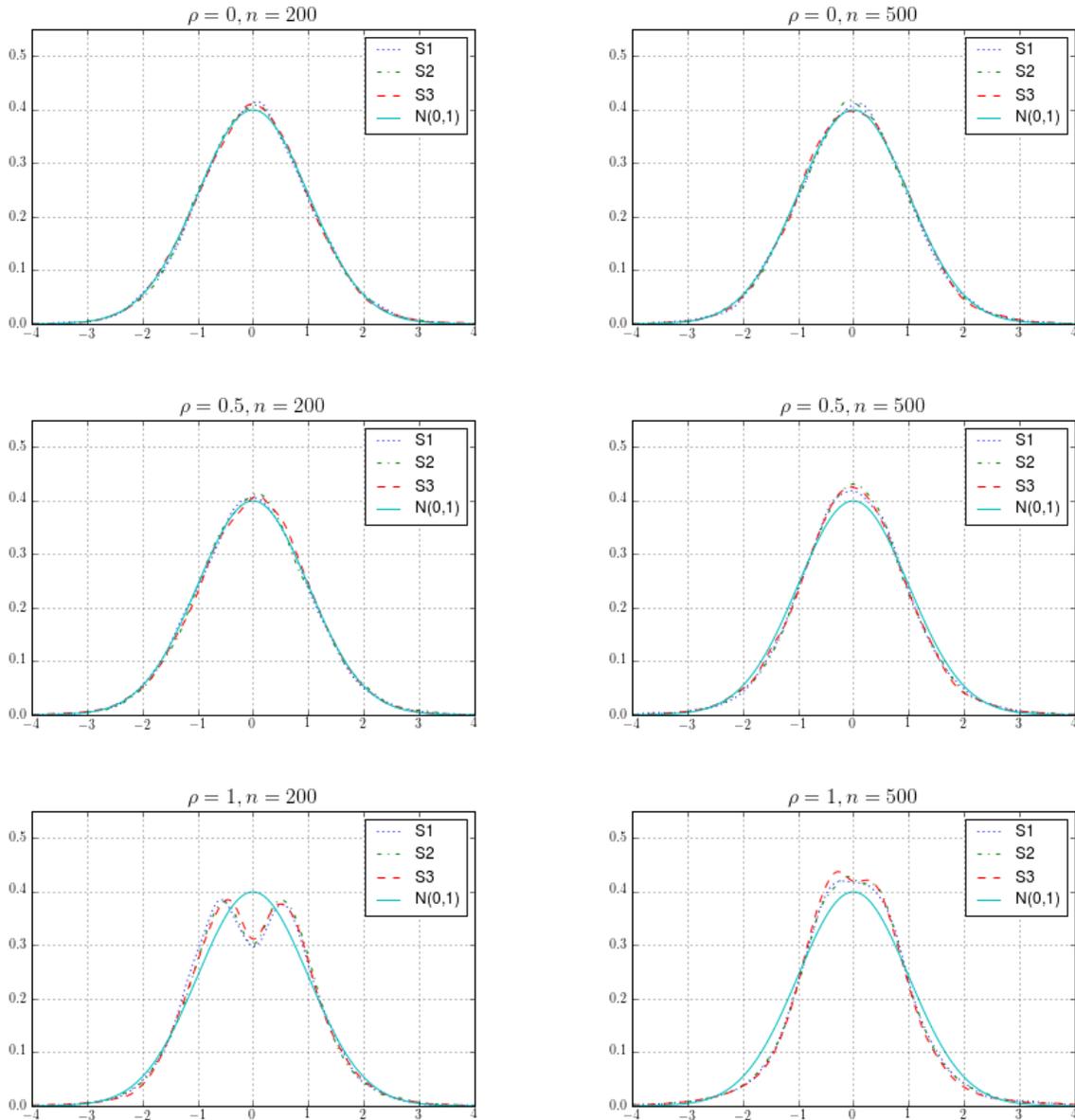
7 Appendix A: Proofs of main results

This section provides proofs of the main results. We start with some preliminaries, which list the limit theorems in common use in the proofs of main results.

7.1 Preliminaries

Denote $\mathcal{N}_\delta(\pi_0) = \{\theta : \|\theta - \pi_0\| < \delta\}$, where $\pi_0 \in \Theta$ is fixed.

LEMMA 7.1. *Under (4) and Assumption 2.1, we have*

$$\sup_{\theta \in \mathcal{N}_\delta(\pi_0)} \kappa_n^{-2} \sum_{t=1}^n \left(|f(x_t, \theta) - f(x_t, \pi_0)| + |f(x_t, \theta) - f(x_t, \pi_0)|^2 \right) \quad \to_P \quad 0 \tag{34}$$

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

$$\sup_{\theta \in \mathcal{N}_\delta(\pi_0)} \kappa_n^{-2} \sum_{t=1}^n |[f(x_t, \theta) - f(x_t, \pi_0)]| \, |u_t| \to_P 0 \tag{35}$$

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

Proof. Simple and the details are omitted. □

LEMMA 7.2. *Under Assumptions 3.1 (i) and 3.2, we have*

$$\frac{d_n}{n} \sup_{\theta \in \mathcal{N}_\delta(\theta_0)} \sum_{t=1}^n \left| \dot{f}_i(x_t, \theta) \, \dot{f}_j(x_t, \theta) - \dot{f}_i(x_t, \theta_0) \, \dot{f}_j(x_t, \theta_0) \right| \quad \to_P \quad 0, \tag{36}$$

$$\frac{d_n}{n} \sup_{\theta \in \mathcal{N}_\delta(\theta_0)} \sum_{t=1}^n \left| \ddot{f}_{ij}(x_t, \theta) \left[f(x_t, \theta) - f(x_t, \theta_0) \right] \right| \quad \to_P \quad 0. \tag{37}$$

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

$$\frac{d_n}{n} \sup_{\theta \in \mathcal{N}_\delta(\theta_0)} \left| \sum_{t=1}^n \ddot{f}_{ij}(x_t, \theta) \, u_t \right| \to_P 0 \tag{38}$$

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

The results are still true if Assumption 3.1 (i) and 3.2 are replaced by 3.1* (i) and 3.2* respectively, and d_n/n is replaced by $a(n)^{-1}$.

Proof. First note that, by Assumption 3.2,

$$\sup_{\theta \in \Theta} |f_i(x_t, \theta)| \le \sup_{\theta \in \Theta} |\dot{f}_i(x_t, \theta) - \dot{f}_i(x_t, \theta_0)| + |\dot{f}_i(x_t, \theta_0)|$$

$$\le \sup_{\theta \in \Theta} h(\|\theta - \theta_0\|) T(x_t) + |\dot{f}_i(x_t, \theta_0)| \le C$$

It follows that

$$\left| \dot{f}_i(x_t, \theta) \, \dot{f}_j(x_t, \theta) - \dot{f}_i(x_t, \theta_0) \, \dot{f}_j(x_t, \theta_0) \right|$$

$$\le \quad \left| \dot{f}_i(x_t, \theta) \right| \left| \dot{f}_j(x_t, \theta) - \dot{f}_j(x_t, \theta_0) \right| + \left| \dot{f}_j(x_t, \theta_0) \right| \left| \dot{f}_i(x_t, \theta) - \dot{f}_i(x_t, \theta_0) \right|$$

$$\le \quad C \left| \dot{f}_j(x_t, \theta) - \dot{f}_j(x_t, \theta_0) \right| + C_1 \left| \dot{f}_i(x_t, \theta) - \dot{f}_i(x_t, \theta_0) \right|$$

19

Therefore, (36) follows immediately from Lemma 7.1. (37) is similar to (36) and thus we omit the details.

For (38), note that, for any $\theta \in \mathcal{N}_\delta(\theta_0)$, $1 \leq i, j \leq m$,

$$\left| \sum_{t=1}^n \ddot{f}_{ij}(x_t, \theta) u_t \right| \leq \left| \sum_{t=1}^n [\ddot{f}_{ij}(x_t, \theta) - \ddot{f}_{ij}(x_t, \theta_0)] u_t \right| + \left| \sum_{t=1}^n \ddot{f}_{ij}(x_t, \theta_0) u_t \right|$$

Then by (35) in Lemma 7.1 (with $\pi_0 = \theta_0$) and Lemma 7.3 below, we obtain the required (38).

Finally, it can be easily seen that if Assumption 3.1 (i) and 3.2 are replaced by 3.1* (i) and 3.2* and d_n/n replaced by $a(n)^{-1}$, the above arguments will hold true, due to Lemma 7.5 instead of Lemma 7.3. $\square$

LEMMA 7.3. *Suppose Assumption 3.1 holds. Then, for any $g(x)$ and $g_1(x)$ satisfying $\int_{-\infty}^{\infty}(|g(x)| + |g(x)|^{4+\gamma})dx < \infty$ for some $\gamma > 0$, $\int_{-\infty}^{\infty}(|g_1(x)| + |g_1(x)|^2)dx < \infty$, $\int_{-\infty}^{\infty} g(x)dx \neq 0$ and $\int_{-\infty}^{\infty} g_1(x)dx \neq 0$, we have*

$$\left\{ \left(\frac{d_n}{n}\right)^{1/2} \sum_{t=1}^n g(x_t) u_t, \frac{d_n}{n} \sum_{t=1}^n g_1(x_t) \right\} \to_D \left\{ \tau_1 N L_G^{1/2}(1,0), \tau_2 L_G(1,0) \right\}, \quad (39)$$

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

Proof. Theorem 2.2 of Wang (2011) provides (39) with $g_1(x) = g^2(x)$. It is not difficult to see that (39) still holds for general $g_1(x)$. We omit the details. $\square$

LEMMA 7.4. *Under Assumption 4.1, for any bounded $g_i(x)$, $i = 1, 2$, satisfying $\int_{-\infty}^{\infty} |g_i(x)|dx < \infty$ and $\int_{-\infty}^{\infty} g_i(x)dx \neq 0$, then*

$$\sum_{t=1}^n g_1(x_t)|u_t| = O_P(n/d_n) \quad (40)$$

and

$$\left\{ (n/d_n)^{-1/2} \sum_{t=1}^n g_1(x_t)u_t, (n/d_n)^{-1} \sum_{t=1}^n g_2(x_t) \right\}$$
$$\to_D \left\{ \tau_3 N L_G^{1/2}(1,0), \tau_4 L_G(1,0) \right\} \quad (41)$$

where $\tau_3^2 = (2\pi)^{-1} \int_{-\infty}^{\infty} \hat{g}_1(\mu)^2 [Eu_0^2 + 2\sum_{r=1}^{\infty} E(u_0 u_r e^{-i\mu x_r})] d\mu$, $\hat{g}_1(\mu) = \int_{-\infty}^{\infty} e^{i\mu x} g_1(x)dx$ and $\tau_4 = \int_{-\infty}^{\infty} g_2(s)ds$.

Proof. See Theorem 3 and 5 of Jeganathan (2008). $\square$

20

LEMMA 7.5. *Under Assumption 3.1* (i), 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*

$$\sum_{t=1}^{n} g(x_t) = O_P[a(n)], \quad \left[\sum_{t=1}^{n} g(x_t)\right]^{-1} = O_P[a(n)^{-1}] \tag{42}$$

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

$$\left\{ a(n)^{-1/2} \sum_{t=1}^{n} g_1(x_t) u_t, \; a(n)^{-1} \sum_{t=1}^{n} g_2(x_t) \right\} \to_D \left\{ \tau_5 \, N \, \Pi_\beta^{1/2}, \tau_6 \, \Pi_\beta \right\} \tag{43}$$

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

   Proof. The proof of (42) sees Theorem 2.1 of Chen (1999). To prove (43), we first impose an additional assumption that $g_2(x) = g_1^2(x)$. Denote

$$\Delta_n^2 = a(n)^{-1} \sum_{t=1}^{n} g_1^2(x_t), \quad Z_{nt} = a(n)^{-1/2} g_1(x_t) \Delta_n^{-1}, \quad W_n = \sum_{t=1}^{n} Z_{nt} u_k \tag{44}$$

Recalling that $E(u_t | \mathcal{F}_{nt}) = 0$, it is readily seen that given $\{x_1, ..., x_n\}$, $\{Z_{nt} u_t, \mathcal{F}_{nt}\}_{t=1}^{n}$ forms a martingale difference sequence. The result (43) will follow if we prove,

$$\sup_x \left| P\big(W_n \leq x \,|\, x_1, ..., x_n\big) - \Phi(x) \right| \to_P 0 \tag{45}$$

Indeed, by noting that Δ_n^2 is measurable with respect to $\sigma(x_1, ...x_n)$, we have, for any $\alpha, \gamma \in R$,

$$\left| E\big[e^{i\alpha W_n + i\beta \Delta_n^2}\big] - e^{-\frac{1}{2}\alpha^2} E\big[e^{i\beta \tau_5 \Pi_\gamma}\big] \right|$$
$$\leq E\left| E\big(e^{i\alpha W_n} \,|\, x_1, ..., x_n\big) - e^{-\frac{1}{2}\alpha^2} \right| + e^{-\frac{1}{2}\alpha^2} \left| E e^{i\gamma \Delta_n^2} - E e^{i\gamma \tau_5 \Pi_\beta} \right| \to 0$$

by dominated convergence theorem, due to (45) and $\Delta_n^2 \to_D \tau_5 \Pi_\beta$ (see, e.g., Theorem 2.3 of Chen (1999)). This implies that

$$\{W_n, \Delta_n^2\} \to_D \{N, \tau_5 \Pi_\beta\}$$

where N is a standard normal random variable independent of Π_β. Hence, by continuous mapping theorem, we have

$$\left\{ a(n)^{-1/2} \sum_{t=1}^{n} g_1(x_t) u_t, \; a(n)^{-1} \sum_{t=1}^{n} g_1^2(x_t) \right\} = \left\{ \Delta_n W_n, \Delta_n^2 \right\} \to_D \{\tau_5^{1/2} N \Pi_\beta^{1/2}, \tau_5 \Pi_\beta\}$$

which implies the required (43).

We now prove (45). By Theorem 3.9 ((3.75) there) in Hall and Heyde (1980) with $\delta = q/2 - 1$ that

$$\sup_x \left| P\big(W_n \leq x \,|\, x_1, ..., x_n\big) - \Phi(x) \right| \leq A(\delta)\mathcal{L}_n^{1/(1+q)} \quad a.s.$$

where $A(\delta)$ is a constant depending only on δ and $q > 2$, and (set $\mathcal{F}_n^* = \sigma(x_1, ..., x_n)$)

$$\mathcal{L}_n = \Delta_n^{-q} \sum_{k=1}^{n} |Z_{nk}|^q E(|u_k|^q|\mathcal{F}_n^*) + E\left[\left| \Delta_n^{-2} \sum_{k=1}^{n} Z_{nk}^2 [E(u_k^2|\mathcal{F}_{nk}) - 1] \right|^{q/2} \Big| \mathcal{F}_n^* \right].$$

Recall from Assumption 3.2 (iv) and the fact that $\Delta_n^2 = \sum_{k=1}^{n} Z_{nk}^2$, we have,

$$E\left[\left| \Delta_n^{-2} \sum_{k=1}^{n} Z_{nk}^2 [E(u_k^2|\mathcal{F}_{nk}) - 1] \right|^{q/2} \Big| \mathcal{F}_n^* \right] \to_P 0$$

by dominated convergence theorem. Hence, routine calculations show that

$$\mathcal{L}_n \leq C\,\Delta_n^{-(q-2)}\, a(n)^{-(q-2)/2} + o_P(1) = o_P(1)$$

because $\Delta_n^{-2} = O_P(1)$ by (42) and $q > 2$. This proves (45), which implies that (43) holds true with $g_2(x) = g_1^2(x)$. Finally, note that, for any $a, b \in R$,

$$a(n)^{-1} \sum_{t=1}^{n} \left\{ ag_1^2(x_t) + bg_2(x_t) \right\} \to_D \int_{-\infty}^{\infty} \left[ag_1^2(s) + bg_2(s) \right] \pi(ds)\, \Pi_\beta$$

due to Theorem 2.3 of Chen (2009), which implies that

$$\left\{ a(n)^{-1} \sum_{t=1}^{n} g_1^2(x_t),\ a(n)^{-1} \sum_{t=1}^{n} g_2(x_t) \right\} \to_D \left\{ \int_{-\infty}^{\infty} g_1^2(s)\pi(ds)\, \Pi_\beta,\ \int_{-\infty}^{\infty} g_2(s)\pi(ds)\, \Pi_\beta, \right\}$$

Hence, by continuous mapping theorem,

$$\frac{\sum_{t=1}^{n} g_1^2(x_t)}{\sum_{t=1}^{n} g_2(x_t)} \to_P \int_{-\infty}^{\infty} g_1^2(s)\, \pi(ds) \Big/ \int_{-\infty}^{\infty} g_2(s)\, \pi(ds)$$

This shows that (43) is still true with general $g_2(x)$. □

LEMMA 7.6. *Let $D_n(\theta, \theta_0) = Q_n(\theta) - Q_n(\theta_0)$. Suppose that, for any $\delta > 0$,*

$$\liminf_{n \to \infty} \inf_{|\theta - \theta_0| \geq \delta} D_n(\theta, \theta_0) > 0 \quad in \quad probability, \tag{46}$$

then $\hat\theta_n \to_P \theta_0$.

Proof. See Lemma 1 of Wu (1981). □

7.2 Proof 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 $Q_n(\theta)$ in Θ, therefore, by Lemma 7.6, proving consistency is equivalent to showing that, for any $0 < \eta < 1$ and $\theta \neq \theta_0$, where $\theta, \theta_0 \in \Theta$, there exist $n_0 > 0$ and $M_1 > 0$ such that

$$P\left(\inf_{\theta \in \Theta \cap \mathcal{N}^c} D_n(\theta, \theta_0) \geq \kappa_n^2/M_1\right) \geq 1 - \eta \tag{47}$$

for all $n > n_0$.

Denote $\mathcal{N}_\delta(\pi_0) = \{\theta : \|\theta - \pi_0\| < \delta\}$. Note that $\Theta \cap \mathcal{N}^c$ is compact, by the finite covering property of compact set, (47) will follow if we prove that, for any fixed $\pi_0 \in \Theta \cap \mathcal{N}^c$,

$$\sup_{\theta \in \mathcal{N}_\delta(\pi_0)} \kappa_n^{-2} \left| D_n(\theta, \theta_0) - D_n(\pi_0, \theta_0) \right| \to_P 0 \tag{48}$$

as $n \to \infty$ first and then $\delta \to 0$, and for $\forall \epsilon > 0$, there exists $M_1 > 0$ and $n > n_0$ such that for all $n > n_0$,

$$P\left(D_n(\pi_0, \theta_0) \geq \kappa_n^2/M_1\right) \geq 1 - 2\eta. \tag{49}$$

The result (48) is simple. Indeed, by Lemma 7.1, it follows that, for each fixed $\pi_0 \in \Theta \cap \mathcal{N}^c$,

$$\sup_{\theta \in \mathcal{N}_\delta(\pi_0)} \kappa_n^{-2} \left| D_n(\theta, \theta_0) - D_n(\pi_0, \theta_0) \right|$$

$$= \sup_{\theta \in \mathcal{N}_\delta(\pi_0)} \kappa_n^{-2} \sum_{t=1}^n (f(x_t, \theta) - f(x_t, \pi_0))^2$$

$$- \sup_{\theta \in \mathcal{N}_\delta(\pi_0)} \kappa_n^{-2} \sum_{t=1}^n |f(x_t, \theta) - f(x_t, \pi_0)||u_t|$$

$$\to_P 0 \tag{50}$$

as $n \to \infty$ first and then $\delta \to 0$, which yields (48).

To prove (49), we write

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

Similar arguments as in Lemma 7.1 yields that

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

Hence, it follows from Assumption 2.3 that, for any $\eta > 0$, there exists $n_0 > 0$ and $M_1 > 0$ such that for all $n > n_0$,

$$P\Big(D_n(\pi_0, \theta_0) \geq M_1^{-1}\Big) \geq P\Big(\sum_{t=1}^{n}(f(x_t, \pi_0) - f(x_t, \theta_0))^2 \geq \kappa_n^2 M_1^{-1}/2\Big) - \eta$$

$$\geq 1 - 2\eta$$

which implies (49).

Finally for the error estimate $\hat{\sigma}_n^2$, note that by Assumption 2.2 and strong law of large number, $Q(\theta_0) = n^{-1}\sum_{t=1}^{n} u_t^2 \to_P \sigma^2$. Then it immediately follows from the consistency of $\hat{\theta}_n$ and (48) that

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

$\square$

Proof of Theorem 2.2. It follows from Lemma 7.3 that

$$\frac{d_n}{n}\sum_{t=1}^{n}(f(x_t, \theta) - f(x_t, \theta_0))^2 \to_D \Big(\int_{-\infty}^{\infty}(f(s, \theta) - f(s, \theta_0))^2 ds\Big) L_G(1, 0)$$

where

$$G(t) = \begin{cases} W_{\mu - 3/2}(t), & \text{under C1,} \\ W(t), & \text{under C2.} \end{cases} \tag{51}$$

The result (5) follows from the well known fact that $P(L_G(1, 0) > 0) = 1$ and $\int_{-\infty}^{\infty}(f(s, \theta) - f(s, \theta_0))^2 ds > 0$, for any $\theta \neq \theta_0$.

By noting $E[T(x_k) + T^2(x_k)] \leq C d_k^{-1}$ due to Lemma 3.2 of Wang and Phillips (2009b), simple calculations show that

$$\frac{d_n}{n}\sum_{k=1}^{n} E[T(x_k) + T^2(x_k)] \leq C$$

which implies (4). $\square$

Proof of Theorem 2.3. The results (4) and (5) follow from Lemma 7.5 with $g(x) = T(x) + T^2(x)$ and $g(x) = (f(x, \theta) - f(x, \theta_0))^2$, respectively. $\square$

Proof of Theorem 2.4. The proof is essentially the same as that given in Theorems 4.2 and 4.3 of PP. We only provide a outline for $f(x, \theta)$ satisfying the **Case 2**, which is related to the verification of the conditions in Lemma 7.6.

Let $\Theta_0 = \{\|\theta - \theta_0\| \geq \delta\}$ where $\delta > 0$ is a constant. By virtue of Lemma 7.6, it suffices to prove that, for any $\eta, M_0 > 0$, there exist a $n_0 > 0$ such that, for all $n > n_0$,

$$P\left(n^{-1} \inf_{\theta \in \Theta_0} D_n(\theta, \theta_0) > M_0\right) > 1 - \eta. \tag{52}$$

To prove (52), first note that $\sum_{t=1}^{n} u_t^2/n \leq M_0$ in probability, for some $M_0 > 0$, due to the Assumption 2.2 (i). This, together with Cauchy-Schwarz Inequality, yields that

$$
\begin{aligned}
& n^{-1} D_n(\theta, \theta_0) \\
& = \frac{1}{n} \sum_{t=1}^{n} (f(x_t, \theta) - f(x_t, \theta_0))^2 - \frac{2}{n} \sum_{t=1}^{n} (f(x_t, \theta) - f(x_t, \theta_0)) u_t \\
& \geq \frac{1}{n} \sum_{t=1}^{n} (f(x_t, \theta) - f(x_t, \theta_0))^2 - \frac{2}{n} \left(\sum_{t=1}^{n} (f(x_t, \theta) - f(x_t, \theta_0))^2\right)^{1/2} \left(\sum_{t=1}^{n} u_t^2\right)^{1/2} \\
& \geq M_n(\theta, \theta_0) \left[1 - \frac{2\sqrt{M_0 + o_P(1)}}{M_n(\theta, \theta_0)^{1/2}}\right],
\end{aligned}
\tag{53}
$$

where $M_n(\theta, \theta_0) = \frac{1}{n} \sum_{t=1}^{n} (f(x_t, \theta) - f(x_t, \theta_0))^2$. Hence, for any equivalent process x_k^* of x_k (i.e., $x_k^* =_d x_k, 1 \leq k \leq n, n \geq 1$, where $=_d$ denotes equivalence in distribution), we have

$$P\left(n^{-1} \inf_{\theta \in \Theta_0} D_n(\theta, \theta_0) > M_0\right) \geq P\left(\inf_{\theta \in \Theta_0} M_n^*(\theta, \theta_0) \left[1 - \frac{2\sqrt{M_0 + o_P(1)}}{M_n^*(\theta, \theta_0)^{1/2}}\right] > M_0\right), \tag{54}$$

where $M_n^*(\theta, \theta_0) = \frac{1}{n} \sum_{t=1}^{n} (f(x_t^*, \theta) - f(x_t^*, \theta_0))^2$.

Recalling $x_{[nt]}/d_n \Rightarrow G(t)$ on $D[0,1]$ and $G(t)$ is a continuous Gaussian process, by the so-called Skorohod-Dudley-Wichura representation theorem (e.g., Shorack and Wellner, 1986, p. 49, Remark 2), we can chose an equivalent process x_k^* of x_k so that

$$\sup_{0 \leq t \leq 1} |x_{[nt]}^*/d_n - G(t)| = o_P(1). \tag{55}$$

For this equivalent process x_t^*, it follows from the structure of $f(x, \theta)$ that

$$
\begin{aligned}
m(d_n, \theta)^2 & := \frac{1}{nv(d_n, \theta)^2} \sum_{t=1}^{n} f(x_t^*, \theta)^2 \\
& = \frac{1}{n} \sum_{t=1}^{n} h(x_t^*/d_n, \theta)^2 + o_P(1) \\
& = \int_0^1 h(x_{[ns]}^*/d_n, \theta)^2 ds + o_P(1) \\
& \to_P \int_0^1 h(G(s), \theta)^2 ds =: m(\theta)^2,
\end{aligned}
\tag{56}
$$

uniformly in $\theta \in \Theta$. Due to (56), the same arguments as in the proof of Theorem 4.3 in PP yields that

$$\inf_{\theta \in \Theta_0} M_n^*(\theta, \theta_0) \to \infty, \quad \text{in probability,}$$

which, together with (54), implies (52). $\square$

Proof of Theorem 3.1. According to the introductory remark of Section 3, it suffices to verify the conditions (14)–(18) with

$$\kappa_n^2 = n/d_n, \quad c_{i,j} = \int_{-\infty}^{\infty} \dot{f}_i(s, \theta_0) \dot{f}_j(s, \theta_0) ds / \int_{-\infty}^{\infty} f(s, \theta_0) ds,$$

$$Y = \Sigma^{1/2} \mathbf{N} L_G^{1/2}(1, 0), \quad Z = \int_{-\infty}^{\infty} f(s, \theta_0) ds\, L_G(1, 0),$$

under Assumptions 3.1 and 3.2.

Recalling $\hat{\theta}_n \to_P \theta_0$ due to Theorems 2.1 and 2.2, $\hat{\theta}_n$ falls in $\mathcal{N}_\delta(\theta_0) = \{\theta : \|\theta - \theta_0\| < \delta\}$ in probability. This, together with Lemma 7.2, yields (14)–(16).

It follows from Lemma 7.3 with $g(x) = \alpha' \dot{f}(x, \theta_0)$ and $g_1(x) = f(x, \theta_0)$ that

$$\left\{ \left(\frac{d_n}{n}\right)^{1/2} \sum_{t=1}^{n} \alpha' \dot{f}(x_t, \theta_0) u_t, \frac{d_n}{n} \sum_{t=1}^{n} f(x_t, \theta_0) \right\}$$

$$\to_D \left\{ \tau N L_G^{1/2}(1, 0), Z \right\} =_D \left\{ \alpha' Y, Z \right\} \tag{57}$$

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

$$\tau N = \left(\int_{-\infty}^{\infty} [\alpha' \dot{f}(s, \theta_0)]^2 ds \right)^{1/2} N$$

$$=_D \alpha' \left(\int_{-\infty}^{\infty} \dot{f}(s, \theta_0) \dot{f}(s, \theta_0)' ds \right)^{1/2} \mathbf{N} = \alpha' \Sigma^{1/2} \mathbf{N}.$$

This proves (17).

Finally, it follows from Lemma 7.3 with $g_1(x) = \beta_1 \dot{f}_i(x, \theta_0) \dot{f}_j(x, \theta_0) + \beta_2 f(x, \theta_0)$ that

$$\frac{d_n}{n} \sum_{t=1}^{n} \left[\beta_1 \dot{f}_i(x_t, \theta_0) \dot{f}_j(x_t, \theta_0) + \beta_2 f(x_t, \theta_0) \right] \to_D \tau_0 L_G(1, 0),$$

where $\tau_0 = \int_{-\infty}^{\infty} \left[\beta_1 \dot{f}_i(x, \theta_0) \dot{f}_j(x, \theta_0) + \beta_2 f(x, \theta_0) \right] dx$. This yields (18), that is,

$$\left\{ \frac{d_n}{n} \sum_{t=1}^{n} \dot{f}_i(x_t, \theta_0) \dot{f}_j(x_t, \theta_0), \frac{d_n}{n} \sum_{t=1}^{n} f(x_t, \theta_0) \right\} \to_D \left\{ c_{i,j} Z, Z \right\}.$$

$\square$

Proof of Theorem 3.2. As in Theorem 3.1, we need to verify the conditions (14)–(18), but with

$$\kappa_n^2 = a(n), \quad c_{i,j} = \int_{-\infty}^{\infty} \dot{f}_i(s,\theta_0)\dot{f}_j(s,\theta_0)\pi(ds) \Big/ \int_{-\infty}^{\infty} f(s,\theta_0)\pi(ds),$$

$$Y = \Sigma_\pi^{1/2}\mathbf{N}\Pi_\beta, \quad Z = \int_{-\infty}^{\infty} f(s,\theta_0)\pi(ds)\,\Pi_\beta,$$

under Assumptions 3.1* and 3.2*. The details are similar to that of Theorem 3.1, with Lemma 7.3 replaced by Lemma 7.5, and hence are omitted. □

Proof of Theorem 4.1. We first establish the consistency result. The proof goes along the same line as in Theorem 2.1. It suffices to show that for any fixed $\pi_0 \in \Theta \cap \mathcal{N}^c$,

$$\sup_{\theta \in \mathcal{N}_\delta(\pi_0)} \frac{d_n}{n} \sum_{t=1}^{n} |[f(x_t,\theta) - f(x_t,\pi_0)]u_t| \to_P 0 \tag{58}$$

as $\delta \to 0$ uniformly for all large n, and

$$\sum_{t=1}^{n} (f(x_t,\pi_0) - f(x_t,\theta_0))u_t = o_P(n/d_n). \tag{59}$$

For (58). By (40) in Lemma 7.4 with $g_1(x) = |T(x)|$, we have,

$$\sup_{\theta \in \mathcal{N}_\delta(\pi_0)} \frac{d_n}{n} \sum_{t=1}^{n} |[f(x_t,\theta) - f(x_t,\pi_0)]u_t| \leq \sup_{\theta \in \mathcal{N}_\delta(\pi_0)} h(\|\theta - \pi_0\|) \frac{d_n}{n} \sum_{t=1}^{n} |T(x_t)||u_t|$$

$$\leq C \sup_{\theta \in \mathcal{N}_\delta(\pi_0)} h(\|\theta - \pi_0\|)\, O_P(1) \to_P 0$$

as $\delta \to 0$.

Also, by (41) in Lemma 7.4 with $g_1(x) = f(x,\pi_0) - f(x,\theta_0)$, we have that

$$\sum_{t=1}^{n} (f(x_t,\pi_0) - f(x_t,\theta_0))u_t = O_P[(n/d_n)^{1/2}]$$

which implies the required (59) immediately.

We next give the limit distribution. As in Theorem 3.1, we need to verify the conditions (14)–(18), but with

$$\kappa_n^2 = n/d_n, \quad c_{i,j} = \int_{-\infty}^{\infty} \dot{f}_i(s,\theta_0)\dot{f}_j(s,\theta_0)ds \Big/ \int_{-\infty}^{\infty} f(s,\theta_0)ds,$$

$$Y = \Lambda^{1/2}\mathbf{N}L_G^{1/2}(1,0), \quad Z = \int_{-\infty}^{\infty} f(s,\theta_0)ds\, L_G(1,0),$$

under Assumptions 4.1.

The proof for (14), (15) and (18) are exactly same as that of Theorem 3.1, and using similar arguments, (17) will follow from Lemma 7.3.

For (16). Because $\hat\theta_n \to_P \theta_0$, by (58) with $\pi_0 = \theta_0$, and (59), we have, for any $1 \le i, j \le m$,

$$
\left| \frac{d_n}{n} \sum_{t=1}^{n} \ddot{f}_{ij}(x_t, \theta_n) u_t \right| \le \left| \frac{d_n}{n} \sum_{t=1}^{n} [\ddot{f}_{ij}(x_t, \theta_n) - \ddot{f}_{ij}(x_t, \theta_0)] u_t \right| + \left| \frac{d_n}{n} \sum_{t=1}^{n} \ddot{f}_{ij}(x_t, \theta_0) u_t \right|
$$
$$
= o_P(1)
$$

which yields the required (16). □

8   Appendix B: Definition of regular function

H is called a regular function if there exists for each $\epsilon > 0$ continuous functions $\underline{H}_\epsilon$, $\overline{H}_\epsilon$, and a constant $\delta_\epsilon > 0$ such that $\underline{H}_\epsilon(x) \le H(y) \le \overline{H}_\epsilon(x)$ for all $|x - y| < \delta_\epsilon$ on K, a compact set of R, and such that $\int_K (\overline{H}_\epsilon - \underline{H}_\epsilon)(x) dx \to 0$ as $\epsilon \to 0$.

REFERENCES

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

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

Chang, Y., Park, J. Y., 2001. Index models with integrated time series. Journal of Econometrics, 114, 73–106.

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.

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., Tjøstheim, D., 2009. Specification testing in nonlinear and nonstationary time series autoregression. The Annals of Statistics, 37, 3893–3928.

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

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

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

Jaganathan, 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 Protter, 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.

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

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

Teräsvirta, T., Tjøstheim, D., 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.

Wang, Q., 2011. Martingale limit theorems revisited and non-linear cointegrating regression. Working Paper.

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, forthcoming.

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

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