Statistica Sinica Preprint No: SS-2021-0426	
Title	Weighted Nonlinear Regression With Nonstationary Time Series
Manuscript ID	SS-2021-0426
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202021.0426
Complete List of Authors	Chunlei Jin and Qiying Wang
Corresponding Authors	Qiying Wang
E-mails	qiying.wang@sydney.edu.au

Statistica Sinica

Weighted nonlinear regression with nonstationary time series

Chunlei Jin and Qiying Wang

The University of Sydney

Abstract: This study investigates a weighted least squares (WLS) estimation in a nonlinear cointegrating regression. In a nonlinear regression model, where the regressors include nearly integrated arrays and stationary processes, we show that the WLS estimator has a mixed Gaussian limit, and the corresponding Studentized statistic converges to a standard normal distribution. The WLS estimator is free of the memory parameter, even when a fractional process is included in the regressors. We also consider an ordinary least squares estimation in a nonlinear cointegrating regression. Compared with the WLS estimator, the limit distribution of the ordinary least squares estimator is non-Gaussian, and depends on the nuisance parameters from the regressors when the regression function is non-integrable.

Key words and phrases: Cointegration, Nonlinear cointegrating regression, Weighted least squares estimation, A mixture of normal distributions, Nonstationarity.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

1. Introduction

It is well known that nonstandard asymptotic behavior appears in nonlinear (linear) cointegrating regressions. A fundamental issue in such a regression model with nonstationary time series is that the limiting distribution of the least squares (LS) often depends on various nuisance parameters, and/or such a limit result is cumbersome in the relevant asymptotic inferences. To illustrate, we consider the following cointegrating regression model:

$$y_k = \alpha\, x_k + \eta_k, \quad x_k = x_{k-1} + \epsilon_k, \qquad k = 1, 2, ..., \tag{1.1}$$

where $v_k := (\epsilon_k, \eta_{k-1})$ is assumed to be a sequence of independent and identically distributed (i.i.d.) random vectors with $Ev_1 = 0$ and $\Omega := cov(v_1, v_1') = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. The standard LS estimator $\hat{\alpha}_n$ of the unknown parameter α in model (1.1) has a nonstandard limit distribution, namely,

$$n\,(\hat{\alpha} - \alpha) = \frac{\sum_{k=1}^n \eta_k x_k}{\sum_{k=1}^n x_k^2} \to_D \frac{\int_0^1 B_{1t} dB_{2t}}{\int_0^1 B_{1t}^2 dt}, \tag{1.2}$$

where $\big(B_{1t}, B_{2t}\big)$ is a two-dimensional Brownian motion with covariance matrix Ω. In practice, ρ (or Ω) is usually an unknown parameter. Noting that ρ is hidden in the functional $\int_0^1 B_{1t} dB_{2t}$, result (1.2) cannot be used directly in inference theory, in which the relevant asymptotic critical value usually depends on the standard normal distribution.

To solve this problem, several instrumental estimators have been proposed in literature as alternatives to the standard LS. Earlier contributions include Phillips and Hansen (1990) and Phillips (1995), who consider the fully modified LS. In a different direction, Magdalinos and Phillips (2009) (see also Kostakis et al. (2015) for refined statements) introduced an IVX estimator, using linear filtering to transform the regressor x_k into a mildly integrated process. In comparison with the standard LS, this IVX estimator has a mixed Gaussian limitation,

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

such that the corresponding Student t statistic converges to a standard normal distribution, enabling many works on conventional inference theory. For other contributions in this area, see Jansson and Moreira (2006), Phillips and Lee (2013), Ellott et al. (2015), Bae and de Jong (2007), Yang et al. (2020), and Demetrescu et al. (2022). The latter works generalize the IVX method to multi-regression (linear) models with nonstationary time series, and the method has been used to test the episodic predictability in stock returns. More recently, for a simple nonlinear in-variables cointegrating regression model, the locally trimmed LS was introduced in Hu et al. (2021) and Kasparis and Phillips (2020) investigated the model with a single covariate heavy-tailed regressor.

Our study has a similar goal to the aforementioned works, but focus on nonlinear parametric cointegrating regression. Nonlinear cointegrating regression was initially introduced in Park and Phillips (2001). Since then, significant developments have occurred in parametric, nonparametric, and semiparametric specifications of such models. These developments have provided a framework for econometric estimation and inference for a wide class of nonlinear, nonstationary relationships: see, for instance, Wang and Phillips (2009a), Wang and Phillips (2009b), Wang and Phillips (2016), Duffy (2016), Duffy (2020), Wang et al. (2021), Chang et al. (2001), Bae and de Jong (2007), Kim and Kim (2012), Dong et al. (2016), Dong and Linton (2018), Lin et al. (2020), and Wang (2021), together with the references therein. It is now well known that the conventional kernel estimator in a nonparametric cointegrating regression has a mixed Gaussian limitation, even when the regressors are nearly integrated, but that the behaviors of parametric regression estimators are asymptotically nonpivotal. The limit distribution of the standard LS estimator in nonlinear (parametric) cointegrating regression is not only non-Gaussian, but also depends on the unknown degree of persistence of the regressor, posing difficulties in inference theory. As a result, it is desirable to develop an alternative estimation

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

theory to the standard LS, so that the limit distribution of the suggested estimator is pivotal in regression models with nonstationary time series.

The aim of this study is to investigate the weighted least squares (WLS) estimation in a nonlinear cointegrating regression. For some selected weight functions, our results show that the WLS estimator has a mixed Gaussian limit, and that the corresponding Studentized statistic converges to a standard normal distribution. Such a WLS estimator is free of the memory parameter, even when a fractional process is included in the regressors, enabling us to apply much classical inference theory directly. In comparison with the ordinary LS estimator, there is a slightly low convergence rate for the WLS estimator (less than $(\log n)^{-1}$, say). This deduction in the convergence rate is necessary for a standard normal limitation. For further explanation, refer to Remark 2.

The remainder of this paper is organized as follows. We present the main results in Section 2. Section 2.1 introduces a nonlinear cointegrating regression model and the corresponding WLS estimator. Our model is more general than those of previous works by allowing for both nonstationary and stationary regressors. Our assumptions and some preliminaries are given in Section 2.2, and the asymptotic theory of the WLE estimators is developed in Section 2.3. In Section 3, we investigate the asymptotics of the ordinary LS estimators for a comparison. A numerical example is given in Section 4 to illustrate our asymptotics. This simple example is designed to illustrate the effect of different weighted functions and the performance of the Studentized statistic given in (2.7). Section 5 concludes the paper. All technical proofs are given in the Appendix, where we also provide a framework for the WLS estimation in a general nonlinear regression model, and collect some general results on convergence to local time and a mixture of normal distributions.

Throughout this paper, we denote constants as $C, C_1, C_2, ...$, although their values may

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

vary between instances.

2. Main results

2.1 Model and estimation

We consider a nonlinear cointegrating regression model:

$$y_k = f(x_k, w_k, \theta) + u_k, \quad k = 1, \ldots, n \tag{2.1}$$

where y_k is the dependent variable, w_k is an $m \times 1$ stationary random vector, x_k is a unit-root nonstationary regressor, u_k is the residual, $f(\cdots)$ is a given smooth regression function, and $\theta = (\theta_1, ..., \theta_q)$ is a vector of unknown parameters such that $\theta \in \Theta$, where Θ is a compact set of R^q.

When the regression function $f(\cdots)$ is linear, model (2.1) is well studied in the literature. As a result, linear cointegrating regression has become an important framework in which to capture long-term relationships among many macroeconomic time series. However, in spite of their importance and convenience in implementation, the linear structure in the related theory is often too restrictive in practice. For empirical examples, see Granger and Teräsvirta (1993) and Teräsvirta et al. (2011). To overcome this problem, various nonlinear parametric cointegrating regression models have been introduced. For instance, Park and Phillips (2001), Chang et al. (2001), and Chan and Wang (2015) allow for the regression function $f(\cdots)$ to be integrable or a class of homogeneous functions. Recently, Hu et al. (2021) considered the power regression function; see also Li et al. (2016) and Wang (2021).

The regression given in (2.1) is similar to the models studied previously, but allows for the presence of stationary regressors. This generalization provides cointegrating relationships that vary or evolve smoothly over time. In particular, our model allows for the additive regression, functional-coefficients nonlinear cointegrating regression, and nonlinear cointegrating regression with finite lags of the $I(1)$ variables x_k. Thus the regression in (2.1) is particularly useful in empirical applications in which there may be a structural evolution in a relationship over time.

We estimate the unknown parameter θ in model (2.1) using the WLS method. Specifically, we define the WLS estimator $\widehat{\theta}_n$ of an interior θ_0 (real value) of Θ to minimize the sum of the weighted squared errors:

$$\widehat{\theta}_n = \arg\min_{\theta\in\Theta} \sum_{k=1}^{n} \big[y_k - f(x_k, w_k, \theta)\big]^2 \lambda(x_k/b_n), \tag{2.2}$$

where $0 < b_n \to \infty$ is a sequence of constants, and $\lambda(.)$ is a bounded weight function. Here, we show that the asymptotics of $\widehat{\theta}_n$ under selected weight functions differ significantly from the usual LSE (i.e., $\lambda(.) = 1$) considered by Park and Phillips (2001), Chan and Wang (2015), and Wang (2021). In particular, when self-normalization is used, this WLS estimator has a standard normal limitation, and is free of the memory parameter even when a fractional process is included in the regressors. Note that the WLE method is widely used in stationary regression. In related work, Li et al. (2016) consider a nonlinear regression with a Harris recurrent Markov chain. However, no stationary regressors are involved.

2.2 Assumptions and preliminaries

Let $v_k := (\epsilon_k, \eta_{k-1})$, for $k \in \mathbb{Z}$, be a sequence of i.i.d. random vectors, with $Ev_0 = 0$, $E\epsilon_0^2 = E\eta_0^2 = 1$, and $\rho = E\epsilon_1\eta_0$. Let $\xi_j = \sum_{k=0}^{\infty} \phi_k\, \epsilon_{j-k}$, for $j \geq 1$, be a linear process, where the coefficients ϕ_k, for $k \geq 0$, satisfy one of the following conditions:

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

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

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

Suppose that $\limsup_{s\to\infty} s^{\delta}|Ee^{is\epsilon_0}| < \infty$ for some $\delta > 0$ throughout this paper. This distributional smooth condition on ϵ_0 is required to establish the convergence to local time for a partial sum process of ξ_j, as shown in Appendix B. To establish the asymptotics of the WLS estimator $\hat{\theta}_n$, we use the following assumptions on the regressors x_k and w_k and the error process u_k.

A1 $x_k = \rho_n x_{k-1} + \xi_k$, where $\rho_n = 1 - \tau n^{-1}$, for some $\tau \geq 0$, and $x_0 = o_P(\sqrt{n})$;

A2 $w_k = (w_{1k}, ..., w_{dk})$, where, for $i = 1, ..., d$, $w_{ik} = \Gamma_i(v_k, ..., v_{k-k_0})$, for some $k_0 \geq 0$, and $\Gamma_i(.)$ are real measurable functions of their components;

A3 $\{u_k, \mathcal{F}_k\}_{k\geq 1}$, where $\mathcal{F}_k$ is an σ-field generated by $v_{k+1}, v_k, ...$, forms a stationary martingale difference, with $E\big(u_k^2 \mid \mathcal{F}_{k-1}\big) = \sigma^2 > 0$ and $\sup_{k\geq 1} E\big(|u_k|^{2+\delta} \mid \mathcal{F}_{k-1}\big) \leq C < \infty$, for some $\delta > 0$

A1 allows for the nearly integrated regressor x_k derived from short memory (under **SM**) and long memory (under **LM**) innovations, which is quite general, in practice. Define

$$d_n^2 \;=\; \mathbb{E}|\sum_{k=1}^{n}\xi_k|^2 \sim \begin{cases} c_\mu\, n^{3-2\mu}a^2(n), & \text{under } \mathbf{LM}, \\ \phi^2\, n, & \text{under } \mathbf{SM}, \end{cases} \tag{2.3}$$

where c_μ is a constant. Standard functional limit theory (see Buchmann and Chan (2007) or Theorem 2.21 of Wang (2015), with a minor modification) shows that

$$\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\epsilon_i,\ \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\epsilon_{-i},\ \frac{1}{d_n}x_{[nt]}\Big) \;\Rightarrow\; \big(B_t,\ B_{-t},\ X_t\big) \tag{2.4}$$

on $D_{\mathbb{R}^3}[0,\infty)$, where $\{B_t\}_{t\geq 0}$ is a standard Brownian motion, $\{B_{-t}\}_{t\geq 0}$ is an independent copy of $\{B_t\}_{t\geq 0}$, and X_t is defined by

$$X_t \;=\; W(t) + \tau\int_0^t e^{-\tau(t-s)}W(s)ds,$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

with $W_k = \begin{cases} G_{3/2-\mu}(t), & \text{under } \mathbf{LM}, \\ G_{1/2}(t), & \text{under } \mathbf{SM}, \end{cases}$ where G_H is a fractional Brownian motion having the following presentation: with $a_+ = \max\{a, 0\}$,

$$G_H(t) \quad = \quad \kappa_H \int_{-\infty}^{t} (t-u)_+^{H-1/2} - (-u)_+^{H-1/2} dB_u,$$

where κ_H is a constant such that $EG_H^2(1) = 1$. Note that X_t is an Ornstein−Uhlenbeck process with a continuous local time $L_X(t,s)$ defined by

$$L_X(t,s) = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_0^t I(|X_r - s| \le \epsilon) dr.$$

A2 is related to stationary regressors, where k_0 (fixed) can be taken as large as necessary, and quite general settings on $\Gamma_i(.)$ are allowed. An extension to general linear processes is possible when certain smoothing conditions on the regression function $f(.)$ are imposed. Details can be found in Corollary 1. **A3** enables model (2.1) to have a martingale structure, which is widely used in the literature; see, for instance, Park and Phillips (2001) and Chan and Wang (2015). An extension that includes varying volatility is possible, but involves quite complicated calculations, and is therefore left to future work.

We next introduce the assumptions on the regression function $f(...)$ and the weight function $\lambda(.)$. Write $\dot{f}(x,y,\theta) = (\dot{f}_1, ..., \dot{f}_q)'$, where $\dot{f}_i = \frac{\partial f(x,y,\theta)}{\partial \theta_i}$, for $i = 1, ..., q$, and let $p(x,y,\theta)$ be one of f and $\dot{f}_i$, for $i = 1, ..., q$.

A4 A measurable function $T_p(x,y)$ and a continuous function $T(x)$ exist such that

(i) for each $\theta, \theta_0 \in \Theta$, $(x,y) \in R^{1+q}$ and for some $\alpha > 0$,

$$|p(x,y,\theta) - p(x,y,\theta_0)| \le ||\theta - \theta_0||^{\alpha}\, T_p(x,y);$$

(ii) for each $\theta \in \Theta$ and $(x, y) \in R^{1+q}$,

$$p(lx, y, \theta) \quad = \quad v_p(l)\, h_p(x, y, \theta) \, + R(lx, y, \theta),$$

where $v_p(l)$ is a positive real function bounded away from zero as l becomes large, $h_p(x, y, \theta)$ is a locally Riemann-integrable function (i.e., Riemann-integrable on any compact set), and as $l \to \infty$,

$$\sup_{(x,y)\in R^{1+q}} \sup_{\theta\in\Theta} \mid R(lx, y, \theta) \mid / T_p(lx, y) \to 0;$$

(iii) for any $l > 0$ and $(x, y) \in R^{1+d}$, and for some $\beta > 0$,

$$\sup_{\theta\in\Theta} |h_p(x, y, \theta)| \le T(x)(1 + ||y||^\beta), \quad T_p(lx, y) \le v_p(l)\, T(x)(1 + ||y||^\beta);$$

(iv) $\lambda(x)\big[1 + T^2(x)\big]$ is a bounded and integrable function;

(v) $\Sigma = \int_{-\infty}^{\infty} \lambda(x)\, E\big[\dot h(x, w_1, \theta_0)\dot h'(x, w_1, \theta_0)\big] dx$ is a positive-definite matrix, where

$$\dot h(x, y, \theta) = \big(h_{\dot f_1}(x, y, \theta), ..., h_{\dot f_q}(x, y, \theta)\big),$$

and, for any $\delta > 0$ such that $\{\theta : ||\theta - \theta_0|| \ge \delta\} \subset \Theta$,

$$\min_{||\theta-\theta_0||\ge\delta} \int_{-\infty}^{\infty} \lambda(x)\, E\big[h_f(x, w_1, \theta) - h_f(x, w_1, \theta_0)\big]^2 dx > 0. \tag{2.5}$$

Assumption **A4**(i)$-$(iii) is a weak condition that is required for the regression function $f(x, y, \theta)$. It includes many common nonlinear (linear) regression functions, and is easy to verify in practice. To give an illustration, let $f(x, y, \theta) = m(x, \theta)\, K(y)$. If $|K(y)| \le 1 + ||y||^\beta$, for some $\beta > 0$, and $m(x, \theta)$ is one of $\theta e^x/(1 + e^x)$, $\theta \log |x|$, $\theta|x|^\alpha$(α is fixed), or $\theta_0 + \theta_1|x| + ... + \theta_k|x|^k$, then $f(x, y, \theta)$ satisfies **A4**(i)$-$(iii). The weight function $\lambda(x)$ is required to satisfy **A4**(iv), which is key to the development of our asymptotics. Such a weight function, together with an additional condition on b_n, reduces the signal of the regression function $f(x, y, \theta)$, enabling us to use the extended martingale limit theorem given in Wang (2014) can be used; see Remark 2 for more details. Because $T(x)$ is continuous, $\lambda(x) = I(|x| \le K)$, for any $K > 0$, satisfies **A4**(iv). Assumption **A4**(v) ensures the consistency of $\hat\theta_n$, where the condition (2.5) is close to necessary, and can therefore be understood as an identification condition in model (2.1).

2.3 Asymptotic theory

This section discusses the limit behavior of the WLS estimator $\hat\theta_n$ defined by (2.2). We provide a pivotal asymptotic distribution. The limit results without an effect from a weight function are considered in Section 3 for a comparison.

Theorem 1. *Suppose* ***A1–A4*** *hold and* $E||w_1||^{4\beta+2} < \infty$*, where* β *is given as in* ***A4*** *(iii). Then, for any* $b_n > 0$ *such that* $c_n := d_n/b_n \to \infty$*, we have*

$$D_n\,(\hat\theta_n - \theta_0) \to_D \sigma\,\Sigma^{-1}\,\Sigma_1^{1/2}\,L_X(1,0)^{-1/2}\,\mathbb{N}, \tag{2.6}$$

where $D_n = \big(n/c_n\big)^{1/2}\, diag\big(v_{\dot f_1}(b_n), ..., v_{\dot f_q}(b_n)\big)$,

$$\Sigma_1 = \int_{-\infty}^{\infty} \lambda^2(x)\, E\big[\dot h(x, w_1, \theta_0)\dot h'(x, w_1, \theta_0)\big]\, dx,$$

and $\mathbb{N}$ *is a standard q-dimensional normal random vector independent of* X_t*. We further have*

$$T_n \quad := \quad \Omega_n \Omega_{1n}^{-1/2}\,(\hat\theta_n - \theta_0) \to_D \sigma\,\mathbb{N}, \tag{2.7}$$

where $\Omega_n = \sum_{k=1}^n \lambda(x_k/b_n)\dot f(x_k, w_k, \theta_0)\dot f(x_k, w_k, \theta_0)'$ *and*

$$\Omega_{1n} = \sum_{k=1}^n \lambda^2(x_k/b_n)\dot f(x_k, w_k, \theta_0)\dot f(x_k, w_k, \theta_0)'.$$

Remark 1. Because $\sigma^2 = Eu_1^2$, under given conditions, a natural consistency estimator $\hat{\sigma}_n^2$ of σ^2 is $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{k=1}^n \big[y_k - f(x_k, w_k, \hat{\theta}_n)\big]^2$. As a result, (2.7) can be rewritten as

$$\hat{\sigma}_n^{-1}\,\Omega_n\Omega_{1n}^{-1/2}\,(\hat{\theta}_n - \theta_0) \to_D \mathbb{N}, \tag{2.8}$$

indicating a pivotal and standard normal limitation. Recall **A1**. The regressor x_k in model (2.1) allows for a nearly integrated process derived from short memory (under **SM**) and long memory (under **LM**) innovations. The result (2.8) is free of all parameters, such as τ, μ, and ψ_k raised in the nearly integrated regressor x_k, and is therefore extremely convenient in inference theory.

Remark 2. In the usual LS estimation theory for nonlinear cointegrating regression (i.e., the estimator $\hat{\theta}_n$ is given by (2.2) with $\lambda(x) \equiv 1$), the standard convergence rate for the asymptotics is $\widetilde{D}_n = n^{1/2}\,\mathrm{diag}\big(v_{\dot{f}_1}(d_n), ..., v_{\dot{f}_q}(d_n)\big)$. See Theorem 3. In comparison, result (2.6) has a low convergence rate, because $c_n = d_n/b_n \to \infty$. This reduction in the convergence rate, together with condition **A4**(iv), is essentially necessary for the standard normal limitation in (2.7). A simple example helps to explain this argument. Consider a nonlinear-in-variables cointegrating regression:

$$y_k = \theta\, g(x_k) + u_k, \tag{2.9}$$

where continuous functions v and H exist such that $g(l\,x) = v(l)\,H(x)$, for any $l \geq 0$ and $x \in R$. Write $x_{nk} = x_k/d_n$ and recall that $c_n = d_n/b_n$. For this simple model, we have

$$\begin{aligned} T_n &= \frac{\sum_{k=1}^n g^2(x_k)\lambda(x_k/b_n)}{\sqrt{\sum_{k=1}^n g^2(x_k)\lambda^2(x_k/b_n)}}\,(\hat{\theta}_n - \theta_0) \\ &= \frac{\sum_{k=1}^n u_k g(x_k)\lambda(x_k/b_n)}{\sqrt{\sum_{k=1}^n g^2(x_k)\lambda^2(x_k/b_n)}} = \frac{\sum_{k=1}^n u_k H(c_n\,x_{nk})\lambda(c_n x_{nk})}{\sqrt{\sum_{k=1}^n H^2(c_n\,x_{nk})\lambda^2(c_n x_{nk})}}. \end{aligned} \tag{2.10}$$

When $\lambda(x)(1 + H^2(x))$ is assumed to be bounded and integrable (an equivalent condition to that of **A4**(iv)), both $\lambda(x)H(x)$ and $\lambda^2(x)H^2(x)$ are bounded and integrable. In this case, to ensure $T_n \to_D N(0,\sigma^2)$ (i.e, (2.7) holds), $c_n \to \infty$ and $c_n/n \to 0$ are necessary. See Wang (2014) or Theorem 7 in Appendix B. Indeed, if $c_n = 1$, we have

$$T_n \quad \to_D \quad \frac{\sigma \int_0^1 H(X_t)\lambda(X_t)dB_t}{\sqrt{\int_0^1 H^2(X_t)\lambda^2(X_t)dt}} \;\neq_D\; N(0,\sigma^2),$$

provided that $(x_{n,[nt]}, \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]} u_j) \Rightarrow (X_t, \sigma B_t)$ on $D_{R^2}[0,1]$.

Remark 3. The simple model (2.9) also helps to reveal the mechanism of our weighting scheme. In fact, if $g(x)$ is bounded and integrable, we have

$$\frac{\sum_{k=1}^n u_k g(x_k)}{\sqrt{\sum_{k=1}^n g^2(x_k)}} \to_D N(0,\sigma^2), \tag{2.11}$$

as shown in Theorem 2. As a result, the usual Studentized statistic T_n (i.e., $\lambda(x) = 1$ applies) has a standard normal limitation. However, result (2.11) is incorrect when $g(x)$ is a non-integrable function. By employing a weight function $\lambda(x)$, we can enable both $g(x)\lambda(x)$ and $g^2(x)\lambda^2(x)$ to be integrable, and thus use the "integrable function" asymptotics, yielding a standard normal limitation for T_n. In our weighting scheme, using $\lambda(x/b_n)$ instead of $\lambda(x)$ improves the convergence rate for $\hat{\theta}_n$, as seen in (2.6).

Remark 4. Although the results in (2.7) and (2.8) have theoretical and practice advantages, it seems to be difficult to determine optimal choices of $\lambda(x)$ and b_n in finite-sample simulations. To provide an illustration, we assume in (2.10) that $H(x) = |x|^{1/2}$ and $\lambda(x) = I(|x| \le K)$, where $K > 0$ is a constant. In this case, we have

$$T_n \quad = \quad \frac{\sum_{k=1}^n u_k |x_{nk}|^{1/2} I(|x_{nk}| \le K/c_n)}{\sqrt{\sum_{k=1}^n |x_{nk}| I(|x_{nk}| \le K/c_n)}} \to_D N(0,\sigma^2),$$

for any fixed $K > 0$ and $b_n > 0$ satisfying $c_n = \sqrt{n}/b_n \to \infty$ or $Kc_n^{-1} \to 0$. Although the limit distribution of T_n is free of the values of K when b_n is given such that $c_n \to \infty$, the performance of T_n in finite-sample simulations (with fixed n) depends on K/c_n (being close to zero or not) rather than on K or c_n individually; see Section 4 for detailed numerical examples. This simple example indicates the complexity of choosing optimal $\lambda(x)$ and b_n in a finite sample size. This topic deserves further consideration, but the solution might require an accurate estimate for the distribution function of the Studentized statistic T_n, such as an ideal Berry−Esseen bound or an Edgeworth expansion can be established. This left for future work.

When it is difficult to choose the optimal $\lambda(x)$ and b_n, we suggest using a truncated weight function $\lambda(x) = I(|x| \le K)$ in practice (b_n can be taken close to d_n, such that $d_n/b_n \to \infty$). This kind of weight function reduces the signal by removing partial data. Thus, we can apply Wang's extended martingale limit theorem for a standard normal limitation (e.g., Wang (2014)), which closely matches the mechanism of our weighting scheme. Furthermore, as shown in Section 4, a reasonable normal density approximation can be achieved with $K = 1/10$ for a wide class of regressors, even for a sample size as small as $n = 100$.

Remark 5. If the regression function $f(x, y, \theta)$ satisfies further smoothness conditions, the stationary regressor w_k given in **A2** can be extended to include general linear processes. The following is a corollary for such an extension.

Corollary 1. *If in addition to **A1**, **A3**, and **A4**, for all $x \in R$ and $y, t \in R^d$,*

$$||\dot{h}(x, y, \theta_0) - \dot{h}(x, t, \theta_0)|| \le T(x)||y - t||^{\delta}(1 + ||y||^{\beta} + ||t||^{\beta}), \quad \textit{for some } \delta > 0, \tag{2.12}$$

*the results (2.6) and (2.7) still hold when we replace **A2** with*

***A2**** $w_k = (w_{1k}, ..., w_{dk})$*, where, for* $i = 1, ..., d$, $w_{ik} = \sum_{j=0}^{\infty} \psi_{i,j}\, \lambda'_{k-j}$*, with* $\psi_{i,j} = (\psi_{i,1j}, \psi_{i,2j})$ *satisfying* $\sum_{j=0}^{\infty}(|\psi_{i,1j}| + |\psi_{i,2j}|) < \infty$.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

3. Asymptotics for the standard LS estimator (without a weight effect)

Theorem 1 provides a useful pivotal limit result for the WLS estimator $\hat{\theta}_n$ defined by (2.2). For a comparison, this section considers the asymptotics of the standard LS estimator in model (2.1), denoted by $\hat{\theta}_n$: that is, $\hat{\theta}_n$ is defined by (2.2) with $\lambda(.) \equiv 1$. Note that the standard LS in nonlinear cointegrating regression has been investigated by Park and Phillips (2001), Chang et al. (2001), and Chan and Wang (2015). See also Wang (2021) for a nonlinear regression with nonstationarity and heteroscedasticity. The results in this section generalize these previous works by allowing for more general settings. In particular, we allow for a stationary regressor w_k in the model (2.1). Because there are essential differences between integrable and nonintegrable functions, we present the asymptotics in two separate subsections.

3.1 Integrable function

This section considers the limit distribution of $\hat{\theta}_n$ defined by (2.2) with $\lambda(.) \equiv 1$ when $f(x, y, \theta)$ is an integrable function, for each fixed y and $\theta \in \Theta$. As in Section 2.3, we define $\dot{f}(x, y, \theta) = (\dot{f}_1, ..., \dot{f}_q)'$, where $\dot{f}_i = \frac{\partial f(x,y,\theta)}{\partial \theta_i}$, for $i = 1, ..., q$, and $p(x, y, \theta)$ is one of f and $\dot{f}_i$, for $i = 1, ..., q$.

Theorem 2. *In addition to* ***A1***$-$***A3****, suppose that*

(i) a bounded and integrable function $T(x)$ exists such that for each $\theta, \theta_0 \in \Theta$, $(x, y) \in R^{1+d}$ and for some $\alpha > 0$,

$$\begin{aligned} |p(x, y, \theta) - p(x, y, \theta_0)| &\leq ||\theta - \theta_0||^{\alpha}\, T(x)(1 + ||y||^{\beta}) \quad \textit{and} \\ \sup_{\theta \in \Theta} |p(x, y, \theta)| &\leq T(x)(1 + ||y||^{\beta}); \end{aligned} \tag{3.1}$$

(ii) $\Sigma_2 = \int_{-\infty}^{\infty} E\big[\dot f(x, w_1, \theta_0)\dot f'(x, w_1, \theta_0)\big]dx$ *is a positive-definite matrix and, for any* $\delta > 0$ *such that* $\{\theta : ||\theta - \theta_0|| \geq \delta\} \subset \Theta$,

$$\min_{||\theta-\theta_0||\geq\delta} \int_{-\infty}^{\infty} E\big[f(x, w_1, \theta) - f(x, w_1, \theta_0)\big]^2 dx > 0.$$

Then, as $n \to \infty$,

$$\sqrt{n/d_n}(\hat\theta_n - \theta_0) \to_D \sigma\, \Sigma_2^{-1/2}\, \mathbb{N}\, L_X^{-1/2}(1, 0), \tag{3.2}$$

where $\mathbb{N}$ *is a standard* q*-dimensional normal random vector, independent of* X_t*. We further have*

$$\Omega_{2n}^{1/2}\, (\hat\theta_n - \theta_0) \to_D \sigma\, \mathbb{N}, \tag{3.3}$$

where $\Omega_{2n} = \sum_{k=1}^{n} \dot f(x_k, w_k, \theta_0)\dot f(x_k, w_k, \theta_0)'$*. Furthermore, if, for all* $x \in R$ *and* $y, t \in R^d$,

$$||\dot f(x, y, \theta_0) - \dot f(x, t, \theta_0)|| \leq T(x)||y - t||^{\delta}(1 + ||y||^{\beta} + ||t||^{\beta}), \quad \textit{for some } \delta > 0,$$

the results (3.2) and (3.3) still hold if we replace ***A2*** *with* ***A2*** *.

Remark 6. Condition (3.1) indicates that $f(x, y, \theta)$ is an integrable function for each fixed y and $\theta \in \Theta$. This result improves and generalizes Theorem 3.2 of Chan and Wang (2015) by using less smoothness in the condition on the regression function $f(x, y, \theta)$ and allowing for a stationary regressor w_k in model (2.1). Note that, as in Theorem 1, result (3.3) has a pivotal limit distribution if a consistency estimator for the conditional variance σ^2 is given (see Remark 1 for such a consistency estimator). Therefore it is usually unnecessary to use the WLS method for such nonlinear cointegrating regression models with integrable regression functions.

3.2 Nonintegrable regression function

This section establishes the limit distribution of $\hat{\theta}_n$ defined by (2.2) with $\lambda(.) \equiv 1$ when f is nonintegrable, for each fixed y and $\theta \in \Theta$. For technical reasons, we require that the regression function $f(x, y, \theta)$ satisfies a certain additional structure. Specifically, we assume that $f(x, y, \theta) = m(x, \theta)K(y)$, where $|K(y)| \leq 1 + ||y||^{\beta}$ and $m(x, \theta)$ satisfies the following condition **A4***.

Write $\dot{m}(x, \theta) = (\dot{m}_1, ..., \dot{m}_q)'$, where $\dot{m}_i = \frac{\partial m(x,\theta)}{\partial \theta_i}$, for $i = 1, ..., q$, and let $\chi(x, \theta)$ be one of m and $\dot{m}_i$, for $i = 1, ..., q$.

A4* Real continuous functions $T_\chi(x)$ and $T(x)$ exist such that

(i) $|\chi(x, \theta) - \chi(x, \theta_0)| \leq ||\theta - \theta_0||^{\alpha}\, T_\chi(x)$, for each $\theta, \theta_0 \in \Theta$ and some $\alpha > 0$;

(ii) for any bounded x,

$$\chi(lx, \theta) \quad = \quad v_\chi(l)\, \widetilde{h}_\chi(x, \theta) \,+\, R(l\, x, \theta),$$

where $v_\chi(l)$ is a positive real function bounded away from zero as l becomes large, $\widetilde{h}_\chi(x, \theta)$, for each $\theta \in \Theta$, is a continuous function and $\sup_{\theta \in \Theta} |R(lx, \theta)|/T_\chi(lx) = o(1)$ as $l \to \infty$.

(iii) $T_\chi(lx) \leq v_\chi(l)\, T(x)$ as $|lx| \to \infty$.

Theorem 3. *Suppose* ***A1****–****A3*** *and* ***A4**** *hold. Suppose that, for each* $\eta > 0$,

$$\int_{|x| \leq \eta} \left[\widetilde{h}_m(x, \theta) - \widetilde{h}_m(x, \theta_0)\right]^2 dx \neq 0, \ \textit{for any } \theta \neq \theta_0, \ \textit{ and}$$
$$\int_{|x| \leq \eta} \widetilde{h}(x, \theta_0)\, \widetilde{h}'(x, \theta_0) dx \ \ \textit{is a positive-definite matrix,} \tag{3.4}$$

where $\widetilde{h}(a, \theta) = \big(\widetilde{h}_{\dot{m}_1}(a, \theta), ..., \widetilde{h}_{\dot{m}_q}(a, \theta)\big)$. *Then, as* $n \to \infty$,

$$\widetilde{D}_n\, (\hat{\theta}_n - \theta_0) \to_D \left(EK^2(w_1) \int_0^1 \Psi(t)\Psi(t)' dt\right)^{-1} \int_0^1 \Psi(t)\, dU_t, \tag{3.5}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

---

where $\widetilde{D}_n = \sqrt{n}\, diag\big(v_{\dot{m}_1}(d_n), ..., v_{\dot{m}_q}(d_n)\big)$*,* $\Psi(t) = \widetilde{h}(X_t, \theta_0)$*,* B_t *and* X_t *are defined as in Section 2.2, and* $(U_t, B_t)_{t\geq 0}$ *is a bivariate Brownian motion with covariance matrix*

$$\Omega = \begin{pmatrix} \widetilde{\sigma}^2 & \rho \\ \rho & 1 \end{pmatrix}, \quad \textit{where } \widetilde{\sigma}^2 = E\big[u_1 K(w_1)\big]^2 \textit{ and } \rho = E\big[\epsilon_1 u_1 K(w_1)\big].$$

Remark 7. In a special case where $K(x) = 1$, Theorem 3 provides a similar result to Theorem 3.4 of Chan and Wang (2015); see also Theorem 3.2 of Wang (2021). Note that the limit result (3.5) is not pivotal, because the unknown covariance ρ is hidden in the joint distribution (U_t, X_t) (e.g., (U_t, B_t), see result (2.4)). Furthermore, the limit distribution in (3.5) is a functional of (U_t, X_t), which is inconvenient in practice, particularly in inference theory where the relevant asymptotic critical value usually depends on it being standard normal.

4. Numerical example

We provide a numerical example to illustrate our asymptotics. This example is designed to show the effects of different weighted functions (in particular, the effects of the different parameters in each weight function) and the performance of the Studentized statistic T_n given in (2.7) for finite sample sizes.

Example 1. Consider the cointegrating regression model defined by

$$y_k \quad = \quad \theta\, |x_{k-1}|^{1/2} + \epsilon_k, \quad \theta_0 = 1, \quad k = 1, 2, ..., n, \tag{4.1}$$

where x_k is generated by one of the following three scenarios:

S1: $x_k = x_{k-1} + \epsilon_k$;

S2: $x_k = x_{k-1} + \xi_k, \quad \xi_k = \rho\, \xi_{k-1} + \epsilon_k$ with $|\rho| < 1$;

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

S3: $x_k = x_{k-1} + \xi_k, \quad (1-B)^d \xi_k = \epsilon_k$, where $0 < d < 1/2$ and $(1-B)^d$ is the fractional difference operator.

In this design, model (4.1) has a martingale structure, as required in theory, and the regressor x_k is generated from a simple random walk by S1, and from short and long integrated processes by S2 and S3, respectively. For the WLS estimator $\widehat{\theta}_n$ defined by (2.2), we consider the following weight functions for comparison:

NW: $\lambda(t) = 1$, that is, no weight function is used;

W1: $\lambda(x) = I(|x| \leq K)$, where $K > 0$ is a constant;

W2: $\lambda(x) = (1 + a\,|x|^4)^{-1}$, where $a > 0$ is a constant:

W3: $\lambda(x) = \exp(-b\,|x|)$, where $b > 0$ is a constant.

It is readily seen that W1 reduces the signal that may affect the asymptotics by removing partial data. Note that $I(|x| \leq K) \to 1$ as $K \to \infty$. As noted in Remark 2, in theory, the consistency of the WLS estimator is better when K is large. For W2 and W3, we use the full data set, but the weight functions have heavy and light tails, respectively. The power 4 is taken in W2 so that the condition in our theorems is satisfied. This can be modified slightly, but is not important to our discussion. By using Remark 2 again, it is expected, in theory, that the consistency of the WLS estimator using W1 and W2 is better when a and b, respectively, are small.

In the simulations, for simplicity of implementation, we assume that $x_0 = 0$, ϵ_k are i.i.d. $N(0,1)$, $b_n = n^{1/3}$, $\rho = 1/2$ in S2, and $d = 0.3$ in S3. Other settings are similar, and the results are available upon request. All simulations have 10000 repetitions. Table 1 shows that the consistency of the WLS estimators is good for all weight functions, even when the sample size n is as small as 100. The consistency is better when the parameter K in W1 is larger, and when a and b are smaller in W2 and W3, respectively, supporting our theoretical results. The best performance is for $K = \infty$ under W1 and $a = b = 0$ under W2 and W3, respectively. Note that $K = \infty$ ($a = b = 0$ as well) corresponds to the consistency of the standard LS estimator (i.e., no weight is used). This indicates that the standard LS estimator has a fast convergence rate, as explained in Remark 2. In Table 2, we further consider the effect of the sample size on the consistency (here, we take $K = 2$, $a = 5$, and $b = 2$). As expected, the consistency of the WLS estimators is always better with a larger sample size, and there is essentially no difference between the weight functions and three scenarios when the sample size is large enough ($n = 500$, say).

Although the standard LS estimator provides a fast consistency rate, the corresponding Student t-statistic $\widetilde{T}_n$ has a nonstandard limit distribution. In fact, under the model (4.1) with the regressor x_k generated by S1, we have

$$\widetilde{T}_n = \frac{\sum_{k=1}^n |x_{k-1}|^{1/2}\epsilon_k}{\sqrt{\sum_{k=1}^n |x_{k-1}|}} \to_D \frac{\int_0^1 |B_t|^{1/2} dB_t}{\sqrt{\int_0^1 |B_t| dt}},$$

where $\{B_t\}_{t\geq 0}$ is a standard Brownian motion; that is, the limit distribution is a functional of a standard Brownian motion rather than of a standard normal. This fact is confirmed by the simulations. Indeed, the density of NW (no weight function, i.e., the density of $\widetilde{T}_n$) in Figure 1 (see also Figures 2 and 3) is clearly away from the standard normal for different sample sizes $n = 100, 500$, and 1000. Figure 1 also provides the densities of the Studentized statistic T_n given in (2.7) with different $K = 1, 1/2$ and $1/10$ using W1 under S1. The density of T_n with $K = 1/10$ is close to the standard normal, and the performance improves as the sample sizes increase. When $K = 1$ and $1/2$, the densities of T_n are not as good as that of $K = 1/10$ (particularly for $K = 1$), and the improvement is not so obvious when the sample size increases from 100 to 1000.

This can be explained theoretically. In fact, under S1 (the discussion is similar under S2

doi:10.5705/ss.202021.0426

Table 1: Means and standard errors of $\hat{\theta}_n - \theta_0$, $n = 100$

W1	$K = 1$	$K = 3$	$K = 5$	$K = \infty$
S1	0.00055 (0.18554)	0.00018 (0.05546)	0.00015 (0.04039)	0.00012 (0.03935)
S2	-0.00249 (0.17897)	-0.00060 (0.06265)	-0.00060 (0.04010)	-0.00034 (0.02831)
S3	-0.00079 (0.18395)	-0.00057 (0.06443)	-0.00048 (0.04192)	-0.00046 (0.02857)
W2	$a = 10$	$a = 5$	$a = 1$	$a = 0$
S1	-0.00865 (0.42538)	-0.00764 (0.34801)	-0.00444 (0.21719)	-0.00011 0.03870
S2	0.00564 (0.35219)	0.00452 (0.29266)	0.00309 (0.19724)	0.00063 (0.02857)
S3	0.00250 (0.47346)	0.00128 (0.37809)	-0.00032 (0.24435)	1.874e-05 (2.404e-02)
W3	$b = 10$	$b = 5$	$b = 1$	$b = 0$
S1	-0.00024 (1.79907)	0.00491 (0.92793)	0.00135 (0.24665)	-0.00037 (0.03920)
S2	0.01353 (1.66340)	0.01023 (0.75282)	0.00208 (0.24818)	3.776e-05 (2.863e-02)
S3	0.00191 (2.73886)	-0.00269 (0.99212)	-0.00021 (0.28759)	0.00017 (0.02403)

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

Table 2: Means and standard errors of $\hat{\theta}_n - \theta_0$

sample size	$n = 100$	$n = 200$	$n = 500$
		$S1$	
NW	-0.00043 (0.03918)	-0.00024 (0.02230)	0.00006 (0.01087)
W1 ($K = 2$)	-0.00233 (0.18209)	-0.00029 (0.12166)	-0.00053 (0.07160)
W2 ($a = 5$)	-0.00217 (0.35129)	0.00014 (0.23513)	-0.00095 (0.14389)
W3 ($b = 2$)	-0.00361 (0.24858)	-0.00058 (0.16568)	-0.00061 (0.09980)
		$S2$	
NW	-9.555e-05 (2.825e-02)	-0.00025 0.01647	3.533e-05 7.908e-03
W1 ($K = 2$)	-0.01260 (0.18550)	0.00309 (0.12791)	0.00061 (0.08158)
W2 ($a = 5$)	-0.00610 (0.30113)	0.00387 (0.21615)	0.00201 (0.14196)
W3 ($b = 2$)	-0.00151 (0.24739)	0.00435 (0.19203)	0.00122 (0.13118)
		$S3$	
NW	-0.00063 0.02416	-0.00015 (0.01328)	-0.00010 (0.00589)
W1 ($K = 2$)	0.00024 (0.22817)	0.00290 (0.17250)	0.00114 (0.12037)
W2 ($a = 5$)	0.00486 (0.37433)	0.00157 (0.27732)	0.00103 (0.19601)
W3 ($b = 2$)	-0.00071 (0.29167)	0.00413 (0.22780)	0.00036 (0.16937)

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

and S3), it follows from (2.7) that (recalling W1 is used)

$$
\begin{aligned}
T_n &= \frac{\sum_{k=1}^n I(|x_{k-1}|/b_n \le K)|x_{k-1}|^{1/2}\epsilon_k}{\sqrt{\sum_{k=1}^n I(|x_{k-1}|/b_n \le K)|x_{k-1}|}} \\
&= \frac{\sum_{k=1}^n I(|x_{k-1}|/\sqrt{n} \le Kc_n^{-1})|x_{k-1}|^{1/2}\epsilon_k}{\sqrt{\sum_{k=1}^n I(|x_{k-1}|/\sqrt{n} \le Kc_n^{-1})|x_{k-1}|}} \to_D N(0,1), \qquad (4.2)
\end{aligned}
$$

for any fixed $K > 0$ and $b_n > 0$ satisfying $c_n = \sqrt{n}/b_n \to \infty$ or $Kc_n^{-1} \to 0$. As noted in Remark 4, although the limit distribution of T_n is free of the values of K when b_n is given such that $c_n \to \infty$, the performance of T_n depends on K in finite-sample simulations. Indeed, when b_n is given ($b_n = n^{1/3}$, say), we have

$$
\begin{array}{rcllll}
 & & K=1, & K=1/2, & K=1/10, & \\
Kc_n^{-1} = Kn^{-1/6} & = & 0.4641598, & 0.2320794, & 0.04641528, & \text{if } n=100, \\
 & & 0.3549537, & 0.1774768, & 0.03549537, & \text{if } n=500, \\
 & & 0.3162278, & 0.1581139, & 0.03162278, & \text{if } n=1000.
\end{array}
$$

Because Kc_n^{-1} with $K = 1/10$ is close to meeting the condition in establishing (4.2) (i.e., $Kc_n^{-1} \to 0$), the performance of T_n in a finite sample size, such as $n = 100, 500$, and 1000, is expected to be better when $K = 1/10$ than when $K = 1$ or $K = 1/2$. Figure 1 confirms this asymptotic theory. Furthermore, when $K = 1$, even under $n = 1000$, $Kc_n^{-1} = 0.3162278$ is far from the required condition in establishing (4.2) (i.e., $Kc_n^{-1} \to 0$). Hence, it is natural that the densities of T_n using W1 with $K = 1$ do not perform good, under given sample sizes $n = 100, 500$ and 1000, as shown in Figure 1.

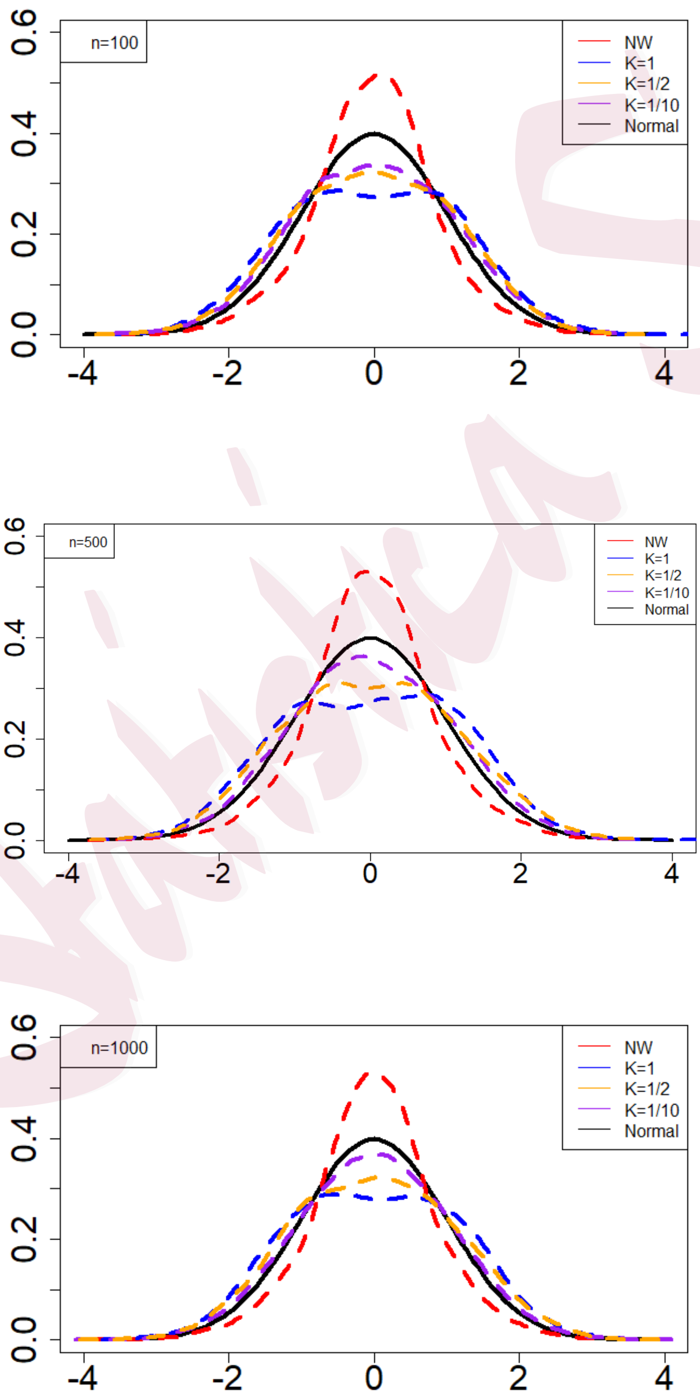
In Figures 2 and 3, the densities of T_n are simulated using W2 and W3, respectively, under S1. As shown in Figures 2 and 3, the ideal results for finite sample sizes can be achieved by using large a and b values, respectively. The theoretical explanation for this is similar to that of W1, and hence the details are omitted. In summary, finite-sample simulations confirm the

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

asymptotic normality of the Studentized statistic T_n given in (2.7). The difference between the weight functions W1, W2, and W3 is not significant, but the choice of the parameter (K, a, and b, respectively) in each weight function has a big effect on the performance of T_n in finite sample sizes. This seems to be natural for the Studentized statistics generated from nonstationary time series.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

Figure 1: T_n densities under W1 with different K



Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

Figure 2: T_n densities under W2 with different a

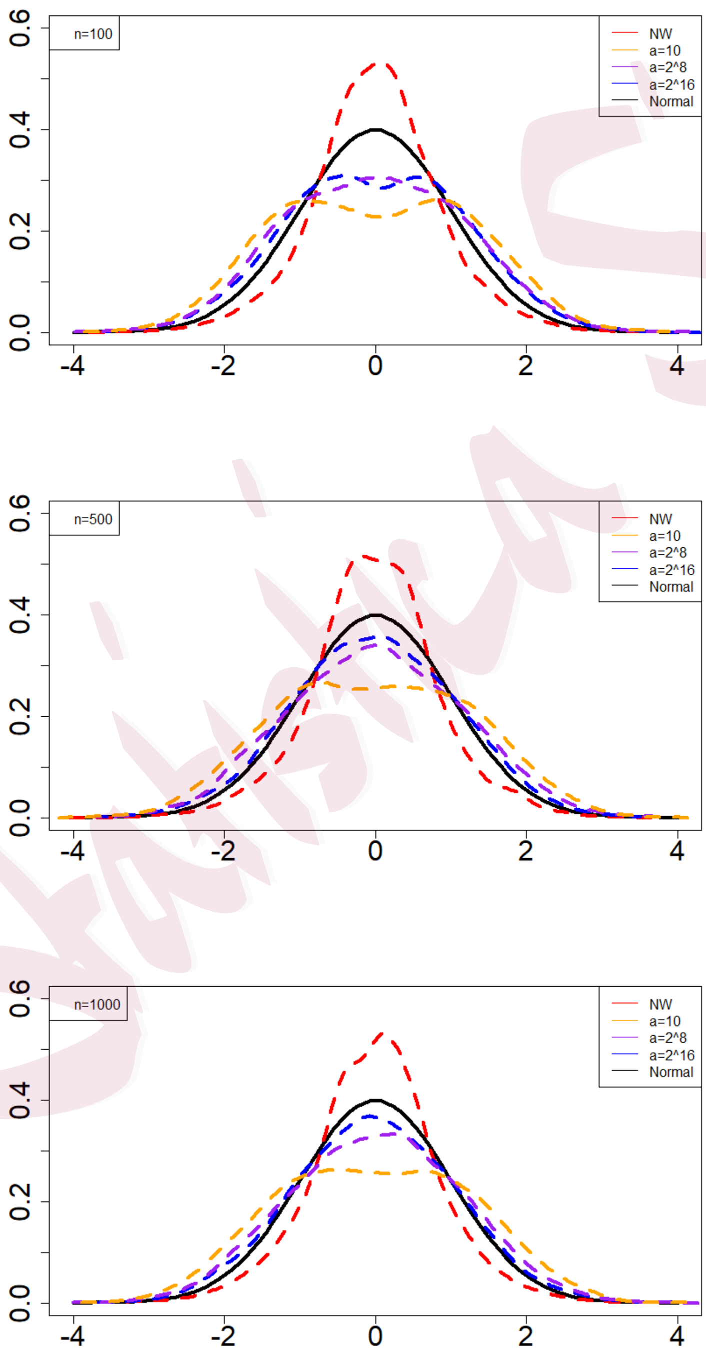
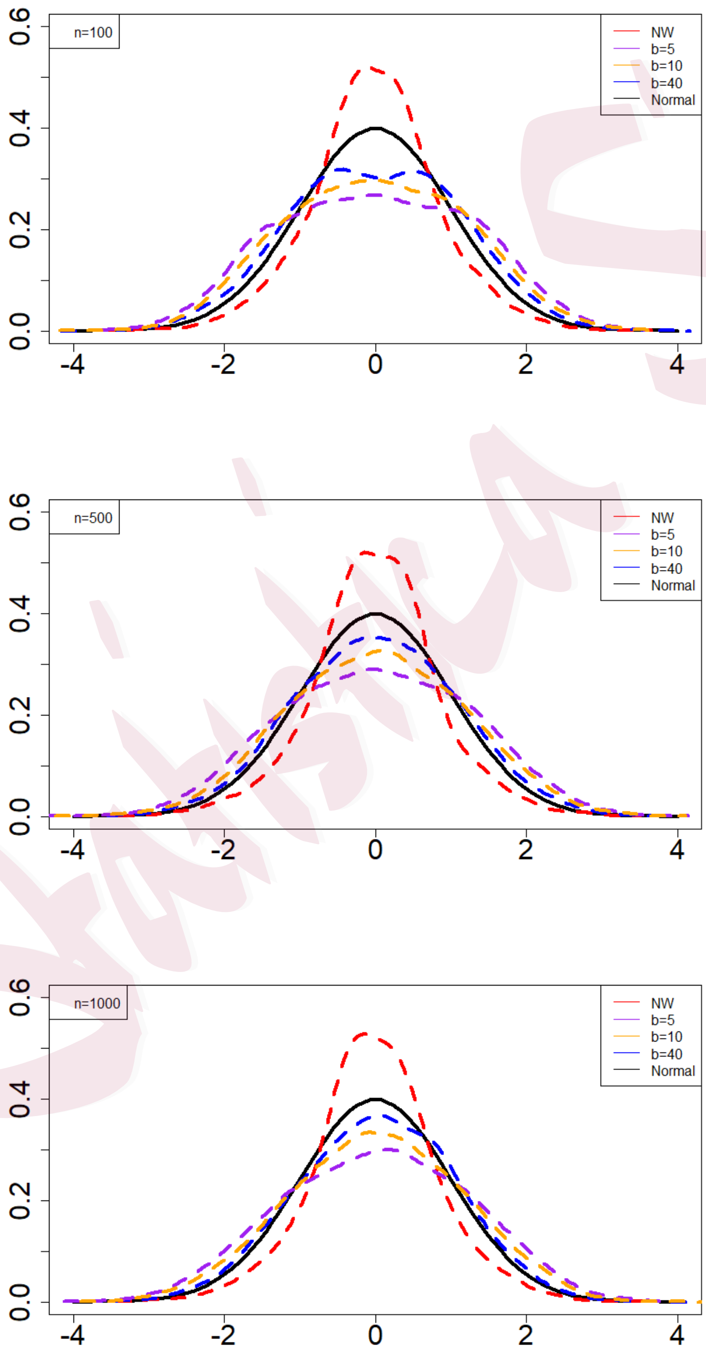


Figure 3: T_n densities under W3 with different b



5. Conclusion

Nonstandard asymptotic behavior appears in regression models with nonstationary time series. The limiting distribution of the standard LS estimator in such regression models often depends on various nuisance parameters, so that the limit results are cumbersome in the relevant asymptotic inferences. In this study, we investigate the WLS estimation in a nonlinear cointegrating regression. Comparied with the standard LS estimator, the WLS estimator has a mixed Gaussian limit, so that the corresponding Studentized statistic converges to a standard normal distribution. This result has advantages in applications, because it is not only convenient in inference theory, but is also free of the memory parameter, even when a fractional process is included in the regressors. There is a convergence rate loss for such a WLS estimator, but it is controllable by using a suitable weight. The ideas presented here work in other areas; such as specification testing related to a cointgerating regression. This is left to future work.

Acknowledgments

The authors are grateful to the editor Professor Rong Chen, associate editor and three anonymous referees for their valuable and constructive comments. Wang's research was supported by the Australian Research Council Discovery Project.

References

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

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

Breitung, J. and Demetrescu, M. (2015). Instrumental variable and variable addition based inference in predictive regressions. *Journal of Econometrics,* **187**(1), 358-375.

Buchmann, B. and Chan, N. H. (2007). Asymptotic theory of least squares estimators for nearly unstable processes under strong dependence. *Ann. Statist.*, **35**, 2001-2017.

Chan, N. and Wang, Q. (2015). Nonlinear regression with nonstationary time series, *Journal of Econometrics*, **185**, 182-195.

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.

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

Ellott, G., Müller, U. K. and Watson, M.W. (2015). Nearly optimal tests when a nuisance parameter is present under the null hypothesis. *Econometrica,* **83**, 771-811.

Demetrescu, M., Georgiev, I., Rodrigues, P. and Taylor, R. (2022). Testing for episodic predictability in stock returns. *Journal of Econometrics*, **227**, 85-113.

Duffy, J. A. (2016). A uniform law for convergence to the local times of linear fractional stable motions. *The Annals of Applied Probability*, **25(1)**, 45-72.

Duffy, J. A. (2020). Asymptotic theory for kernel estimators under moderate deviations from a unit room, with an application to the asymptotic size of nonparametric tests. *Econometric Theory,* **36,** 559-582.

Dong, C. and Linton, O. (2018). Additive nonparametric models with time variables and both stationary and nonstationary regressors. *Journal of Econometrics*, **207**, 212-236.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

Dong, C. Gao, J. and Tjostheim, D. (2016). Estimation for single-index and partially linear single-index integrated model. *Ann. Statist*, **44**, 425-453.

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

Hu, Z., Phillips, P., and Wang, Q. (2021). Nonlinear cointegrating power regression with endogeneity. *Econometric Theory*, **37(6)**, 1173-1213.

Jacob, C. (2010). Conditional least squares estimation in nonstationary nonlinear stochastic regression models. *Ann. Statist.* **38,** 566–597.

Jansson, M. and Moreira, M.J. (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica*, **74**, 681-714.

Kim, S. K. and Kim, I. M. (2012). Partial parametric estimation for nonstationary nonlinear regressions. *Journal of Econometrics*, **167**, 448-457.

Kasparis, I. and Phillips, P.C.B. (2020). Smooth instrumental variables: robust inference in the presence of persistent and heavy-tailed data. In preparation.

Kostakis, A., Magdalinos, T. and Stamatogiannis, M.P. (2015). Robust econometric inference for stock return predictability. *Review of Financial Studies*, **28**(5), 1506-1553.

Kurtz, T. G. and Protter, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability*, **19**, 1035–1070.

Li, D., Tjostheim, D. and Gao, J. (2016). Estimation in nonlinear regression with Harris recurrent Markov chains. *Ann. Statist.*, **44,** 1957-1987.

Lin, Y., Tu, Y. and Yao, Q. (2020). Estimation for double-nonlinear cointegration. *Journal of Econometrics*, **216**, 175-191.

Magdalinos, T. and Phillips, P.C.B. (2009). Econometric inference in the vicinity of unity. *Mimeo*, Singapore Management University.

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

Phillips, P.C.B. (1995). Fully modified least squares and vector autoregression. *Econometrica*, **63**(5) 1023-1078.

Phillips, P.C.B. and Hansen, B. (1990). Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies*, **57**(1), 99-125

Phillips, P. C. B. and Lee, J. H. (2013). Predictive regression under various degrees of persistence and rubust long-horizon regression *Journal of Econometrics*, **177**, 250-264.

Philips P.C.B., Li, D. and Gao, J. (2017). Estimating smooth structural change in cointegration models. *Journal of Econometrics*, **196**, 180-195.

Pollard, D. and Radchenko, P. (2006). Nonlinear least squares estimation. *Journal of Multi. Analysis,* **97**, 548–562.

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

Wang, Q. (2014). Martingale limit theorems revisited and non-linear cointegrating regression. *Econometric Theory*, 30, 509-535.

Wang, Q. (2015). *Limit Theorems for Nonlinear Cointegrating Regression.* World Scientific.

Wang, Q. (2021). Least squares estimation for nonlinear regression models with heteroscedasticity. *Econometric Theory*, **37(6)**, 1267-1289

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

Wang, Q. and Phillips, P.C.B. (2009b). Structural Nonparametric Cointegrating Regression, *Econometrica*, **25**, 1901-1948.

Wang, Q. and Phillips, P.C. B.(2016). Nonparametric cointegrating regression with endogeneity and long memory. *Econometric Theory,* **32,** 359–401.

Wang, Q., Phillips, P.C. B. and Kasparis, I. (2021). Latent Variable Nonparametric Cointegrating Regression, *Econometric Theory,* 37, 138-168.

Yang, B., Long, W., Peng, L. and Cai, Z. (2020). Testing the predictability of U.S. housing price index returns based on an IVX-AR model, *Journal of the American Statistical Association*, **115,** 1598-1619.

School of Mathematics and Statistics, The University of Sydney, NSW 2006, Australia. chunlei.jin@sydney.edu.au

School of Mathematics and Statistics, The University of Sydney, NSW 2006, Australia. E-mail: qiying.wang@sydney.edu.au

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

Appendix

A. Weighted least squares estimation – a framework

Consider a nonlinear regression model having the form:

$$y_k = g_k(\theta) + u_k, \quad k = 1, 2, ..., n, \tag{A.1}$$

where $g_k(.)$ can be random or deterministic functions, $\theta = (\theta_1, ..., \theta_q)$ is a vector of unknown parameters and u_k are the unobservable random disturbances. We assume $\theta \in \Theta$, where Θ is a compact of R^q. The weighted least squares estimator $\widehat{\theta}_n$ of θ in model (A.1) is defined by

$$\widehat{\theta}_n = \text{argmin}_{\theta\in\Theta} Q_n(\theta), \quad \text{where } Q_n(\theta) = \sum_{k=1}^{n} \big[y_k - g_k(\theta)\big]^2 \, l_k \tag{A.2}$$

and l_k is a weight function allowing for dependence on n and g_k, but free of the unknown parameter θ.

The asymptotics of $\widehat{\theta}_n$ without weight effects (i.e., $l_k \equiv 1$) has been widely investigated in literature. See, for instance, Andrews and Sun (2004), Pollard and Radchenko (2006), Jacob (2010) and Chan and Wang (2015). Recently, Wang (2021) established a new framework on the asymptotics of $\widehat{\theta}_n$ that can be easily applied to various nonlinear regression models with heteroscedasticity. This section generalizes and modifies the framework by Wang (2021) by allowing for weight effects and endogeneity in the model (A.1). Throughout this section, let a real parameter value θ_0 of θ be an interior of Θ.

A.1 Consistency of $\widehat{\theta}_n$

To consider the consistency of $\widehat{\theta}_n$, we introduce the following conditions on $g_k(\theta)$.

A5 A sequence of constants $0 < k_n \to \infty$ and a sequence of random variables T_k exist such that

(a) $|g_k(\theta_1) - g_k(\theta_2)| \leq h(||\theta_1 - \theta_2||)\, T_k$ for all $\theta_1, \theta_2 \in \Theta$, where $h(x)$ is a continuous function satisfying $\lim_{x\downarrow 0} h(x) = 0$;

(b) $\frac{1}{k_n}\sum_{k=1}^n l_k T_k\big(|u_k| + T_k\big) = O_P(1)$ and $\frac{1}{k_n}\sum_{k=1}^n l_k\left[g_k(\theta) - g_k(\theta_0)\right] u_k = o_P(1)$ for each $\theta \in \Theta$;

(c) the finite dimensional distributions of $\frac{1}{k_n}\sum_{k=1}^n l_k\left[g_k(\theta) - g_k(\theta_0)\right]^2$ converge to those of $G(\theta)$, where $G(\theta), \theta \in \Theta$ is a stochastic process of θ satisfying $P(\inf_{\substack{||\theta-\theta_0||\geq\delta \\ \theta\in\Theta}} G(\theta) > 0) = 1$ for each $\delta > 0$.

Theorem 4. *Suppose **A5** holds. Then $\hat{\theta}_n$ is a consistent estimator of θ_0, i.e., $||\widehat{\theta}_n - \theta_0|| = o_P(1)$.*

Proof. Write $L_{n,\theta} = \sum_{k=1}^n l_k d_k(\theta)\, u_k$, $D_{n,\theta} = \sum_{k=1}^n l_k d_k^2(\theta)$ and $Q_n(\theta) = \sum_{k=1}^n l_k\big[y_k - g_k(\theta)\big]^2$, where $d_k(\theta) = g_k(\theta) - g_k(\theta_0)$. Given that θ_0 is a real value of θ in the model (A.1), we have

$$\sum l_k u_k^2 \quad = \quad Q_n(\theta_0) \geq Q_n(\hat{\theta}_n) = \sum l_k u_k^2 + D_{n,\hat{\theta}} - 2L_{n,\hat{\theta}}\,.$$

Hence, for any $\epsilon > 0$, we have

$$\begin{aligned} P(||\widehat{\theta}_n - \theta_0|| \geq \epsilon) \quad &\leq \quad P\Big(\sup_{||\theta-\theta_0||\geq\epsilon} |L_{n,\theta}|\,/D_{n,\theta} \geq 1/2\Big) \\ &\leq \quad P\Big(\sup_{\theta\in\Theta} |L_{n,\theta}| \geq 1/2 \inf_{||\theta-\theta_0||\geq\epsilon} D_{n,\theta}\Big), \end{aligned}$$

and Theorem 4 will follow if we prove

(i) $\sup_{\theta\in\Theta} |L_{n,\theta}| = o_P(k_n)$;

(ii) $k_n^{-1} \inf_{\substack{||\theta-\theta_0||\geq\delta, \\ \theta\in\Theta}} D_{n,\theta}$ for any $\delta > 0$, is away from 0 with probability one, as $n \to \infty$;

We only prove (i) since (ii) is the same as that of Proposition 2.1 in Wang (2021) except some routine notation changes. Denote $\mathcal{N}_\eta(\theta_1) = \{\theta : ||\theta - \theta_1|| < \eta\}$. Since Θ is compact, by the finite covering property of compact set, (i) will follow if we prove $L_{n,\theta_1} = o_P(k_n)$ for each $\theta_1 \in \Theta$ and

$$\sup_{\theta \in \mathcal{N}_\eta(\theta_1)} |L_{n,\theta} - L_{n,\theta_1}| \quad \to_P \quad 0, \tag{A.3}$$

as $n \to \infty$ first and then $\eta \to 0$. In fact, since

$$\begin{aligned}\sup_{\theta \in \mathcal{N}_\eta(\theta_1)} |L_{n,\theta} - L_{n,\theta_1}| &\leq \sup_{\theta \in \mathcal{N}_\eta(\theta_1)} \sum_{k=1}^n l_k |g_k(\theta) - g_k(\theta_1)|\, |u_k| \\ &\leq \sup_{\theta \in \mathcal{N}_\eta(\theta_1)} h(||\theta - \theta_1||) \sum_{k=1}^n l_k\, T_k\, |u_k|\end{aligned}$$

by using **A5**(a), result (A.3) follows immediately from **A5**(b), because $h(x)$ is continuous with $h(x) \to 0$ as $x \to 0$. □

Remark 8. Assumption **A5** allows for endogeneity in model (A.1), enabling Theorem 4 quite useful in nonlinear (cointegrating) regression. This result extends Theorem 5.8 of Wang (2015) by allowing for the use of the weight function l_k in (A.2). If $h(x)$ satisfies more smoothness condition and model (A.1) has a martingale structure, we have the following extension of Theorem 2.2 in Wang (2021).

Theorem 5. *Suppose that $\{u_k, \mathcal{F}_k\}_{t\geq 1}$ is a martingale difference with $E(u_k^2 \mid \mathcal{F}_{k-1}) < \infty$, a.s. for each $k \geq 1$. Suppose that l_k and $g_k(\theta)$ for each $\theta \in \Theta$ are adapted to $\mathcal{F}_{k-1}$. Suppose that a sequence of constants $0 < k_n \to \infty$ and a sequence of random variables T_k (adapted to $\mathcal{F}_{k-1}$) exist such that*

(a)′ $|g_k(\theta_1) - g_k(\theta_2)| \leq ||\theta_1 - \theta_2||^\alpha\, T_k$ *for all* $\theta_1, \theta_2 \in \Theta$ *and for some* $\alpha > 0$*;*

(b)′ $\frac{1}{k_n} \sum_{k=1}^n l_k T_k^2 = O_P(1)$ *and* $\sum_{k=1}^n l_k^2\, T_k^2 \big[1 + E(u_k^2 \mid \mathcal{F}_{k-1})\big] = O_P(k_n^2/\log^2 n)$*;*

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

(c)′ ***A5*** *(c) holds.*

Then $\hat{\theta}_n$ is a consistent estimator of θ_0.

Note that $(b)'$ in Theorem 5 is a weaker condition than that of (b) in Assumption **A5**. This is a trade off among other assumptions. The proof of Theorem 5 is similar to Theorem 2.2 of Wang (2021) and hence the details are omitted.

A.2 Asymptotic distribution of $\widehat{\theta}_n$

Let $\dot{g}_k(\theta) = \big(\frac{\partial g_k(\theta)}{\partial \theta_1}, ..., \frac{\partial g_k(\theta)}{\partial \theta_q}\big)'$ be the first derivative of $g_k(\theta)$, $Z_n(\theta) = (D_n^{-1})' \sum_{k=1}^n l_k\, \dot{g}_k(\theta)\, u_k$ and

$$Y_n = (D_n^{-1})' \sum_{k=1}^{n} l_k\, \dot{g}_k(\theta_0) \dot{g}_k(\theta_0)'\, D_n^{-1},$$

where $D_n = \text{diag}(\widetilde{D}_n, ..., d_{qn})$ is a sequence of diagonal matrices satisfying $n^{-\delta} \min_{1\le j\le q} d_{jn} \to \infty$ for some $\delta > 0$. For the asymptotic distribution of $\widehat{\theta}_n$, we have the following extension of Theorem 2.1 in Wang (2021).

Theorem 6. *Suppose that*

(i) $Y_n \to_D M$, *where the smallest eigenvalue of M is almost surely positive (i.e., $M > 0$, a.s.);*

(ii) $Z_n(\theta_0) = O_P(1)$, $\sup_{\theta\in\Theta} ||Z_n(\theta) - Z_n(\theta_0)|| = O_P(\log^{1/2} n)$ *and*

$$\sup_{||D_n(\theta-\theta_0)|| \le \log n} ||Z_n(\theta) - Z_n(\theta_0)|| = o_P(1);$$

(iii) $\sup_{||\theta-\theta_0||\le\delta} \sum_{k=1}^n l_k ||D_n^{-1}[\dot{g}_k(\theta) - \dot{g}_k(\theta_0)]||^2 = o_P(1)$, *as $n \to \infty$ first and then $\delta \to 0$.*

For any estimator $\widehat{\theta}_n$ of θ_0 satisfying (A.2) such that $\widehat{\theta}_n \to_P \theta_0$, we have

$$D_n(\hat{\theta}_n - \theta_0) = Y_n^{-1} Z_n(\theta_0) + o_P(1). \tag{A.4}$$

If in addition $(Y_n, Z_n) \to_D (M, Z)$ where $M > 0$, a.s., then $D_n(\hat{\theta}_n - \theta_0) \to_D M^{-1} Z$.

Proof. The idea is similar to Theorem 2.1 of Wang (2021) since, except obvious notation changes, Assumption 2.1 and Assumption 2.2 (i) in the cited paper are only used to prove the conditions (ii) and (iii) of Theorem 6. We omit the details. □

Remark 9. As noticed in Remark 8, Theorem 6 is useful in nonlinear cointegrating regression as endogeneity is allowed in model (A.1). When model (A.1) has a martingale structure, we have the following corollary, providing an extension of Theorem 2.1 of Wang (2021) to weighted LS. Let $\{\mathcal{F}_k\}_{k\geq 0}$ be an increasing sequence of σ-fields on some probability space $(\Omega, \mathcal{F}, P)$ with $\mathcal{F}_0 = \sigma(\phi, \Omega)$.

Corollary 2. *Suppose that*

(i) $\{u_k, \mathcal{F}_k\}_{k\geq 1}$ *forms a martingale difference with* $E(u_k^2 \mid \mathcal{F}_{k-1}) < \infty$, *a.s. for each* $k \geq 1$;

(ii) $g_k(\theta)$ *for each* $\theta \in \Theta$ *and* l_k *are adapted to* $\mathcal{F}_{k-1}$;

(iii) $||D_n^{-1}\big[\dot{g}_k(\theta_1) - \dot{g}_k(\theta_2)\big]|| \leq ||\theta_1 - \theta_2||^\alpha\, T_{nk}$ *for some* $0 < \alpha \leq 1$ *and for any* $\theta_1, \theta_2 \in \Theta$, *where* T_{nk} *is adapted to* $\mathcal{F}_{k-1}$ *for each* $n \geq 1$, *satisfying*

$$\sum_{k=1}^n l_k^2\, T_{nk}^2 \big[1 + E(u_k^2|\mathcal{F}_{k-1})\big] + \sum_{k=1}^n l_k\, T_{nk}^2 = O_P(1); \tag{A.5}$$

(iv) $Z_n(\theta_0) = O_P(1)$ *and* $Y_n \to_D M$, *where* $M > 0$, *a.s., i.e., the smallest eigenvalue of* M *is almost surely positive.*

Then, for any estimator $\widehat{\theta}_n$ of θ_0 satisfying (A.2) such that $\widehat{\theta}_n \to_P \theta_0$, result (A.4) still holds.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

Proof. It follows from the same arguments as that of Corollary 2.1 in Wang (2021), i.e., under conditions (i)-(iii), we may prove $\sup_{\theta\in\Theta}||Z_n(\theta)-Z_n(\theta_0)||=O_P(\log^{1/2}n)$ and

$$\sup_{||D_n(\theta-\theta_0)||\le\log n}||Z_n(\theta)-Z_n(\theta_0)||=o_P(1),$$

by using Theorem 2.3 of Wang (2021). Then result follows from Theorem 6. □

B. Convergence to local time and a mixture of normal distributions

The proofs of Theorems 1, 2 and 3 depend on certain fundamental results on convergence to local time and a mixture of normal distributions, which are summarized in this section. Except mentioned explicitly, notation is the same as in Section 2.

Recall $(u_k,\mathcal{F}_k)_{k\ge1}$, where $\mathcal{F}_k=\sigma(v_{k+1},v_k,...)$, is a sequence of stationary martingale differences with $E(u_k^2|\mathcal{F}_{k-1})=\sigma^2<\infty$. Let $x_{nk}=x_k/d_n$, where $d_n^2=var(\sum_{k=1}^n\xi_k)$. Let $g(.),g_1(.),...$ and other related $G(.),T(.)$, etc be Borel measurable functions on their components. For the convergence to a local time process, the following result comes from Theorem 2.21 of Wang (2015).

Lemma 1. *Write $\widetilde{g}(x)=\sup_{0\le s\le1}|g(s,x)|$ and suppose that $\widetilde{g}(x)$ is a bounded and integrable function on R. Then, for any $c_n\to\infty$ satisfying $c_n/n\to0$, we have*

$$\begin{aligned}&\Big(\frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]}\epsilon_k,\ \frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]}\epsilon_{-k},\ x_{n,[nt]},\ \frac{c_n}{n}\sum_{k=1}^{n}g\big(k/n,\ c_n\,x_{nk}\big)\Big)\\&\Rightarrow\Big(B_t,\ B_{-t},\ X_t,\ \int_0^1G(s)dL_X(s,0)\Big),\end{aligned}\tag{B.1}$$

on $D_{R^4}[0,\infty)$, where $G(s)=\int_{-\infty}^{\infty}g(s,x)dx$.

By using Lemma 1, together with the extended martingale limit theorem by Wang (2014)

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

(also see Chapter 3 of Wang (2015)), we may establish following Theorems 7 and 8, which are used for the purpose of this paper.

Let $V_{k,m} = (v_k, ..., v_{k-m})$ and $V_m = V_{m,m}$, where $m \geq 0$ is a fixed integer, and write

$$S_n = \frac{c_n}{n}\sum_{k=1}^n g_1(k/n, c_n\, x_{nk}, V_{k,m}), \quad M_n = \big(\frac{c_n}{n}\big)^{1/2}\sum_{k=1}^n g_2(k/n, c_n\, x_{nk}, V_{k,m})\, u_k,$$

where c_n is a sequence of positive constants.

Theorem 7. *Suppose that, for any $x \in R$ and $y \in R^{m+1}$ and for some $\beta > 0$,*

$$\sup_{0\leq s\leq 1}|g_1(s,x,y)| + \sup_{0\leq s\leq 1}|g_2(s,x,y)| \leq T(x)(1+||y||^{\beta}), \tag{B.2}$$

where $T(x)$ is a bounded and integrable function. Suppose that $E||v_0||^{2\wedge(2\beta)} < \infty$. Then, for any $c_n \to \infty$ satisfying $c_n/n \to 0$, we have

$$\Big(\frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]}\epsilon_k,\ \frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]}\epsilon_{-k},\ x_{n,[nt]},\ S_n\Big) \Rightarrow \Big(B_t,\ B_{-t},\ X_t,\ \int_0^1 G_1(s)\, dL_X(s,0)\Big) \tag{B.3}$$

on $D_{R^4}[0,\infty)$, where $G_1(s) = \int_{-\infty}^{\infty} Eg_1(s,x,V_m)dx$. In in addition $E||v_0||^{4\beta+2} < \infty$, we further have

$$\big(S_n,\ M_n\big) \quad \to_D \quad \Big(\int_0^1 G_1(s)\, dL_X(s,0), \big[\int_0^1 G_2(s)\, dL_X(s,0)\big]^{1/2} N\Big), \tag{B.4}$$

where $G_2(s) = \int_{-\infty}^{\infty} Eg_2^2(s,x,V_m)dx$ and N is a standard normal variate independent of X.

Proof. By virtue of Lemma 1, to prove (B.3), it suffices to show that, under (B.2) and $E||v_0||^{2\wedge(2\beta)} < \infty$,

$$\begin{aligned} & S_n - \frac{c_n}{n}\sum_{k=1}^n \hat{g}_1\big[k/n, c_n\, x_{nk}\big] \\ = \ & \frac{c_n}{n}\sum_{k=1}^n \big\{g_1\big[k/n, c_n\, x_{nk}, V_{k,l}\big] - \hat{g}_1\big[k/n, c_n\, x_{nk}\big]\big\} = o_P(1), \end{aligned} \tag{B.5}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

where $\hat{g}_1(s,x) = Eg_1(s,x,V_m)$. This is essentially the same as in the proof of (A.20) with $i=2$ in Wang et al. (2021) and hence the details are omitted.

We next prove (B.4). It follows from (B.5) and the similar arguments as in the proof of Theorem 2.19 in Wang (2015) that

$$\begin{aligned} S_n &= \frac{c_n}{n}\sum_{k=1}^n \hat{g}_1\big[k/n, c_n\, x_{nk}\big] + o_P(1) \\ &= \frac{1}{n}\sum_{k=1}^n G_1(k/n)\,\phi_\epsilon(x_{nk}) + R_n(\epsilon) = \int_0^1 G_1(s)\phi_\epsilon(x_{n,[ns]})ds + R_n(\epsilon), \end{aligned}$$

where $\phi_\epsilon(x) = \frac{1}{\epsilon\sqrt{2\pi}}e^{-x^2/(2\epsilon^2)}$ and $R_n(\epsilon) \to_P 0$ as $n \to \infty$ first and then $\epsilon \to 0$. Since $\int_0^1 G_1(s)\phi_\epsilon(Z_{n,[ns]})ds$ is a continuous functional of the process $\{x_{n,[nt]}\}_{0\le t\le 1}$, the continuous mapping theorem indicates that (B.4) will follow if we prove

$$\big(x_{n,[nt]},\ M_n\big) \quad \Rightarrow \quad \Big(X_t,\ \big[\int_0^1 G_2(s)\,dL_X(s,0)\big]^{1/2} N\Big), \tag{B.6}$$

on $D_{R^2}[0,1]$. Note that $M_n = \sum_{k=1}^n m_{nk}u_k$, where $m_{nk} = \big(\frac{c_n}{n}\big)^{1/2} g_2(k/n, c_n\, x_{nk}, V_{k,m})$ depends only on $v_k, v_{k-1}, \ldots$. We may establish (B.6) by using the extended martingale limit theorem given in Wang (2014). Indeed, by noting that, for any $-1\le\delta\le 1/\beta$,

$$\begin{aligned} &\Big(\frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]}\epsilon_k,\ \frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]}\epsilon_{-k},\ x_{n,[nt]},\ \frac{c_n}{n}\sum_{k=1}^n |g_2(k/n, c_n\, x_{nk}, V_{k,m})|^{2+\delta}\Big) \\ &\Rightarrow \Big(B_t,\ B_{-t},\ X_t,\ \int_0^1 \widetilde{G}_2(s)dL_X(s,0)\Big), \end{aligned}$$

(using (B.3) with $g_1(.) = |g_2(.)|^{2+\delta}$) where $\widetilde{G}_2(s) = \int_{-\infty}^{\infty} E|g_2(s,x,V_m)|^{2+\delta}dx$, it is readily seen that

$$\max_{1\le k\le 1}|m_{nk}| \quad \le \quad \Big[\big(\frac{c_n}{n}\big)^{1+1/(2\beta)}\sum_{k=1}^n |g_2(k/n, c_n\, x_{nk}, V_{k,m})|^{2+1/\beta}\Big]^{2\beta/(1+2\beta)} = o_P(1)$$

and

$$\frac{1}{\sqrt{n}}\sum_{k=1}^n |m_{nk}| \quad = \quad c_n^{-1/2}\,\frac{c_n}{n}\sum_{k=1}^n |g_2(k/n, c_n\, x_{nk}, V_{k,m})| = o_P(1),$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

due to $c_n \to \infty$ and $c_n/n \to 0$. Result (B.6) follows from Theorem 2.1 of Wang (2014). $\square$

We next establish a more general result than that of Theorem 7. Write

$$\begin{aligned} S_{1n} &= \frac{c_n}{n}\sum_{k=1}^{n} g_3(k/n, c_n\, x_{nk}, w_k + \nu_{nk}), \\ M_{1n} &= \big(\frac{c_n}{n}\big)^{1/2}\sum_{k=1}^{n} g_4(k/n, c_n\, x_{nk}, w_k + \nu_{nk})\, u_k, \end{aligned}$$

where the following additional conditions are used:

C1 (i) $w_k = (w_{1k}, ..., w_{dk})$, where, for $i = 1, ..., d$, $w_{ik} = \sum_{j=0}^{\infty} \psi_{i,j}\, v'_{k-j}$ with $\psi_{i,j} = (\psi_{i,1j}, \psi_{i,2j})$ satisfying $\sum_{j=0}^{\infty}(|\psi_{i,1j}| + |\psi_{i,2j}|) < \infty$;

(ii) $\nu_{nk} = (\nu_{1,nk}, ..., \nu_{d,nk})$, where, for $i = 1, ..., d$, $|\nu_{i,nk}| \le \delta_n \sum_{j=0}^{\infty}\big(|\varphi_{1j}\, \epsilon_{k-j}| + |\varphi_{2j}\, \eta_{k-j}|\big)$ with $\sum_{j=0}^{\infty}(|\varphi_{1j}| + |\varphi_{2j}|) < \infty$ and $\delta_n \to 0$, as $n \to \infty$;

C2 there exist $\delta > 0$, integer $\beta \ge 1$ and a bounded and integrable function $T(x)$ such that, for all $x \in R$ and $y, t \in R^d$ and for $i = 3$ and 4,

(i) $\sup_{s\in[0,1]} |g_i(s, x, y)| \le T(x)(1 + ||y||^{\beta})$;

(ii) $\sup_{s\in[0,1]} |g_i(s, x, y) - g_i(s, x, t)| \le T(x)||y - t||^{\delta}\,(1 + ||y||^{\beta} + ||t||^{\beta})$.

Theorem 8. *Suppose that **C1** and **C2** hold and $E||v_0||^{2\beta+2} < \infty$ where β is given as in **C2**. For any $c_n \to \infty$, $c_n/n \to 0$ and $z \in R$, we have*

$$\Big(\frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]} \epsilon_k,\ \frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]} \epsilon_{-k},\ x_{n,[nt]},\ S_{1n}\Big) \ \Rightarrow\ \Big(B_t,\ B_{-t},\ X_t,\ \int_0^1 G_3(s)\, dL_X(s, 0)\Big) \qquad \text{(B.7)}$$

on $D_{R^4}[0, \infty)$, where $G_3(s) = \int_{-\infty}^{\infty} Eg_3(s, x, w_0)dx$. If in addition $E||v_0||^{4\beta+2} < \infty$, we further have

$$(S_{1n},\ M_{1n}) \quad \to_D \quad \Big(\int_0^1 G_3(s)\, dL_X(s, 0), \big[\int_0^1 G_4(s)\, dL_X(s, 0)\big]^{1/2} N\Big), \qquad \text{(B.8)}$$

where $G_4(s) = \int_{-\infty}^{\infty} Eg_4^2(s, x, w_0)dx$ and N is a standard normal variate independent of X.

Proof. We only prove (B.7), since the proof of (B.8) is the same as that of (B.4) by an application of the extended martingale limit theorem given in Wang (2014). For any $l \geq 0$, write $w_k(l) = \big(w_{1k}(l), ..., w_{dk}(l)\big)$, where, for $i = 1, ..., d$, $w_{ik}(l) = \sum_{j=0}^{l} \psi_{i,j}\, v'_{k-j}$. It follows from Theorem 7 that, for each $l \geq 1$,

$$\frac{c_n}{n} \sum_{k=1}^{n} g_3\big[k/n, c_n(x_{nk} + c'_n z), w_k(l)\big] \quad \to_D \quad \int_0^1 G_{3,l}(s) d\, \widetilde{L}_X(s, z), \tag{B.9}$$

where $G_{3,l}(s) = \int_{-\infty}^{\infty} E g_3(s, x, w_0(l)) dx$. Using **C2** (ii), we have

$$\begin{aligned} \sup_{s\in[0,1]} |G_{3,l}(s) - G_3(s)| \quad &\leq \quad C\, E\big(||w_0(l) - w_0||^{\delta} (1 + ||w_0||^{\beta} + ||w_0(l)||^{\beta})\big) \\ &\leq \quad C\, \big(\sum_{i=0}^{d} \sum_{j=l}^{\infty} |\psi_{i,j}|^2\big)^{1/2} \big(E||w_0||^{2\beta+2}\big)^{1/2} \to 0, \end{aligned}$$

as $l \to \infty$, i.e.,

$$\int_0^1 G_{3,l}(s) d\, \widetilde{L}_X(s, z) = \int_0^1 G_3(s) d\, \widetilde{L}_X(s, z) + o_P(1), \tag{B.10}$$

as $l \to \infty$. In terms of (B.9) and (B.10), (B.7) will follow if we prove

$$S_{1n} - \frac{c_n}{n} \sum_{k=1}^{n} g_3\big[k/n, c_n(x_{nk} + c'_n z), w_k(l)\big] = o_P(1),$$

as $n \to \infty$ first and then $l \to \infty$. Under **C2**(ii), this is similar to that of (A.20) of Wang et al. (2021) with $i = 1$ and hence the details are omitted. □

C. Proofs of the main results

We prove Theorems 1, 2, 3 and Corollary 1 by checking the conditions of Theorem 6, where we require Theorems 7 and 8 for the results on the convergence to local time and a mixture of normal distributions.

Proof of Theorem 1. We start with a outline. Let $D_n = \big(n/c_n\big)^{1/2} \operatorname{diag}\big(v_{\dot f_1}(b_n), ..., v_{\dot f_q}(b_n)\big)$, $c_n = d_n/b_n, x_{nk} = x_k/d_n$ and

$$\begin{aligned} Z_n(\theta) &= D_n^{-1} \sum_{k=1}^n \lambda(x_k/b_n)\, \dot f(x_k, w_k, \theta)\, u_k, \\ Y_n(\theta) &= \big(D_n^{-1}\big)' \sum_{k=1}^n \lambda(x_k/b_n) \dot f(x_k, w_k, \theta) \dot f(x_k, w_k, \theta)'\, D_n^{-1}. \end{aligned}$$

Note that, owing to **A4** (i) and (iii),

$$\begin{aligned} & ||D_n^{-1}\big[\dot f(x_k, w_k, \theta_1) - \dot f(x_k, w_k, \theta_2)\big]|| \\ \le\ & ||\theta_1 - \theta_2||^\alpha\, (n/c_n)^{-1/2} \sum_{i=1}^q v_{\dot f_i}^{-1}(b_n) T_{\dot f_i}(x_k, w_k) \\ \le\ & q\, ||\theta_1 - \theta_2||^\alpha\, (n/c_n)^{-1/2}\, T(c_n x_{nk})(1 + |w_k|^\beta), \end{aligned}$$

for any $\theta_1, \theta_2 \in \Theta$. It follows from Corollary 2 with $g_k(\theta) = f(x_k, w_k, \theta)$ that, to prove Theorem 1, it suffices to show that

(i) $\hat\theta_n$ is a consistent estimator of θ_0;

(ii) for any bounded and integrable function $g(x)$,

$$\frac{c_n}{n} \sum_{k=1}^n E\big[g(c_n\, x_{nk})(1 + |w_k|^{2\beta})\big] = O(1); \tag{C.1}$$

(iii) for a q-dimensional standard normal vector $\mathbb{N}$ that is independent of X,

$$\big\{Z_n(\theta_0),\ Y_n(\theta_0)\big\} \quad \to_D \quad \big\{\sigma \Sigma_1^{1/2}\, L_X(1,0)^{1/2}\mathbb{N},\ \Sigma\, L_X(1,0)\big\}. \tag{C.2}$$

Result (C.1) is well-known in literature. See, for instance, (7.22) of Wang et al. (2021).

We next prove (C.2). We may write

$$\begin{aligned} Z_n(\theta_0) &= D_n^{-1} \sum_{k=1}^n \lambda(x_k/b_n)\, \dot f(x_k, w_k, \theta_0)\, u_k, \\ &= \big(\frac{c_n}{n}\big)^{1/2} \sum_{k=1}^n \lambda(c_n\, x_{nk})\, \dot h(c_n x_{nk}, w_k, \theta_0)\, u_k + R_n \\ &:= Z_{n1}(\theta_0) + R_n, \end{aligned} \tag{C.3}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

where $R_n = \sum_{k=1}^n a_{nk} u_k$ with

$$a_{nk} = \lambda(x_k/b_n)\big[D_n^{-1}\dot f(x_k, w_k, \theta_0) - \big(\frac{c_n}{n}\big)^{1/2}\, \dot h(x_k/b_n, w_k, \theta_0)\big].$$

It follows from **A4**(ii) and (iii) with $l = b_n$ that, for any $\epsilon > 0$, there exists a n_0 so that when $n \geq n_0$,

$$||a_{nk}|| \leq \epsilon\, q\, \big(\frac{c_n}{n}\big)^{1/2} \lambda(x_k/b_n)\, T(x_k/b_n)(1 + ||w_k||^\beta).$$

This, together with (C.1) with $g(x) = \lambda^2(x)T^2(x)$, yields that, for any $\epsilon > 0$, there exists a n_0 so that when $n \geq n_0$,

$$ER_n^2 \leq C \sum_{k=1}^n E||a_{nk}||^2 \leq C\epsilon\, \frac{c_n}{n} \sum_{k=1}^n E\big[\lambda^2(x_k/b_n)\, T^2(x_k/b_n)(1 + ||w_k||^{2\beta})\big] \leq C\epsilon,$$

i.e., $||R_n|| = o_P(1)$. Similarly, we have

$$\begin{aligned} Y_n(\theta_0) &= \big(D_n^{-1}\big)' \sum_{k=1}^n \lambda(x_k/b_n)\dot f(x_k, w_k, \theta_0)\dot f(x_k, w_k, \theta_0)'\, D_n^{-1} \\ &= \frac{c_n}{n} \sum_{k=1}^n \lambda(c_n\, x_{nk})\dot h(c_n\, x_{nk}, w_k, \theta_0)\dot h(c_n\, x_{nk}, w_k, \theta_0)' + R_{1n} \\ &= Y_{n1}(\theta_0) + R_{1n}, \end{aligned} \tag{C.4}$$

where $||R_{1n}|| = o_P(1)$.

By virtue of (C.3) and (C.4), result (C.2) follows from the fact: for any $\alpha_i' = (\alpha_{i1}, ..., \alpha_{iq}) \in \mathbb{R}, i = 1, 2, 3,$

$$\big(\alpha_1'\, Y_{n1}(\theta_0)\alpha_2,\ \ \alpha_3' Z_{n1}(\theta_0)\big) \ \ \to_D \ \ \big(\alpha_1'\, \Sigma\, \alpha_2\, L_X(1,0),\ \ \big[\alpha_3'\, \Sigma_1\, \alpha_3\, L_X(1,0)\big]^{1/2} N\big), \tag{C.5}$$

where we have used Theorem 7.

We finally prove $\hat\theta_n$ is a consistent estimator of θ_0. Note that $\frac{c_n}{n v_f^2(b_n)} \sum_{k=1}^n T^2(x_k/b_n)(1 + |w_k|^\beta)^2 = O_P(1)$ and, owing to **A4** (i) and (iii) with $p(.) = f(.)$,

$$|f(x_k, w_k, \theta_1) - f(x_k, w_k, \theta_2)| \leq ||\theta_1 - \theta_2||^\alpha\, v_f(b_n)\, T(x_k/b_n)(1 + |w_k|^\beta),$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

for any $\theta_1, \theta_2 \in \Theta$. Using Theorem 4 with $g_k(\theta) = f(x_k, w_k, \theta)$ and $k_n = \frac{c_n}{n v_f^2(b_n)}$, it suffices to show that, for any $\theta_j \in \Theta$ and $\alpha_j \in R$, $j = 1, 2, ..., q$,

$$\sum_{j=1}^{q} \alpha_j \, G_n(\theta_j) \quad \to_D \quad \sum_{j=1}^{q} \alpha_j G(\theta_j), \tag{C.6}$$

where

$$\begin{aligned} G_n(\theta) &= \frac{c_n}{n v_f^2(b_n)} \sum_{k=1}^{n} \lambda(x_k/b_n) \left[f(x_k, w_k, \theta) - f(x_k, w_k, \theta_0) \right]^2, \\ G(\theta) &= \int_{-\infty}^{\infty} \lambda(x) E\big[h_f(x, w_1, \theta) - h_f(x, w_1, \theta_0) \big]^2 dx \, L_X(1, 0), \end{aligned}$$

since $P\big(\min_{||\theta-\theta_0|| \geq \delta} G(\theta) > 0\big) = 1$ for each $\delta > 0$, owing to **A4** (iv) and the fact that $P(L_X(1,0) > 0) = 1$.

In order to establish (C.6), we write

$$\begin{aligned} R_n(\theta) &= \frac{c_n}{n v_f^2(b_n)} \sum_{k=1}^{n} \lambda(x_k/b_n) \left[f(x_k, w_k, \theta) - v_f(b_n) h_f(x_k/b_n, w_k, \theta) \right]^2, \\ G_{1n}(\theta) &:= \frac{c_n}{n} \sum_{k=1}^{n} \lambda(x_k/b_n) \left[h_f(x_k/b_n, w_k, \theta) - h_f(x_k/b_n, w_k, \theta_0) \right]^2. \end{aligned}$$

As that of $R_n = o_P(1)$, it follows from **A4** (ii) and (iii) with $l = b_n$ that

$$\begin{aligned} \sup_{\theta \in \Theta} |R_n(\theta)| &\leq \frac{o(1)\, c_n}{n v_f^2(b_n)} \sum_{k=1}^{n} \lambda(x_k/b_n) T_f^2(x_k, w_k) \\ &= o(1) \frac{c_n}{n} \sum_{k=1}^{n} \lambda(x_k/b_n) T^2(x_k/b_n)(1 + |w_k|^\beta) \\ &= o_P(1), \end{aligned} \tag{C.7}$$

as $n \to \infty$. On the other hand, as in the proof of (C.5),

$$\sum_{j=1}^{q} \alpha_j \, G_{1n}(\theta_j) \quad \to_D \quad \sum_{j=1}^{q} \alpha_j \, G(\theta_j). \tag{C.8}$$

Now, by noting $G_n(\theta)$ can be decomposed into $G_n(\theta) = G_{1n}(\theta) + \Delta_n(\theta)$, where

$$\Delta_n(\theta) \quad \leq \quad 2 \sup_{\theta \in \Theta} |R_n(\theta)| + 4 \, G_{1n}^{1/2}(\theta) \sup_{\theta \in \Theta} |R_n(\theta)|^{1/2},$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

we have $\sum_{j=1}^q \alpha_j \Delta_n(\theta_j) = o_P(1)$ and (C.6).

The proof of Theorem 1 is now complete. $\square$

Proof of Corollary 1. It is similar to that of Theorem 1 except using Theorem 8 instead of Theorem 7. We omit the details. $\square$

Proof of Theorem 2. For the first part, i.e., the proofs of (3.2) and (3.3), the idea is essentially the same as that of Theorem 1. Indeed, by recalling that $T(x)$ is bounded and integrable and

$$||\dot{f}(x_k, w_k, \theta_1) - \dot{f}(x_k, w_k, \theta_2)|| \quad \leq \quad ||\theta_1 - \theta_2||^\alpha T(x_k)(1 + ||w_k||^\beta),$$

as in Theorem 1 by using Corollary 2 with $l_k \equiv 1$ and $g_k(\theta) = f(x_k, w_k, \theta)$, it suffices to show: (a) $\hat{\theta}_n$ is a consistent estimator of θ_0; (b)

$$\frac{d_n}{n} \sum_{k=1}^n E\left[g(x_k)(1 + |w_k|^{2\beta})\right] = O(1)$$

for any bounded and integrable function $g(x)$; and (c)

$$\Big\{\widehat{Y}_n(\theta_0),\ \widehat{Z}_n(\theta_0)\Big\} \to_D \Big\{\Sigma_2\, L_X(1,0),\ \sigma\, \mathbb{N}\left[\Sigma_2 L_X(1,0)\right]^{1/2}\Big\},$$

where

$$\widehat{Z}_n(\theta) \quad = \quad (\frac{d_n}{n})^{1/2} \sum_{k=1}^n \dot{f}(x_k, w_k, \theta)\, u_k,$$
$$\widehat{Y}_n(\theta) \quad = \quad \frac{d_n}{n} \sum_{k=1}^n \dot{f}(x_k, w_k, \theta) \dot{f}(x_k, w_k, \theta)'.$$

and $\mathbb{N}$ is a standard q-dimensional normal random vector independent of X_t. In fact, by recalling $x_{nk} = x_k/d_n$, result (c) comes from a direct application of Theorem 7; (b) is the same as (C.1) with $c_n = d_n$ and (a) is similar to that of (i) in the proof of Theorem 1 with some routine notation changes.

If **A2** is replaced by **A2** *, to show (3.2) and (3.3) still hold, we only need to replace Theorem 7 by Theorem 8 and hence the details are omitted. $\square$

Proof of Theorem 3. As in Theorem 1, we establish Theorem 3 by using Corollary 2 with $g_k(\theta) = m(x_k, \theta)K(w_k)$, but with different details. Indeed, owing to **A4*** (i) and (iii), we have

$$\begin{aligned} ||\widetilde{D}_n^{-1}\left[\dot{g}_k(\theta_1) - \dot{g}_k(\theta_2)\right]|| &\leq ||\theta_1 - \theta_2||^{\alpha}\, n^{-1/2} \sum_{i=1}^{q} v_{\dot{m}_i}^{-1}(d_n) T_{\dot{m}_i}(x_k)(1 + ||w_k||^{\beta}) \\ &\leq q\, ||\theta_1 - \theta_2||^{\alpha}\, n^{-1/2}\, T(x_{nk})(1 + ||w_k||^{\beta}), \end{aligned}$$

for any $\theta_1, \theta_2 \in \Theta$, where $x_{nk} = x_k/d_n$ and $\widetilde{D}_n = n^{1/2}\,\mathrm{diag}\left(v_{\dot{m}_1}(d_n), ..., v_{\dot{m}_q}(d_n)\right)$. It follows from Corollary 2 with $l_k \equiv 1$ and $g_k(\theta) = m(x_k, \theta)K(w_k)$ that, to prove Theorem 3, it suffices to show that

(i) for any function $\widetilde{K}(x)$ satisfying $E|\widetilde{K}(w_1)|^{1+\delta} < \infty$ for some $\delta > 0$ and for any continuous function $g(x)$,

$$\begin{aligned} \frac{1}{n}\sum_{k=1}^{n} g(x_{nk})\widetilde{K}(w_k) &= \frac{E\widetilde{K}(w_1)}{n}\sum_{k=1}^{n} g(x_{nk}) + o_P(1) \\ &\to_D E\widetilde{K}(w_1)\int_0^1 g(X_t)dt. \end{aligned} \tag{C.9}$$

(ii) $\hat{\theta}_n$ is a consistent estimator of θ_0, or equivalently [using Theorem 5 with $l_k = 1$ and $g_k(\theta) = m(x_k, \theta)K(w_k)$] for any $\theta_j \in \Theta$ and $\alpha_j \in R$, $j = 1, 2, ..., l$,

$$\begin{aligned} &\sum_{j=1}^{l} \alpha_j \frac{1}{n v_m^2(d_n)} \sum_{k=1}^{n} \left[m(x_k, \theta_j) - m(x_k, \theta_0)\right]^2 K^2(w_k) \\ &\qquad \to_D EK^2(w_1) \sum_{j=1}^{l} \alpha_j\, G(\theta_j), \end{aligned} \tag{C.10}$$

where $G(\theta) := \int_0^1 \left[\widetilde{h}_m(X_t, \theta) - \widetilde{h}_m(X_t, \theta_0)\right]^2 dt$.

(iii)

$$\left\{\widetilde{Z}_n(\theta_0),\ \widetilde{Y}_n(\theta_0)\right\} \to_D \left\{\int_0^1 \Phi(t)dU_t,\ EK^2(w_1)\int_0^1 \Phi(t)\Phi(t)'dt\right\}, \tag{C.11}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0426

where

$$\widetilde{Z}_n(\theta) = \widetilde{D}_n^{-1}\sum_{k=1}^n \dot{m}(x_k,\theta)\, K(w_k)\, u_k,$$
$$\widetilde{Y}_n(\theta) = \sum_{k=1}^n \big(\widetilde{D}_n^{-1}\big)'\, \dot{m}(x_k,\theta)\dot{m}(x_k,\theta)'\, \widetilde{D}_n^{-1}\, K^2(w_k).$$

Note that $w_k, k \geq 0$, is a stationary k_0-dependent random sequence. It is routine to show that, for any $m \to \infty$ satisfying $m/n \to 0$,

$$\max_{m\leq j\leq n-m} E\Big|\frac{1}{m}\sum_{k=j+1}^{j+m} \widetilde{K}(w_k) - E\widetilde{K}(w_1)\Big| \to 0.$$

Result (C.9) now follows from Lemma 5.1 of Hu et al. (2021) and standard result on the convergence to stochastic integrals.

The proof of (C.10) is similar to that of (B.3) of Wang (2021), by showing that, for any $\theta \in \Theta$,

$$\begin{aligned}&\frac{1}{nv_m^2(d_n)}\sum_{k=1}^n \big[m(x_k,\theta) - m(x_k,\theta_0)\big]^2 K^2(w_k)\\ =\ &\frac{1}{n}\sum_{k=1}^n \big[\widetilde{h}_m(x_{nk},\theta) - \widetilde{h}_m(x_{nk},\theta_0)\big]^2 K^2(w_k) + \widetilde{\Delta}_n, \end{aligned} \tag{C.12}$$

where $\widetilde{\Delta}_n \to_P 0$. Indeed, by recalling $\widetilde{h}_m(.,\theta)$ is continuous for each θ, (C.10) follows from (C.9), (C.12) and the Cramér–Wold theorem. We provide an outline proof of (C.12) for convenience of the reading. Write $x_k^* = x_k I(|x_k|/d_n \leq A)$,

$$\widetilde{R}_n(\theta) = \sum_{k=1}^n \big[m(x_k,\theta) - v_m(d_n)\widetilde{h}_m(x_k/d_n,\theta)\big]^2 K^2(w_k),$$
$$\widetilde{R}_n^*(\theta) = \sum_{k=1}^n \big[m(x_k^*,\theta) - v_m(d_n)h_m(x_k^*/d_n,\theta)\big]^2 K^2(w_k).$$

For any fixed $A > 0$, it follows from **A4***(ii) and (iii) and (C.9) that

$$\sup_{\theta\in\Theta} |R_n^*(\theta)| \leq o(1)\, v_m^2(d_n)\sum_{k=1}^n T^2(x_k^*/d_n)K^2(w_k) = o(1)nv_m^2(d_n)EK^2(w_1),$$

as $n \to \infty$. This implies that, for any $\epsilon > 0$,

$$\begin{aligned}
& P\big(\frac{1}{nv_m^2(d_n)} \sup_{\theta\in\Theta} |\widetilde{R}_n(\theta)| \ge \epsilon\big) \\
\le\ & P\big(x_k \ne x_k^*, \text{for some k=1,...,n}\big) + P\big(\frac{1}{nv_m^2(d_n)} \sup_{\theta\in\Theta} |\widetilde{R}_n^*(\theta)| \ge \epsilon\big) \\
\le\ & P\big(\max_{1\le k\le n} |x_k|/d_n \ge A\big) + P\big(\frac{1}{nv_m^2(d_n)} \sup_{\theta\in\Theta} |\widetilde{R}_n^*(\theta)| \ge \epsilon\big) \\
\to\ & 0,
\end{aligned}$$

as $n \to \infty$ first and then $A \to \infty$, namely, we have

$$\sup_{\theta\in\Theta} |\widetilde{R}_n(\theta)| = o_P\big[nv_m^2(d_n)\big]. \tag{C.13}$$

Now (C.12) follows (C.9) and (C.13) since

$$\begin{aligned}
& \widetilde{\Delta}_n \le 4 \sup_{\theta\in\Theta} |\widetilde{R}_n(\theta)| \\
& +4 \sup_{\theta\in\Theta} |\widetilde{R}_n(\theta)|^{1/2} \Big(\frac{1}{n}\sum_{k=1}^n \big[\widetilde{h}_m(x_{nk},\theta) - \widetilde{h}_m(x_{nk},\theta_0)\big]^2 K^2(w_k)\Big)^{1/2}.
\end{aligned}$$

We finally prove (C.11). We may write

$$\begin{aligned}
\widetilde{Z}_n(\theta_0) &= \frac{1}{\sqrt{n}} \sum_{k=1}^n \widetilde{h}(x_{nk},\theta_0)\, K(w_k)\, u_k + \widetilde{R}_n \\
&:= \widetilde{Z}_{n1}(\theta_0) + \widetilde{R}_n, \qquad\qquad (C.14)
\end{aligned}$$

where $\widetilde{R}_n = \sum_{k=1}^n \widetilde{a}_{nk} K(w_k) u_k$ with

$$\begin{aligned}
\widetilde{a}_{nk} &= \widetilde{D}_n^{-1} \dot{m}(x_k,\theta_0) - \frac{1}{\sqrt{n}}\, \widetilde{h}(x_{nk},\theta_0) \\
&= \Big[\frac{\dot{m}_1(.) - v_{\dot{m}_1}(d_n) h_{\dot{m}_1}(.)}{\sqrt{n} v_{\dot{m}_1}(d_n)}, \cdots, \frac{\dot{m}_q(.) - v_{\dot{m}_q}(d_n) h_{\dot{m}_q}(.)}{\sqrt{n} v_{\dot{m}_q}(d_n)}\Big]'.
\end{aligned}$$

As in the proof of (C.13), we have

$$\sum_{k=1}^n ||\widetilde{a}_{nk}||^2 K^2(w_k) = o_P(1) \quad \text{and} \quad \max_{1\le k\le n} \big[||\widetilde{a}_{nk}||^2 K^2(w_k)\big] = o_P(1)$$

by (C.9). This yields $||\widetilde{R}_n|| = o_P(1)$ since $\{\widetilde{a}_{nk}u_k, \mathcal{F}_k\}$ forms a martingale difference array with $\sup_{k\geq 1} E\big(|u_k|^{2+\delta} \mid \mathcal{F}_{k-1}\big) \leq C < \infty$ for some $\delta > 0$ under the conditions **A1 - A3**. Similarly, the same arguments used in the proofs of (C.9) and (C.13) yield that

$$\begin{aligned}
\widetilde{Y}_n(\theta_0) &= \sum_{k=1}^n (\widetilde{D}_n^{-1})'\, \dot{m}(x_k,\theta)\dot{m}(x_k,\theta)'\, \widetilde{D}_n^{-1}\, K^2(w_k) \\
&= \frac{1}{n}\sum_{k=1}^n \widetilde{h}(x_{nk},\theta_0)\widetilde{h}(x_{nk},\theta_0)'\, K^2(w_k) + \widetilde{R}_{1n} \\
&= \frac{EK^2(w_1)}{n}\sum_{k=1}^n \widetilde{h}(x_{nk},\theta_0)\widetilde{h}(x_{nk},\theta_0)' + o_P(1) \\
&= \widetilde{Y}_{n1}(\theta_0) + o_P(1). \qquad \text{(C.15)}
\end{aligned}$$

By virtue of (C.14) and (C.15), result (C.11) will follow if we prove: for any $\alpha_i' = (\alpha_{i1}, ..., \alpha_{iq}) \in \mathbb{R}$, $i = 1, 2, 3$,

$$\begin{aligned}
&\big\{\alpha_1'\, Y_{n1}(\theta_0)\alpha_2,\ \alpha_3' Z_{n1}(\theta_0)\big\} \\
\to_D\ &\big\{\alpha_1' \int_0^1 \Phi(t)\Phi(t)' dt \alpha_2 EK^2(w_1),\ \alpha_3' \int_0^1 \Phi(t) dU_t\big\}. \qquad \text{(C.16)}
\end{aligned}$$

Recall $\Psi(t) = \widetilde{h}(X_t, \theta_0)$. The proof of (C.16) is standard, which follows from the classical result on the convergence to stochastic integrals [e.g., Kurtz and Protter (1991)] and the fact: instead of (2.4), we have

$$\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor} K(w_i)u_i,\ \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\epsilon_i,\ \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\epsilon_{-i},\ \frac{1}{d_n}x_{[nt]}\Big) \quad\Rightarrow\quad \big(U_t,\ B_t,\ B_{-t},\ X_t\big),$$

on $D_{\mathbb{R}^4}[0,\infty)$, where B_t, B_{-t} and X_t are given as in Section 2.2 and $(U_t, B_t)_{t\geq 0}$ is a bivariate Brownian motion with covariance matrix:

$$\Omega = \begin{pmatrix} \widetilde{\sigma}^2 & \rho \\ \rho & 1 \end{pmatrix}, \quad \text{where } \widetilde{\sigma}^2 = E\big[u_1 K(w_1)\big]^2 \text{ and } \rho = E\big[\epsilon_1 u_1 K(w_1)\big].$$

The proof of Theorem 3 is complete. □