Journal of Econometrics 114 (2003) 73 – 106
JOURNAL OF Econometrics
www.elsevier.com/locate/econbase

Index models with integrated time series

Yoosoon Chang[a,*], Joon Y. Park[a,b]

[a] *Department of Economics, Rice University, 6100 Main Street-MS 22, Houston, TX 77005-1892, USA*
[b] *School of Economics, Seoul National University, Seoul, 151-742, South Korea*

Received 8 December 1998; received in revised form 9 July 2002; accepted 15 September 2002

Abstract

This paper considers index models, such as simple neural network models and smooth transition regressions, with integrated regressors. The models can be used to analyze various nonlinear relationships among nonstationary economic time series. Asymptotics for the nonlinear least squares (NLS) estimator in such models are fully developed. The estimator is shown to be consistent with a convergence rate that is a mixture of $n^{3/4}$, $n^{1/2}$ and $n^{1/4}$ for simple neural network models, and of $n^{5/4}$, n, $n^{3/4}$ and $n^{1/2}$ for smooth transition regressions. Its limiting distribution is also obtained. Some of its components are mixed normal, with mixing variates depending upon Brownian local time as well as Brownian motion. However, it also has nonGaussian components. It is in particular shown that applications of usual statistical methods in such models generally yield inefficient estimates and/or invalid tests. We develop a new methodology to efficiently estimate and to correctly test in those models. A simple simulation is conducted to investigate the finite sample properties of the (NLS) estimators and the newly proposed efficient estimators.

JEL classification: C22; C45

Keywords: Index model; Integrated time series; Neural network model; Smooth transition regression; Brownian motion and Brownian local time

1. Introduction

Nonlinear models seem to become increasingly popular in econometrics. A wide range of econometric models have been fitted using nonlinear regressions. This is true for both cross section and time series data. The statistical theory of the nonlinear regression model is now well established for the fixed and/or weakly dependent regressors.

* Corresponding author. Tel.: +713-348-2796; fax: +713-348-5278.
E-mail addresses: yoosoon@rice.edu (Y. Chang), jpark@plaza.snu.ac.kr (J.Y. Park).

0304-4076/03/$ - see front matter
PII: S0304-4076(02)00220-8

See Jennrich (1969) and Wu (1981) for its early developments, and Wooldridge (1994) and Andrews and McDermott (1995) for some important later extensions. Moreover, Park and Phillips (2001) and Chang et al. (2001) have recently developed the general theory of nonlinear regressions with integrated time series. They consider nonlinear regressions with separably additive regression function. That is, the regression function is allowed to be nonlinear, but they assume that it can be written as a sum of nonlinear functions each of which includes only a single regressor. For such models, they derive the asymptotic distributions of the nonlinear least squares (NLS) estimators as functionals of Brownian motions and Brownian local time.

We consider in the paper nonlinear index models driven by integrated time series. Our models include as special cases the simple neural network models and the smooth transition regressions. These are two classes of index models, which seem to have most interesting potential applications. The neural network models, which are inspired by features of the way information is processed in the brain, have been widely used in practical applications, since they were advocated by White (1989). The smooth transition regressions are appropriate to model an economic relationship changing from one state to another with a smooth transition function. For its motivation and history, the reader is referred to Granger and Teräsvirta (1993). In our context, they actually represent a longrun cointegrating relationship departing from a longrun equilibrium and smoothly adjusting to a new equilibrium.

In the nonstationary nonlinear index models we consider here, the regression function is in particular allowed to include more than one explanatory variables. For the regressions with integrated time series, the statistical theory of the index type models is vastly different from that of separably additive models. This is because the behavior of a functional of univariate Brownian motion is drastically different from that of a vector Brownian motion. For the index models with integrated time series, we show that the NLS estimators are consistent with convergence rates ranging from $n^{1/4}$ to $n^{3/4}$ for the simple neural network models, and from $n^{1/2}$ to $n^{5/4}$ for the smooth transition regressions. We also derive the limiting distributions of the NLS estimators, and present them as functionals of Brownian motions and Brownian local time.

The usual NLS estimators for such nonstationary index models are generally not efficient in the sense of Phillips (1991) and Saikkonen (1991), just as the usual OLS estimators are not efficient for the linear cointegrating regressions. This is because the usual NLS estimators do not use the information on the presence of the unit roots in the explanatory variables. Moreover, their limiting distributions are nonnormal and dependent upon nuisance parameters, which invalidates the standard chi-square tests. We show in the paper that the methodology developed by Chang et al. (2001) can also be applied to the nonstationary index models. We modify the usual NLS estimators using the correction terms that are in motivation the same as those of Phillips and Hansen (1990) and Park (1992), so that the resulting estimators become efficient and provide standard chi-square tests.

The rest of the paper is organized as follows. In Section 2 we introduce the model, assumptions and preliminary results. The model is presented in a general form, and assumptions are introduced. Also, preliminary lemmas, on which all the subsequent theories heavily rely, are presented. The statistical theory of the model is developed in

Section 3. In particular, the asymptotic theories are fully developed for two classes of models—the simple neural network models and smooth transition regressions. The efficient estimation of and hypothesis testing on the models are considered subsequently in Section 4. To investigate the finite sample behavior of the estimators and test statistics, we perform a simple simulation and report its results in Section 5. Section 6 concludes the paper. Mathematical proofs are collected in Section 7.

2. The model, assumptions and preliminary results

We consider nonlinear regressions of the form

$$y_t = F(x_t, \theta_0) + u_t \tag{1}$$

with the regression function F further modeled as

$$F(x, \theta) = \mu + p(x, \alpha) + q(x, \alpha)G(\nu + x'\beta), \tag{2}$$

where (x_t) is an m-dimensional integrated process of order one, $\theta = (\mu, \alpha', \nu, \beta')'$ is a vector of parameters with the true value denoted by $\theta_0 = (\mu_0, \alpha_0', \nu_0, \beta_0')'$, and (u_t) the stationary error.[1] We assume that $p(\cdot, \alpha)$ and $q(\cdot, \alpha)$ are linear functionals defined on $\mathbf{R}^m$. The nonlinear part of the regression function F is specified as an index model with G, which will be assumed to be a smooth distribution function-like transformation on $\mathbf{R}$.[2]

We now introduce precise assumptions on the data generating processes. As mentioned above, (x_t) is assumed to be an integrated process of order one. More explicitly, we let $v_t = \Delta x_t$ and specify (v_t) as a general linear process given by

$$v_t = \Phi(L)\varepsilon_t = \sum_{k=0}^{\infty} \Phi_k \varepsilon_{t-k}. \tag{3}$$

Moreover, we let $w_t = (u_t, \varepsilon_{t+1}')'$ and define a filtration $(\mathscr{F}_t)_{t \geqslant 0}$ by $\mathscr{F}_t = \sigma((w_s)_{-\infty}^t)$, i.e., the σ-field generated by (w_s) for all $s \leqslant t$. Throughout the paper, the Euclidean norm of a vector will be denoted by $\|\cdot\|$.

Assumption 1. We assume

(a) $(w_t, \mathscr{F}_t)$ is a martingale difference sequence,
(b) $\mathbf{E}(w_t w_t' | \mathscr{F}_{t-1}) = \Sigma > 0$, and
(c) $\sup_{t \geqslant 1} \mathbf{E}(\|w_t\|^r | \mathscr{F}_{t-1}) < \infty$ for some $r > 2$.

---

[1] We may allow for the presence of weakly dependent covariates in our model, though it is not explicitly considered for expositional simplicity. In particular, if they are included linearly as additional regressors and orthogonal to regression errors, their presence would not affect our subsequent asymptotics. This can be shown as in Chang et al. (2001).

[2] Our model here does not allow for (y_t) to be a binary response. The binary choice model with integrated explanatory variables, though it has the regression function which can be regarded as a special case of F in (2), has persistent conditional heterogeneity, and consequently its asymptotics are quite different from those developed in the paper. See Park and Phillips (2000) for details.

The condition in (a) implies, in particular, that (x_t) is predetermined and that $\mathbf{E}(u_t|\mathscr{F}_{t-1})=0$. We therefore have $\mathbf{E}(y_t|\mathscr{F}_{t-1})=F(x_t,\theta_0)$, as is often the case also for the usual nonlinear regression.[3] Note that the regressor (x_t) can be generated by a general serially correlated linear process (v_t), though we require that the regression error (u_t) be devoid of temporal dependence. The moment conditions in (b) and (c), however, do not allow for the presence of conditional heterogeneity in both (u_t) and (v_t).[4] We decompose Σ introduced in (b) conformably with the partition of (w_t), and denote the entries by σ_u^2, $\sigma_{u\varepsilon}$, $\sigma_{\varepsilon u}$ and $\Sigma_{\varepsilon\varepsilon}$.

Assumption 2. We assume

(a) $\Phi(1)$ is nonsingular, and $\sum_{k=0}^\infty k\|\Phi_k\|<\infty$, and
(b) (ε_t) are iid with $\mathbf{E}\|\varepsilon_t\|^r<\infty$ for some $r>8$, and the distribution of (ε_t) is absolutely continuous and has characteristic function φ such that $\varphi(t)=\mathrm{o}(\|t\|^{-\delta})$ as $\|t\|\to\infty$ for some $\delta>0$.

The condition on $\Phi(1)$ in (a) ensures that the spectrum of (v_t) at the origin is nonsingular. This, in turn, implies that (x_t) is an integrated process of full rank, i.e., there is no cointegrating relationship among the component time series in (x_t).[5] The summability condition on (Φ_k) in (a) is commonly imposed for linear processes. The condition in (b) is somewhat strong, and in fact not necessary for some of our subsequent results. However, it is still satisfied by a wide class of data generating processes including all invertible Gaussian ARMA models.

For (u_t) and (v_t), we define stochastic processes

$$U_n(r)=\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}u_t \quad\text{and}\quad V_n(r)=\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}v_t \tag{4}$$

on $[0,1]$, where $[s]$ denotes the largest integer not exceeding s. The process (U_n,V_n) takes values in $D[0,1]^{1+m}$, where $D[0,1]$ is the space of cadlag functions on $[0,1]$. Under Assumptions 1 and 2, an invariance principle holds for (U_n,V_n). That is, we have as $n\to\infty$

$$(U_n,V_n)\to_d (U,V), \tag{5}$$

where (U,V) is $(1+m)$-dimensional vector Brownian motion. It is shown, for instance, by Phillips and Solo (1992).

For the function G in (2) used to model the nonlinear component of the regression (1), we use the notation $\dot G,\ddot G$ and $\dddot G$ respectively to denote its first, second and third derivative, and let $\dot G_i(x),\ddot G_i(x),\dddot G_i(x)=x^i\dot G(x),x^i\ddot G(x),x^i\dddot G(x)$.

[3] For nonlinear regression to work well in the weakly dependent case, we only need $\mathbf{E}(y_t|x_t)=F(x_t,\theta_0)$.

[4] Our subsequent results on the estimates of μ and α hold under much weaker conditions, which allow for cross correlations in (u_t) and (v_t) as well as temporal dependencies and conditional/unconditional heterogeneities in (u_t).

[5] This also implies that the presence of stationary or weakly dependent variables in (x_t) is not allowed.

Assumption 3. We assume

(a) G is bounded with $\lim_{x\to-\infty} G(x)=0$ and $\lim_{x\to\infty} G(x)=1$, and
(b) $\dot{G}, \ddot{G}$ and $\dddot{G}$ exist, and $\dot{G}_i, \ddot{G}_i$ and $\dddot{G}_i$ are bounded and integrable for $0 \leqslant i \leqslant 3$.

We consider G primarily as a function that behaves like a distribution function of a continuous type random variable. The standard normal distribution function $G(x) = (2\pi)^{-1/2}\int_{-\infty}^{x} \mathrm{e}^{-y^2/2}\,\mathrm{d}y$ or the logistic function $G(x)=e^x/(1+e^x)$ are good examples. The function G in our model, however, is not restricted to such a function. One may easily see that any smooth bounded function with well defined asymptotes can be normalized so that it satisfies conditions in Assumption 3.

To develop the limit theory for the model given by (1) and (2), we first rotate the integrated regressor x_t and the associated parameter β using an $(m\times m)$-orthogonal matrix $H=(h_1, H_2)$ with $h_1=\beta_0/\|\beta_0\|$. The components h_1 and H_2 of H are of ranks 1 and $(m-1)$, respectively. More explicitly, we have

$$H'x_t = \begin{pmatrix} h_1'x_t \\ H_2'x_t \end{pmatrix} = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} \quad \text{and} \quad H'\beta = \begin{pmatrix} h_1'\beta \\ H_2'\beta \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \tag{6}$$

where (x_{1t}) and β_1 are scalars, and (x_{2t}) and β_2 are $(m-1)$-dimensional vectors. We accordingly define the limit BMs of (x_{1t}) and (x_{2t}) as

$$V_1 = h_1'V \quad \text{and} \quad V_2 = H_2'V$$

that are of dimensions 1 and $(m-1)$, respectively. We denote respectively by ω_1^2 and Ω_{22} the variances of the Brownian motions V_1 and V_2. Their covariance is denoted by ω_{12} or ω_{21}.

Our subsequent theory relies heavily on the local time of V_1, which we denote by $L_{V_1}(t,s)$, where t and s are respectively time and spatial parameters. We also define the scaled local time of V_1 as

$$L_1(t,s) = (1/\omega_1^2) L_{V_1}(t,s).$$

We will call L_1, instead of L_{V_1}, the local time of V_1 throughout the paper. As will become evident as we move along, the local time L_1 plays an important role in our theory. The reader is referred to Park and Phillips (1999, 2001) for more discussions on the role of Brownian local time on the asymptotic theories of nonlinear models with integrated time series. Our representations of the limiting distributions also involve another vector Brownian motion, denoted by W, which is independent of U and V, and has variance $\sigma_u^2 I$.

We now present lemmas that are important in establishing the asymptotic theories of our model. For $x\in\mathbf{R}^{m-1}$ and $i=0,\ldots,\kappa$, we define x^i to be the i-fold tensor product of x, i.e., $x^i = x\otimes\cdots\otimes x$. By convention, we let $x^0=1$. Also, we let $f_i:\mathbf{R}\to\mathbf{R}$ for

$i = 0, \ldots, \kappa$ and define $K : \mathbf{R}^m \to \mathbf{R}^{m_\kappa}, m_\kappa = 1 + (m-1) + \cdots + (m-1)^\kappa$, by

$$K(x_1, x_2) = \begin{pmatrix} f_0(x_1) \\ f_1(x_1)x_2 \\ \vdots \\ f_\kappa(x_1)x_2^\kappa \end{pmatrix} \tag{7}$$

for $(x_1, x_2) \in \mathbf{R} \times \mathbf{R}^{m-1}$. For the asymptotics of nonstationary index models, we need to analyze the asymptotic behaviors of $\sum_{t=1}^n K(x_{1t}, x_{2t})$ and $\sum_{t=1}^n K(x_{1t}, x_{2t})\, u_t$, which we call the first and second asymptotics of K.

Lemma 1. *Let Assumptions* 1 *and* 2 *hold. If* K *is defined as in* (7) *with* f_i*'s that are bounded, integrable and differentiable with bounded derivatives, then we have*

$$n^{-1/2} \Lambda_n^{-1} \sum_{t=1}^n K(x_{1t}, x_{2t}) \to_d \int_{-\infty}^{\infty} \mathrm{d}s \int_0^1 \mathrm{d}L_1(r,0) K(s, V_2(r))$$

$$n^{-1/4} \Lambda_n^{-1} \sum_{t=1}^n K(x_{1t}, x_{2t}) u_t \to_d \left(\int_{-\infty}^{\infty} \mathrm{d}s \int_0^1 \mathrm{d}L_1(r,0) K(s, V_2(r)) K(s, V_2(r))' \right)^{1/2} \times W(1)$$

where $\Lambda_n = \mathrm{diag}(1, n^{1/2} I_{m-1}, \ldots, n^{\kappa/2} I_{\kappa(m-1)})$.

Lemma 1 gives the asymptotic behavior of K consisting of smooth and bounded f_i's. The asymptotics of K are represented by a Riemann–Stieltjes integral of $K(s, V_2(r))$ with respect to the Lebesgue measure ds and the measure d$L_1(r,0)$ given by the local time L_1 of V_1 at the origin, respectively for s and r. The limiting distribution for the first asymptotics is nonstandard and nonnormal. However, the second asymptotics yield limiting distribution that is mixed normal, with a mixing variate dependent not only on the sample path but also on the local time of the limit Brownian motions.

To investigate the parameter dependency of the limiting distributions in Lemma 1, we may let

$$V_1 = \omega_1 V_1^\circ \quad \text{and} \quad V_2 = \frac{\omega_{21}}{\omega_1} V_1^\circ + \left(\Omega_{22} - \frac{\omega_{21}\omega_{12}}{\omega_1^2} \right)^{1/2} V_2^\circ,$$

where V_1° and V_2° are two independent standard Brownian motions. If we let L_1° be the local time of V_1°, then it follows that

$$L_1(t,s) = \omega_1 L_1^\circ \left(t, \frac{s}{\omega_1} \right).$$

Furthermore, we have due to a well known property of the local time

$$\int_0^1 V_1^\circ(r)\, \mathrm{d}L_1^\circ(r,0) = 0 \quad \text{a.s.}$$

We may therefore represent the first and second asymptotics in Lemma 1 as

$$\int_{-\infty}^{\infty} \mathrm{d}s \int_0^1 \omega_1 \, \mathrm{d}L_1^\circ(r,0) K(s, \Omega_{22\cdot 1}^{1/2} V_2^\circ(r))$$

$$\left(\sigma_u^2 \int_{-\infty}^{\infty} \mathrm{d}s \int_0^1 \omega_1 \mathrm{d}L_1^\circ(r,0) K(s, \Omega_{22\cdot 1}^{1/2} V_2^\circ(r)) K(s, \Omega_{22\cdot 1}^{1/2} V_2^\circ(r))' \right)^{1/2} W^\circ(1),$$

where $\Omega_{22\cdot 1} = \Omega_{22} - \omega_{21}\omega_{12}/\omega_1^2$, i.e., the conditional variance of V_2 given V_1, and W° is defined by $W = \sigma_u W^\circ$ conformably as V_1° and V_2°.

Lemma 2. *Let Assumptions* 1 *and* 2 *hold. If K is defined as in* (7) *with f_i's that are bounded and have asymptotes a_i and b_i as $x \to \mp\infty$, then we have*

$$n^{-1} \Lambda_n^{-1} \sum_{t=1}^{n} K(x_{1t}, x_{2t}) \to_d \int_0^1 K^\circ(V_1(r), V_2(r)) \, \mathrm{d}r,$$

$$n^{-1/2} \Lambda_n^{-1} \sum_{t=1}^{n} K(x_{1t}, x_{2t}) u_t \to_d \int_0^1 K^\circ(V_1(r), V_2(r)) \, \mathrm{d}U(r),$$

where Λ_n is given in Lemma 1 *and K° is defined similarly as K with f_i replaced by f_i°, $f_i^\circ(x) = a_i 1\{x < 0\} + b_i 1\{x \geqslant 0\}$, for $i = 0, \ldots, \kappa$.*

The asymptotics for K with f_i's which have nonzero asymptotes are quite different. Their stochastic orders are bigger than those for K with f_i's vanishing at infinity, which we have seen in Lemma 1. This may well be expected, since integrated time series (x_t) has a growing stochastic trend and thus the orders of its nonlinear transformations are determined by the asymptotes of the transformation functions. The first asymptotics is characterized by a path by path Riemann integral of the limit Brownian motions. The second asymptotics is, however, represented by a stochastic integral. Unlike the corresponding asymptotics for K with vanishing f_i's, the second asymptotics for K does not yield Gaussian limiting distribution. It is nonnormal and biased. It reduces to a mixed normal distribution, only when U is independent of V_1 and V_2. This, however, seems rarely to be the case in practical applications. Notice that the asymptotics for K depend on f_i's only through their asymptotes.

3. Statistical theory

The nonlinear regression (1) can be estimated by NLS. If we let

$$Q_n(\theta) = \frac{1}{2} \sum_{t=1}^{n} (y_t - F(x_t, \theta))^2$$

then the NLS estimator $\hat{\theta}_n$ of θ in (1) is given by

$$\hat{\theta}_n = \operatorname*{argmin}_{\theta \in \Theta} Q_n(\theta), \tag{8}$$

where Θ is the parameter set, which is assumed to be a compact and convex subset of $\mathbf{R}^p$. We let θ_0 be an interior point of Θ. An error variance estimate is given by $\hat{\sigma}_n^2 = (1/n)\sum_{t=1}^n \hat{u}_t^2$, where $\hat{u}_t = y_t - F(x_t, \hat{\theta}_n)$.

Define $\dot{Q}_n = \partial Q_n/\partial\theta$ and $\ddot{Q}_n = \partial^2 Q_n/\partial\theta\partial\theta'$. Then we have

$$\dot{Q}_n(\theta) = -\sum_{t=1}^n \dot{F}(x_t,\theta)(y_t - F(x_t,\theta)),$$

$$\ddot{Q}_n(\theta) = \sum_{t=1}^n \dot{F}(x_t,\theta)\dot{F}(x_t,\theta)' - \sum_{t=1}^n \ddot{F}(x_t,\theta)(y_t - F(x_t,\theta)),$$

where $\dot{F} = \partial F/\partial\theta$ and $\ddot{F} = \partial^2 F/\partial\theta\partial\theta'$. Furthermore, we have from the usual first order Taylor expansion that

$$\dot{Q}_n(\hat{\theta}_n) = \dot{Q}_n(\theta_0) + \ddot{Q}_n(\theta_n)(\hat{\theta}_n - \theta_0), \tag{9}$$

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

The limiting distribution of $\hat{\theta}_n$ can be derived from (9) as in the standard nonlinear regression. For our model given by (1) and (2), we may apply Lemmas 1 and 2 to deduce

$$C_n^{-1}J'\ddot{Q}_n(\theta_0)JC_n^{-1} \to_d A > 0 \quad \text{a.s.} \quad \text{and} \quad -C_n^{-1}J'\dot{Q}_n(\theta_0) \to_d B \tag{10}$$

for an appropriately chosen normalizing sequence (C_n) of symmetric matrices and an orthogonal matrix J. Therefore, we may expect under a suitable set of conditions that

$$C_nJ'(\hat{\theta}_n - \theta_0) = -(C_n^{-1}J'\ddot{Q}_n(\theta_0)JC_n^{-1})^{-1}C_n^{-1}J'\dot{Q}_n(\theta_0) + o_p(1) \to_d A^{-1}B. \tag{11}$$

If we let $C_{n\delta} = n^{-\delta}C_n$ for $\delta > 0$, and define $\Theta_n \subset \Theta$ by

$$\Theta_n = \{\theta : \|C_{n\delta}(\theta - \theta_0)\| \leqslant 1\} \tag{12}$$

then it can be shown for our model given by (1) and (2) that

$$\|C_{n\delta}^{-1}J'(\ddot{Q}_n(\theta) - \ddot{Q}_n(\theta_0))JC_{n\delta}^{-1}\| \to_p 0 \tag{13}$$

uniformly for all $\theta \in \Theta_n$. Given (10), the existence of such $C_{n\delta}$ as in (13) is sufficient to ensure the asymptotics in (11). This is shown in Wooldridge (1994), and used in Park and Phillips (2001) to derive the asymptotics for nonlinear regressions with integrated time series.

Below, we consider two special nonlinear index models, simple neural network models and smooth transition regressions. This is to develop the relevant asymptotics more explicitly. All other models that are specified as (1) and (2) can be analyzed similarly. In what follows, we let

$$H'\hat{\beta}_n = \begin{pmatrix} h_1'\hat{\beta}_n \\ H_2'\hat{\beta}_n \end{pmatrix} = \begin{pmatrix} \hat{\beta}_{1n} \\ \hat{\beta}_{2n} \end{pmatrix} \quad \text{and} \quad H'\beta_0 = \begin{pmatrix} h_1'\beta_0 \\ H_2'\beta_0 \end{pmatrix} = \begin{pmatrix} \|\beta_0\| \\ 0 \end{pmatrix}$$

correspondingly as β_1 and β_2 defined in (6). Also, we define $\dot{G}_0(s) = \dot{G}(v_0 + \|\beta_0\|s)$.

3.1. Simple neural network models

When the nonlinear function F defined in (2) is specified with $\theta=(\mu,\alpha,v,\beta')'$, $p(x,\alpha)\equiv 0$ and $q(x,\alpha)=\alpha$, the model (1) becomes

$$y_t=\mu+\alpha G(v+x_t'\beta)+u_t. \tag{14}$$

It is the prototypical one hidden layer neural network model. The model is motivated by the way that information is believed to be processed in the brain. The following theorem characterizes the asymptotic behaviors of the NLS estimators $\hat{\mu}_n$, $\hat{\alpha}_n$, $\hat{v}_n$ and $\hat{\beta}_n$ of the parameters in the simple neural network model (SNNM) (14). We assume that $\alpha_0\neq 0$, which is necessary for the identifiability of β_0.

Theorem 3. *Let Assumptions* 1–3 *hold, and suppose that the model is given by* (14). *Then we have as* $n\to\infty$

$$\begin{pmatrix} n^{1/2}(\hat{\mu}_n-\mu_0) \\ n^{1/2}(\hat{\alpha}_n-\alpha_0)\end{pmatrix}\to_d\left(\int_0^1 \mathrm{N}(r)\mathrm{N}(r)'\mathrm{d}r\right)^{-1}\int_0^1 \mathrm{N}(r)\,\mathrm{d}U(r),$$

where $\mathrm{N}(r)=(1,1\{V_1(r)\geqslant 0\})'$, *and*

$$\begin{pmatrix} n^{1/4}(\hat{v}_n-v_0) \\ D_nH'(\hat{\beta}_n-\beta_0)\end{pmatrix}\to_d\left(\int_{-\infty}^{\infty}\mathrm{d}s\int_0^1 \mathrm{d}L_1(r,0)M(r,s)M(r,s)'\right)^{-1/2}W(1)$$

where $D_n=\mathrm{diag}(n^{1/4},n^{3/4}I_{m-1})$ *and* $M(r,s)=\alpha_0(\dot{G}_0(s),s\dot{G}_0(s),\dot{G}_0(s)V_2(r)')'$.

All the parameters are estimated consistently in the SNNM (14).[6] Their convergence rates are, however, different. The estimators $\hat{\mu}_n$ and $\hat{\alpha}_n$ for the intercept μ_0 and the coefficient of the index function α_0 converge at the rate $\sqrt{n}$, as in the standard regression model. These are the parameters which determine the asymptotes of the conditional mean of (y_t), i.e., μ_0 and $\mu_0+\alpha_0$ give the lower and upper conditional mean values. The estimators $\hat{v}_n$ and $\hat{\beta}_n$ of the parameters v_0 and β_0 inside the nonlinear function G have convergence rates that are a mixture of $n^{1/4}$ and $n^{3/4}$. Along the hyperplane orthogonal to β_0, $\hat{\beta}_n$ has convergence rate $n^{3/4}$, which is an order of magnitude faster than the other component of $\hat{\beta}_n$ and $\hat{v}_n$.

Theorem 3 shows in particular that (11) holds with

$$C_n=\mathrm{diag}(n^{1/2},n^{1/2},n^{1/4},D_n)\quad\text{and}\quad J=\mathrm{diag}(1,1,1,H) \tag{15}$$

for the SNNM (14). The limiting distributions of $\hat{v}_n$ and $\hat{\beta}_n$ are mixed normal with zero mean. However, $\hat{\mu}_n$ and $\hat{\alpha}_n$ have asymptotic distributions that are biased and nonnormal, unless (x_t) are strictly exogenous. They are biased, due to the presence of correlation between U and V_1. The distributions reduce to normal with mean zero, only when U and V_1 are independent. The two sets of parameters $(\hat{\mu}_n,\hat{\alpha}_n)$ and $(\hat{v}_n,\hat{\beta}_n)$ are asymptotically independent, since W is independent of both U and V.

[6] Note that $\int_{-\infty}^{\infty}\mathrm{d}s\int_0^1\mathrm{d}L_1(r,0)M(r,s)M(r,s)'$ is nonsingular a.s. so long as $\dot{G}$ does not vanish almost everywhere, and this is guaranteed by the conditions in part (a) of Assumption 3.

The results in Theorem 3 imply in particular that the parameters (μ, α) and (v, β) are *separable*. That is, for the estimation of one set of parameters, we may regard the other as being fixed and known. For the estimation of μ and α, we may assume that v and β are known to be v_0 and β_0, and look at the regression

$$y_t = \mu + \alpha G(v_0 + x_t'\beta_0) + u_t$$

and the asymptotic distribution of $\hat{\mu}_n$ and $\hat{\alpha}_n$ are the same as the usual OLS estimators from this regression. Likewise, we may fix μ and α at μ_0 and α_0 for the estimation of v and β and look at the nonlinear regression

$$y_t - \mu_0 = \alpha_0 G(v + x_t'\beta) + u_t$$

with unknown parameters v and β.

3.2. Smooth transition regressions

The model (1) becomes the so-called smooth transition regression (STR) when the function $F(x, \theta)$ in (2) is defined with $p(x, \alpha) = x'\alpha_1$ and $q(x, \alpha) = x'(\alpha_2 - \alpha_1)$, where $\alpha = (\alpha_1', \alpha_2')'$. The resulting regression is written as

$$\begin{aligned} y_t &= \mu + x_t'\alpha_1 + x_t'(\alpha_2 - \alpha_1)G(v + x_t'\beta) + u_t \\ &= \mu + x_t'\alpha_1(1 - G(v + x_t'\beta)) + x_t'\alpha_2 G(v + x_t'\beta) + u_t \end{aligned} \tag{16}$$

and the parameter θ is defined by $\theta = (\mu, \alpha_1', \alpha_2', v, \beta')'$. The STR allows us to model an economic relationship which evolves slowly over time, from one state to the other. The coefficient of the regressor (x_t) is assumed to change from α_1 to α_2 in (16). The transition is specified in such a way that it is also affected by (x_t). We may however let the underlying regressions have one set of variables as explanatory variables, while assuming that the transition is governed by another set of variables. This can be done simply by setting some of the coefficients in α and β to be zero.

Recall that we assume (x_t) is an integrated time series. The regression in (16) therefore models a cointegrating relationship. The above STR describes a longrun relationship that has been changing slowly and smoothly. We may think of two regression coefficients as representing two different equilibrium states. Therefore, the STR in (16) describes an economy moving slowly from one equilibrium to the other. The following theorem presents the limit theory for the NLS estimators $\hat{\mu}_n, \hat{\alpha}_{1n}, \hat{\alpha}_{2n}, \hat{v}_n$ and $\hat{\beta}_n$. We assume that $\alpha_{10} \neq \alpha_{20}$.

Theorem 4. *Let Assumptions* 1–3 *hold, and suppose that the model is given by* (16). *Then we have as* $n \to \infty$

$$\begin{pmatrix} \sqrt{n}(\hat{\mu}_n - \mu_0) \\ n(\hat{\alpha}_{1n} - \alpha_{10}) \\ n(\hat{\alpha}_{2n} - \alpha_{20}) \end{pmatrix} \to_d \left(\int_0^1 \mathrm{N}(r)\mathrm{N}(r)'\,\mathrm{d}r \right)^{-1} \int_0^1 \mathrm{N}(r)\,\mathrm{d}U(r), \tag{17}$$

where $\mathrm{N}(r) = (1, 1\{V_1(r) < 0\}V(r)', 1\{V_1(r) \geqslant 0\}V(r)')'$, *and*

$$\begin{pmatrix} n^{3/4}(\hat{v}_n - v_0) \\ D_n H'(\hat{\beta}_n - \beta_0) \end{pmatrix} \to_d \left(\int_{-\infty}^{\infty} \mathrm{d}s \int_0^1 \mathrm{d}L_1(r,0) M(r,s) M(r,s)' \right)^{-1/2} W(1), \quad (18)$$

where $D_n = \mathrm{diag}(n^{3/4}, n^{5/4} I_{m-1})$ *and*

$$M(r,s) = (\dot{G}_0(s) c' V_2(r), s\dot{G}_0(s)\, c' V_2(r), \dot{G}_0(s) c' V_2(r) V_2(r)')'$$

with $c = H_2'(\alpha_{20} - \alpha_{10})$.

Again, all the parameters are estimated consistently by NLS.[7] Also, the convergence rates vary across different parameters. The estimators $\hat{\mu}_n$, $\hat{\alpha}_{1n}$ and $\hat{\alpha}_{2n}$ converge at the same rates as in the usual linear cointegrating regressions. The convergence rates for $\hat{v}_n$ and $\hat{\beta}_n$ are $\sqrt{n}$-order faster than their counterparts in the SNNM. The limiting distributions of $\hat{\alpha}_{1n}$, $\hat{\alpha}_{2n}$ and $\hat{\mu}_n$ do not depend upon G. This implies, in particular, that the estimators may well be consistent even if our specification on G is incorrect. Indeed, we may show that they have the same limiting distribution regardless of possible misspecification of G, as long as it is a smooth distribution function-like transformation on $\mathbf{R}$.[8] It also makes it clear that we may test on the parameters α_1, α_2 and μ without actually knowing precise functional form of G.

We may easily see from Theorem 4 that (11) holds for the STR in (16) with

$$C_n = \mathrm{diag}(n^{1/2}, nI_m, nI_m, n^{3/4}, D_n) \quad \text{and} \quad J = \mathrm{diag}(1, H, H, 1, H). \quad (19)$$

The limiting distributions of the NLS estimators in the STR are given similarly as those for the corresponding parameters in the simple neural network models. The distributions for $\hat{v}_n$ and $\hat{\beta}_n$ are mixed normal, but $\hat{\mu}_n$, $\hat{\alpha}_{1n}$ and $\hat{\alpha}_{2n}$ have distributions which are generally biased and nonnormal. The latter become mixed normal only if the limiting Brownian motions U and V are independent each other.

Just as in the asymptotics for the SNNM, we have separability for two sets of parameters $(\mu, \alpha_1, \alpha_2)$ and (v, β). For the estimation of the parameters μ, α_1 and α_2, we may set the values of the parameters v and β to v_0 and β_0, respectively. Therefore, we can just look at the model

$$y_t = \mu + x_t'\alpha_1(1 - G(v_0 + x_t'\beta_0)) + x_t'\alpha_2 G(v_0 + x_t'\beta_0) + u_t$$

with unknown parameters μ, α_1 and α_2 only. The model is a regression with nonlinearity only in variables, the asymptotics of which can be derived with relative ease. On the other hand, the asymptotic distributions of $\hat{v}_n$ and $\hat{\beta}_n$ can be obtained from the NLS estimation of

$$y_t - \mu_0 - x_t'\alpha_{10} = x_t'\alpha_{20} G(v + x_t'\beta) + u_t,$$

where μ_0, α_{10} and α_{20} are assumed to be known.

[7] As for the SNNM, $\int_{-\infty}^{\infty} \mathrm{d}s \int_0^1 \mathrm{d}L_1(r,0) M(r,s) M(r,s)'$ is nonsingular a.s. See footnote 5.

[8] The potential misspecification error here is given by a nonlinear transformation of integrated processes, with the transformation function vanishing at infinity. As shown in Chang et al. (2001), the presence of such a transformation of integrated processes does not affect the asymptotic inferences on α_1, α_2 and μ.

When $c = H_2'(\alpha_{20} - \alpha_{10}) = 0$, the asymptotic results for $\hat{v}_n$ and $\hat{\beta}_n$ in Theorem 4 are no longer applicable, since $M = 0$ a.s. in this case. However, it is quite clear from the proof of Theorem 4 that (18) still holds with the rates $n^{3/4}$ and $n^{5/4}$ replaced by $n^{1/4}$ and $n^{3/4}$ respectively, and

$$M(r,s) = c(s\dot{G}_0(s), s^2\dot{G}_0(s), s\dot{G}_0(s)V_2(r)')',$$

where $c = h_1'(\alpha_{20} - \alpha_{10})$. If both $h_1'(\alpha_{20} - \alpha_{10}) = 0$ and $H_2'(\alpha_{20} - \alpha_{10}) = 0$ so that $\alpha_{10} = \alpha_{20}$, then β_0 is unidentified.

4. Inference in index models

In this section we consider the statistical inference in models introduced and analyzed in Section 3. Addressed are the problems of efficient estimation of, and hypothesis testing on those models. In general, the NLS estimator $\hat{\theta}_n$ is not efficient in the sense of Phillips (1991) and Saikkonen (1991), since it does not utilize the information on the presence of unit roots in the explanatory variables. However, following Chang et al. (2001), we may easily obtain the efficient estimator for θ.

Assumption 4. Assume

(a) $\Phi(z)$ is bounded and bounded away from zero for $|z| \leqslant 1$, and
(b) if we write $\Phi(z)^{-1} = 1 - \sum_{k=1}^{\infty} \Pi_k z^k$, then $\ell^s \sum_{k=\ell+1}^{\infty} |\Pi_k|^2 < \infty$ for some $s \geqslant 9$.

To estimate our models efficiently, we first run the regression

$$v_t = \hat{\Pi}_1 v_{t-1} + \cdots + \hat{\Pi}_\ell v_{t-\ell} + \hat{\varepsilon}_{\ell,t},$$

where we let ℓ increase as $n \to \infty$. More precisely, we let $\ell = n^\delta$, and let

$$\frac{r+2}{2r(s-3)} < \delta < \frac{r}{6+8r}, \tag{20}$$

where r is given by the moment condition for (ε_t), i.e., $\mathbf{E}\|\varepsilon_t\|^r < \infty$ for some $r > 8$ as given in Assumption 2. It is easy to see that δ satisfying condition (20) exists for all $r > 8$, if $s \geqslant 9$ as is assumed in Assumption 4. For Gaussian ARMA models, Assumptions 2 and 4 hold for any finite r and s. Then we may choose any δ such that $0 < \delta < 1/8$.

We define

$$y_t^* = y_t - \hat{\sigma}_{u\varepsilon}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\varepsilon}_{\ell,t+1},$$

where

$$\hat{\sigma}_{u\varepsilon} = \frac{1}{n}\sum_{t=1}^{n} \hat{u}_t \hat{\varepsilon}_{\ell,t+1} \quad \text{and} \quad \hat{\Sigma}_{\varepsilon\varepsilon} = \frac{1}{n}\sum_{t=1}^{n} \hat{\varepsilon}_{\ell,t}\hat{\varepsilon}_{\ell,t}'$$

with the first step NLS residual $\hat{u}_t$. Then in place of (1) we consider the regression

$$y_t^* = F(x_t, \theta_0) + u_t^*, \tag{21}$$

where $u_t^* = u_t - \hat{\sigma}_{u\varepsilon}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\varepsilon}_{l,t+1}$. Define $\hat{\theta}_n^*$ to be the NLS estimator for θ_0 from (21). This modified NLS estimator is called the *efficient nonstationary nonlinear least squares* (ENNLS) estimator. We also define W_* to be an independent set of Brownian motions that are independent of V and have variance $\sigma^{*2} = \sigma_u^2 - \omega_{uv}\Omega_{vv}^{-1}\omega_{vu}$ from σ_u^2.

Theorem 5. *Let Assumptions* 1–4 *hold, and the model is given by* (14) *or* (16). *Then we have*

$$C_n J'(\hat{\theta}_n^* - \theta_0) \to_d M^{-1/2} W_*(1),$$

where C_n *and* J *are as given in* (15) *and* (19), *and*

$$M = \mathrm{diag}\left(\int_0^1 \mathrm{N}(r)\mathrm{N}(r)'\,\mathrm{d}r, \int_{-\infty}^{\infty} \mathrm{d}s \int_0^1 \mathrm{d}L_1(r,0) M(r,s) M(r,s)' \right)$$

with $\mathrm{N}(r)$ *and* $M(r,s)$ *defined in Theorems* 3 *and* 4.

The limiting distribution of $\hat{\theta}_n^*$ is mixed normal. Moreover, the variance of mixture normal is reduced from σ_u^2 to σ^{*2}, which is the conditional longrun variance of (u_t) given (v_t). The ENNLS estimator $\hat{\theta}_n^*$ is therefore optimal in the sense of Phillips (1991) and Saikkonen (1991). See Section 5 of Chang et al. (2001) for the efficient estimation in nonlinear regressions with integrated time series.

Now we consider the hypothesis testing. Suppose that a nonlinear hypothesis on θ_0 is given by

$$\mathrm{H}_0 : R(\theta_0) = 0, \tag{22}$$

where $R : \mathbf{R}^p \to \mathbf{R}^q$ is continuously differentiable.[9] We define $\dot{R} = \partial R/\partial\theta'$. The Wald statistic for the hypothesis (22) is given by

$$W_n = \frac{R(\hat{\theta}_n)'(\dot{R}(\hat{\theta}_n)\ddot{Q}_n(\hat{\theta}_n)^{-1}\dot{R}(\hat{\theta}_n)')^{-1}R(\hat{\theta}_n)}{\hat{\sigma}_n^2} \tag{23}$$

in notation defined in Section 3. Since

$$C_n^{-1}J'\ddot{Q}_n(\hat{\theta}_n)JC_n^{-1} = C_n^{-1}J' \sum_{t=1}^{n} \dot{F}(x_t,\hat{\theta}_n)\dot{F}(x_t,\hat{\theta}_n)'JC_n^{-1} + \mathrm{o_p}(1)$$

as shown earlier, we may use $\sum_{t=1}^{n} \dot{F}(x_t,\hat{\theta}_n)\dot{F}(x_t,\hat{\theta}_n)'$ instead of $\ddot{Q}(\hat{\theta}_n)$ in the definition of the Wald test in (23).

For the models that we considered in Section 3 the limiting distribution of the Wald statistic W_n in (23) is in general not chi-square. It also depends on various nuisance parameters. Therefore, the test relying on the traditional chi-square values are generally invalid for such models. There are, however, some special cases where the test has a chi-square limiting distribution. First, if the hypothesis (22) only involves parameters v and β, then the Wald statistic W_n has limiting chi-square distribution. This is because the limiting distributions of $\hat{v}_n$ and $\hat{\beta}_n$ are mixed normal, as shown in Theorems 3

[9] We maintain that $\alpha_0 \neq 0$ for the SNNM, and $\alpha_{10} \neq \alpha_{20}$ for the STR. If this condition is violated, the parameters v and β are not identified.

and 4. Second, even if the hypothesis (22) is on other parameters μ and α, we may have limiting chi-square distribution for W_n when U and V are independent. Note that the distributions of $\hat{\mu}_n$ and $\hat{\alpha}_n$ are mixed normal for both Theorems 3 and 4 in this case, as we explained earlier.

As in Chang et al. (2001), we may use a modified test to avoid the nuisance parameter dependency problem. The modified Wald statistic is defined by

$$W_n^* = \frac{R(\hat{\theta}_n^*)'(\dot{R}(\hat{\theta}_n^*)\ddot{Q}_n(\hat{\theta}_n^*)^{-1}\dot{R}(\hat{\theta}_n^*)')^{-1}R(\hat{\theta}_n^*)}{\hat{\sigma}_n^{*2}}, \tag{24}$$

where $\hat{\theta}_n^*$ is the ENNLS estimator introduced above, and

$$\hat{\sigma}_n^{*2} = \hat{\sigma}_n^2 - \hat{\omega}_{uv}\hat{\Omega}_{vv}^{-1}\hat{\omega}_{vu}$$

with consistent estimates $\hat{\omega}_{uv}, \hat{\omega}_{vu}$ and $\hat{\Omega}_{vv}$ of covariances of U and V, and variance of V. Just as for the usual Wald statistic in (23), we may use $\sum_{t=1}^n \dot{F}(x_t, \hat{\theta}_n^*)\dot{F}(x_t, \hat{\theta}_n^*)'$ instead of $\ddot{Q}(\hat{\theta}_n^*)$ in (24).

Corollary 6. *Let Assumptions* 1–3 *hold. For the models considered in Section* 3, *we have*

$$\hat{\sigma}_n^2 \to_p \sigma_u^2$$

as $n \to \infty$.

Theorem 7. *Let Assumptions* 1–3 *hold. For the models considered in Section* 3, *we have*

$$W_n^* \to_d \chi_q^2$$

as $n \to \infty$.

We may also consider other tests based on the likelihood ratio-like (LR) statistic (or distance metric statistic in the terminology of Newey and McFadden, 1994) and Lagrange multiplier (LM) statistic. Denote them respectively by LR_n and LM_n. They require the estimation of the model with restrictions. If we denote by $\tilde{\theta}_n^*$ the restricted NLS estimator, corresponding to the unrestricted NLS estimator $\hat{\theta}_n^*$, of θ_0 based on the modified regression, then the statistics are given by

$$\mathrm{LR}_n = 2(Q_n(\tilde{\theta}_n^*) - Q_n(\hat{\theta}_n^*)),$$

$$\mathrm{LM}_n = \frac{\dot{Q}_n(\tilde{\theta}_n^*)'\ddot{Q}_n(\tilde{\theta}_n^*)^{-1}\dot{Q}_n(\tilde{\theta}_n^*)}{\hat{\sigma}_n^{*2}}.$$

In the definition of the LM_n statistic, we may replace $\hat{\sigma}_n^{*2}$ with $\tilde{\sigma}_n^{*2}$, say, which is computed from the restricted model. Given our previous results, it is quite clear that

$$\mathrm{LR}_n, \mathrm{LM}_n \to \chi_q^2$$

if the restricted models satisfy all the assumptions that we require for the corresponding unrestricted models.

5. Simulation

In this section we perform a set of simulations to investigate the finite sample properties of the NLS and the newly proposed ENNLS estimators in nonstationary index models. For the simulations, we consider the SNNM

$$y_t = \mu_0 + \alpha_0 G(x_{1t}\beta_{10} + x_{2t}\beta_{20}) + u_t, \tag{25}$$

where $G(x) = e^x/(1+e^x)$ is a logistic function. The true values of the parameters are set at $\mu_0 = 0, \alpha_0 = 1$ and

$$\beta_0 = (\beta_{10}, \beta_{20})' = (1, 0)'.$$

The regression error (u_t) and the regressors (x_t) are generated by

$$u_t = \varepsilon_{0,t+1}/\sqrt{2} + (\varepsilon_{1,t+1} + \varepsilon_{2,t+1})/2$$

and

$$\Delta x_t = v_t = \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} + \begin{pmatrix} 0.2 & 0 \\ 0 & 0.6 \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{pmatrix},$$

where (ε_{0t}), (ε_{1t}) and (ε_{2t}) are i.i.d. samples drawn from independent standard normal distributions.

By construction, the regression error (u_t) is an i.i.d. sequence and has no serial correlation. However, it is asymptotically correlated with the innovations (v_t) that generate the regressors (x_t), rendering their limit Brownian motions U and V dependent each other. With our choice of β_0 given above, the rotated regressors are simply given by

$$h_1' x_t = (1, 0)x_t = x_{1t} \quad \text{and} \quad h_2' x_t = (0, 1)x_t = x_{2t}$$

with the rotation matrix $\mathrm{H} = (h_1, h_2) = I_2$.

The limit theories of Theorems 3 and 5 readily apply to the NLS and ENNLS estimators for the parameters in our model (25). The NLS estimators of the intercept μ and the index function coefficient α converge at a rate $n^{1/2}$ to limit distributions that are biased and nonnormal, which implies that the limit distributions of the t-statistics based on them are nonstandard. In contrast, the NLS estimates of the parameters inside the index function, β_1 and β_2, converge to zero-mean mixed normal distributions at the rates $n^{1/4}$ and $n^{3/4}$, and consequently the t-statistics constructed from them have standard normal distributions. On the other hand, the limit distributions of the ENNLS estimators for μ, α, β_1 and β_2 are all mixed normal. Therefore, the standard test statistics based upon the ENNLS estimators are distributed asymptotically as standard normal or chi-square in all directions. Moreover, the ENNLS estimators have reduced longrun variances, and they are asymptotically more efficient than the NLS estimators.

Samples of sizes 250 and 500 are drawn 5,000 times to compare the finite sample performances of the NLS and ENNLS estimators and the t-statistics based on these estimators. The ENNLS correction terms are constructed from the one-period ahead fitted innovations $\hat{\varepsilon}_{t+1}$, which are obtained from the ℓth order vector autoregressions of v_t with $\ell = 1$ and 2, respectively for $n = 250$ and 500. For the NLS estimation,

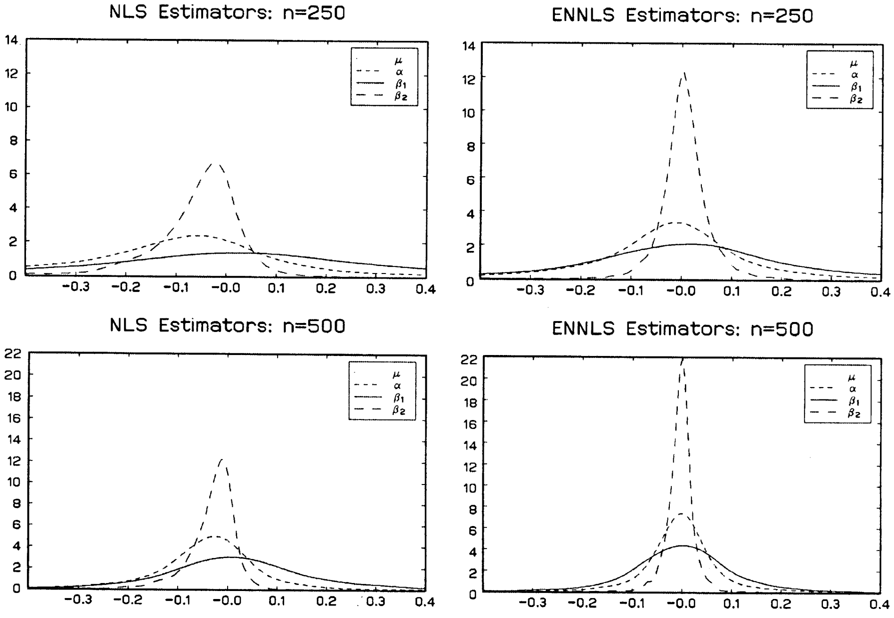

Fig. 1. Densities of estimators.

GAUSS optimization application with Gauss–Newton algorithm is used. Fig. 1 shows the density estimates of the NLS and ENNLS estimators for $n = 250$ and 500. The estimated densities of the t-statistics computed from the NLS and ENNLS estimators are given in Fig. 2 for $n = 250$ and 500.

Finite sample behavior of the NLS and ENNLS estimators are mostly consistent with the limit theories given in the previous sections. As can be seen clearly from Fig. 1, the finite sample distributions of the estimators with faster convergence rates do seem more concentrated than those with slower rates. The density estimates for the estimators of β_2 are most concentrated, while those of β_1 are most dispersed. As expected from the limit theory, the NLS estimators for both μ and α suffer from biases. Finite sample distribution of the NLS estimator for β_1, on the other hand, is well centered and symmetric, which again is expected from its asymptotics. However, the observations from the finite sample distribution of the NLS estimator for β_2 do not seem to support the limit theory. It has a noticeable bias, which does not seem to go away as the sample size increases. We may therefore say that the asymptotic approximation for the NLS estimator of β_2 is poor.

Finite sample performances of the ENNLS estimators are also as expected. As is clear from Fig. 1 again, all of the density estimates for the ENNLS estimators are very well centered and symmetric, which is quite in contrast with our earlier observations on the density estimates for the NLS estimators. The ENNLS estimators are also noticeably more concentrated around the true parameter values, as our theory suggests. It is worth noting that for the estimation of β_2 our correction for the ENNLS estimator does not

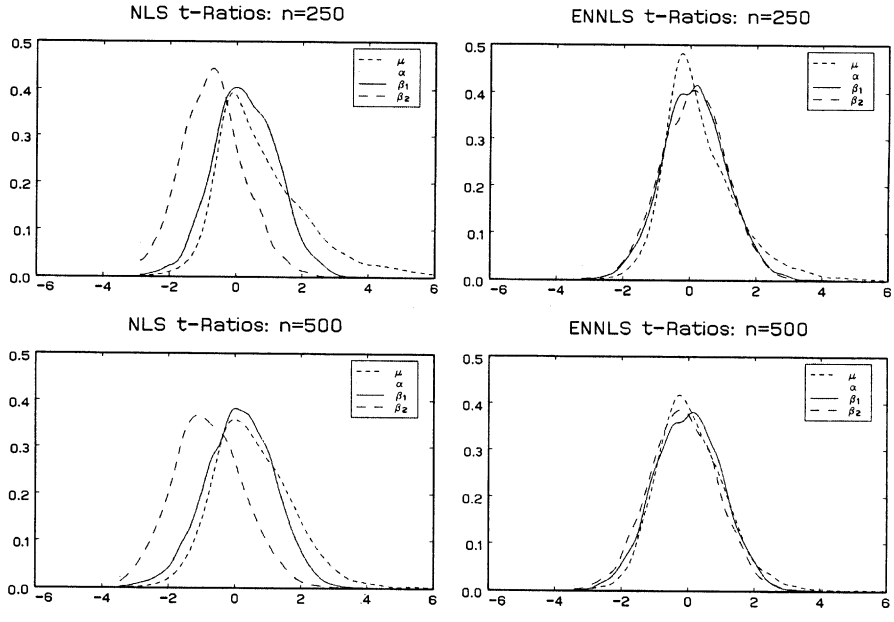

Fig. 2. Densities of t-ratios.

just reduce the sampling variance. It also effectively removes the finite sample bias and the distributional asymmetry of the NLS estimator of β_2. Our ENNLS procedure seems to improve the finite sample properties also for the estimators that are asymptotically mixed normal.

As can be seen clearly from the density estimates given in Fig. 2, the simulation study of the t-ratios based on the NLS and ENNLS estimators also corroborate our theoretical findings. As expected, the empirical distributions of the t-statistics based on the NLS estimators for β_1 and all of the ENNLS estimators indeed quite well approximate their limit standard normal distribution, and the approximation improves as the sample size increases. The finite sample distributions of the t-ratios constructed from the ENNLS estimators for β_1 and β_2 seem to approximate more closely their standard normal limit distribution than those constructed from the ENNLS estimators for μ and α.

The finite sample distribution of the t-statistics based on the NLS estimator for β_2, however, does not seem to properly approximate its limit standard normal distribution. It suffers from bias even in large samples, though it becomes quite symmetric as the sample size increases. This is expected from the poor asymptotic approximation of the NLS estimator for β_2 that we mentioned earlier. The sampling distributions of the t-ratios based on the NLS estimators for μ and α are nonstandard both in small and large samples, as is expected from their limit theories.

6. Conclusion

In this paper, we have established the statistical theories for nonstationary index models driven by integrated time series. The specification of our model is flexible enough to include simple neural network models and smooth transition regressions, which seem to have many potential applications. For these models, complete asymptotic results are provided. The usual NLS estimators are shown to be consistent, and have well defined asymptotic distributions which can be represented as functionals of Brownian motion and Brownian local time. Some components of the NLS estimators have limiting distributions that are mixed normal. However, they also have components whose asymptotic distributions are nonGaussian, biased and nuisance parameter dependent. In particular it is shown that applications of the usual statistical methods in such models generally yield inefficient estimates and/or invalid tests. We propose in the paper a new methodology to solve this problem. The new ENNLS procedure yields efficient estimators and allows us to perform the usual standard normal or chi-square tests.

7. Mathematical proofs

Proof of Lemma 1. As in Park and Phillips (2001), we may assume that

$$(U_n, V_n) \to_{\text{a.s.}} (U, V)$$

in $D[0,1]^m$ with uniform topology. Moreover, we may let U_n be given by

$$U_n\left(\frac{t}{n}\right) = U\left(\frac{\tau_{nt}}{n}\right),$$

where (τ_{nt}) is an increasing sequence of stopping times with $\tau_{n0} = 0$ a.s. and

$$\sup_{1\leqslant t\leqslant n}\left|\frac{\tau_{nt}-t}{n}\right| \to_{\text{a.s.}} 0 \tag{26}$$

as $n \to \infty$. See Park and Phillips (2001, Lemma 2.1) for details.

To prove the first part, we let

$$f_n(x) = \sum_{k=-\kappa_n}^{\kappa_n} f(k\delta_n)1\{k\delta_n \leqslant x < (k+1)\delta_n\},$$

where κ_n and δ_n are sequences of numbers satisfying conditions in the proof of Theorem 5.1 in Park and Phillips (1999). In particular, $\kappa_n \to \infty$ and $\delta_n \to 0$. Also, we let $\pi_n = \delta_n/\sqrt{n}$. It follows that

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n} f(x_{1t})x_{2t}^{i} = \sqrt{n}\int_0^1 f(\sqrt{n}V_{1n}(r))V_{2n}^{i}(r)\,\mathrm{d}r$$

$$= \sqrt{n}\int_0^1 f_n(\sqrt{n}V_{1n}(r))V_{2n}^{i}(r)\,\mathrm{d}r + \mathrm{o_p}(1)$$

$$
\begin{aligned}
&= \sqrt{n} \sum_{k=-\kappa_n}^{\kappa_n} f(k\delta_n) \int_0^1 1\{k\pi_n \leqslant V_{1n}(r) \leqslant (k+1)\pi_n\} \\
&\quad \times V_{2n}^i(r)\,\mathrm{d}r + \mathrm{o_p}(1) \\
&= \left(\delta_n \sum_{k=-\kappa_n}^{\kappa_n} f(k\delta_n) \right) \pi_n^{-1} \int_0^1 1\{0 \leqslant V_{1n}(r) < \pi_n\} \\
&\quad \times V_{2n}^i(r)\,\mathrm{d}r + \mathrm{o_p}(1) \\
&= \left(\int_{-\infty}^{\infty} f(s)\,\mathrm{d}s \right) \pi_n^{-1} \int_0^1 1\{0 \leqslant V_1(r) < \pi_n\} V_2^i(r)\,\mathrm{d}r + \mathrm{o_p}(1) \\
&= \left(\int_{-\infty}^{\infty} f(s)\,\mathrm{d}s \right) \int_0^1 \int_0^1 V_2^i(r)\,\mathrm{d}L_1(r,\pi_n s)\,\mathrm{d}s + \mathrm{o_p}(1) \\
&= \left(\int_{-\infty}^{\infty} f(s)\,\mathrm{d}s \right) \int_0^1 V_2^i(r)\,\mathrm{d}L_1(r,0) + \mathrm{o_p}(1)
\end{aligned}
$$

jointly for all i, $0 \leqslant i \leqslant \kappa$. Each step can be shown rigorously following the arguments in the proof of Lemma 5.1 of Park and Phillips (1999).

We now prove the result in the second part. In what follows, we let $m=2$ and $\kappa=1$, so that $K(x_1,x_2)=(f_0(x_1), f_1(x_1)x_2)'$. This is just to ease the exposition. The proof for the general case is essentially identical. For the general case with vector-valued (x_{2t}) and higher tensor product terms (x_{2t}^i) can be dealt with by considering their arbitrary linear combination. For $c=(c_1,c_2)\in\mathbf{R}^2$, we let

$$
T_n(x_1,x_2) = c_1 n^{-1/4} f_0(x_1) + c_2 n^{-3/4} f_1(x_1) x_2
$$

and write $T_n(V_n) = T_n(V_{1n}, V_{2n})$ subsequently. Define

$$
\begin{aligned}
M_n(r) = \sqrt{n} \sum_{i=1}^{t-1} T_n\left(\sqrt{n} V_n\left(\frac{i}{n}\right)\right)\left(U\left(\frac{\tau_{ni}}{n}\right) - U\left(\frac{\tau_{n,i-1}}{n}\right)\right) \\
+ \sqrt{n} T_n\left(\sqrt{n} V_n\left(\frac{t}{n}\right)\right)\left(U(r) - U\left(\frac{\tau_{n,t-1}}{n}\right)\right)
\end{aligned}
$$

for $\tau_{n,t-1}/n < r \leqslant \tau_{nt}/n$, where τ_{nt}, $t=1,\ldots,n$, are the stopping times introduced in Lemma 2.1 of Park and Phillips (2001). We may easily see that M_n is a continuous martingale such that

$$
\sum_{t=1}^{n} T_n(x_{1t}, x_{2t}) u_t = M_n\left(\frac{\tau_{nn}}{n}\right). \tag{27}
$$

Moreover,

$$
\sup_{1 \leqslant t \leqslant n} \left| \left(\frac{\tau_{nt}}{n} - \frac{\tau_{n,t-1}}{n} \right) - \frac{1}{n} \right| = \mathrm{o}(1) \quad \text{a.s.} \tag{28}
$$

which follows from (26).

Let $[M_n]$ be the quadratic variation of M_n. We have

$$[M_n](r) = n\sigma_u^2 \sum_{i=1}^{t-1} T_n^2\left(\sqrt{n}V_n\left(\frac{i}{n}\right)\right)\left(\frac{\tau_{ni}}{n} - \frac{\tau_{n,i-1}}{n}\right)$$

$$+ n\sigma_u^2 T_n^2\left(\sqrt{n}V_n\left(\frac{t}{n}\right)\right)\left(r - \frac{\tau_{n,t-1}}{n}\right)$$

$$= n\sigma_u^2 \int_0^r T_n^2(\sqrt{n}V_n(s))\,\mathrm{d}s + \mathrm{o_p}(1)$$

uniformly in $r \in [0,1]$, due to (28). Therefore,

$$[M_n](r) \to_p c'\left(\int_{-\infty}^{\infty} \mathrm{d}s \int_0^r \mathrm{d}L_1(t,0)K(s,V_2(t))K(s,V_2(t))'\right)c \tag{29}$$

uniformly in $r \in [0,1]$. Furthermore, if we denote by $[M_n, V]$ the covariation of M_n and V, then

$$[M_n, V](r) = \sqrt{n}\omega_{uv} \sum_{i=1}^{t-1} T_n\left(\sqrt{n}V_n\left(\frac{i}{n}\right)\right)\left(\frac{\tau_{ni}}{n} - \frac{\tau_{n,i-1}}{n}\right)$$

$$+ \sqrt{n}\omega_{uv}\, T_n\left(\sqrt{n}V_n\left(\frac{i}{n}\right)\right)\left(r - \frac{\tau_{n,t-1}}{n}\right)$$

$$= n^{-1/4}\left(n^{3/4}\omega_{uv} \int_0^r T_n(\sqrt{n}V_n(s))\,\mathrm{d}s + \mathrm{o_p}(1)\right)$$

uniformly in $r \in [0,1]$, due to (28). However,

$$\left|n^{3/4} \int_0^r T_n(\sqrt{n}V_n(s))\,\mathrm{d}s\right| \leqslant n^{3/4} \int_0^1 |T_n(\sqrt{n}V_n(s))|\,\mathrm{d}s = \mathrm{O_p}(1)$$

and we have

$$[M_n, V](\rho_n(r)) \to_p 0, \tag{30}$$

where $\rho_n(r) = \inf\{s \in [0,1]\colon [M_n](s) > r\}$ is a time change. The stated result now follows from (27), (29) and (30) as in the proof of Theorem 5.1 of Park and Phillips (1999). In particular, we have independence between W and V, due to (30).

The Brownian motion W is also independent of U. To see this, we look at the covariation $[M_n, U]$ of M_n and U. We have, exactly as for $[M_n, V]$ in (29) above,

$$[M_n, U](r) = \sqrt{n}\sigma_u^2 \sum_{i=1}^{t-1} T_n\left(\sqrt{n}V_n\left(\frac{i}{n}\right)\right)\left(\frac{\tau_{ni}}{n} - \frac{\tau_{n,i-1}}{n}\right)$$

$$+ \sqrt{n}\sigma_u^2 T_n\left(\sqrt{n}V_n\left(\frac{i}{n}\right)\right)\left(r - \frac{\tau_{n,t-1}}{n}\right)$$

$$= n^{-1/4}\left(n^{3/4}\sigma_u^2 \int_0^r T_n(\sqrt{n}V_n(s))\,\mathrm{d}s + \mathrm{o_p}(1)\right) \to_p 0$$

uniformly in $r \in [0,1]$. □

Proof of Lemma 2. Let $g_i = f_i - f_i^\circ$. Note that g_i's are bounded and vanish at infinity. We have

$$\frac{1}{n^{1+j/2}}\sum_{t=1}^{n} f_i(x_{1t})x_{2t}^j = \frac{1}{n^{1+j/2}}\sum_{t=1}^{n} f_i^\circ(x_{1t})x_{2t}^j + \frac{1}{n^{1+j/2}}\sum_{t=1}^{n} g_i(x_{1t})x_{2t}^j$$
$$= \frac{1}{n^{1+j/2}}\sum_{t=1}^{n} f_i^\circ(x_{1t})x_{2t}^j + \mathrm{o_p}(1)$$

due to Lemma A4 in Park and Phillips (2001). Apply the continuous mapping theorem to get

$$\frac{1}{n^{1+j/2}}\sum_{t=1}^{n} f_i^\circ(x_{1t})x_{2t}^j \to_d \int_0^1 f_i^\circ(V_1(r))V_2^j(r)$$

which proves the first part.

To show the second part, we notice from Lemma A4 in Park and Phillips (2001) that

$$\frac{1}{n^{(j+1)/2}}\sum_{t=1}^{n} f_i(x_{1t})x_{2t}^j u_t = \frac{1}{n^{(j+1)/2}}\sum_{t=1}^{n} f_i^\circ(x_{1t})x_{2t}^j u_t + \frac{1}{n^{(j+1)/2}}\sum_{t=1}^{n} g_i(x_{1t})x_{2t}^j u_t$$
$$= \frac{1}{n^{(j+1)/2}}\sum_{t=1}^{n} f_i^\circ(x_{1t})x_{2t}^j u_t + \mathrm{o_p}(1).$$

However, we have due to Kurz and Protter (1994)

$$\frac{1}{n^{(j+1)/2}}\sum_{t=1}^{n} f_i^\circ(x_{1t})x_{2t}^j u_t \to_d \int_0^1 f_i^\circ(V_1(r))V_2^j(r)\,\mathrm{d}U(r)$$

since $U_n \to_d U$ in $D[0,1]$ and

$$f_i^\circ(V_{1n})V_{2n}^j \to_d f_i^\circ(V_1)V_2^j$$

in $D[0,1]^{j(m-1)}$, jointly for all i and j, $0 \leqslant i,j \leqslant \kappa$. □

Lemma A1. *Let Assumptions 1 and 2 hold, and consider model (14). Assume that $\theta \in \Theta_n$, where Θ_n is defined in (12) with C_n given by either (15) or (19). For $f : \mathbf{R} \to \mathbf{R}$, we define $\dot{f}(x) = \mathrm{d}f(x)/\mathrm{d}x$ and $\dot{f}_i(x) = |x|^i \dot{f}(x)$. We let x^i be the i-times tensor product of x with itself, if x is a vector. Write $f_t = f(v + x_t'\beta)$ and $f_t^0 = f(v_0 + x_t'\beta_0)$ for notational simplicity.*

(a) *If f_i is bounded and integrable, then we have*

$$\sum_{t=1}^{n} f_t x_{1t}^i x_{2t}^j, \quad \sum_{t=1}^{n} f_t x_{1t}^i x_{2t}^j u_t = \mathrm{O_p}(n^{(j+1)/2})$$

uniformly in $\theta \in \Theta_n$, for all $i,j \geqslant 0$.

(b) *If $\dot f$ exists and if $\dot f_i$ and $\dot f_{i+1}$ are bounded and integrable, then we have*

$$\sum_{t=1}^{n}(f_t - f_t^0)x_{1t}^i x_{2t}^j, \quad \sum_{t=1}^{n}(f_t - f_t^0)x_{1t}^i x_{2t}^j u_t = \mathrm{O_p}(n^{(2j+1)/4+\delta})$$

uniformly in $\theta \in \Theta_n$, for all $i, j \geqslant 0$.

(c) *If $\dot f$ exists and if $\dot f_i$ and $\dot f_{i+1}$ are bounded and integrable, then we have*

$$\sum_{t=1}^{n}(\alpha^k f_t - \alpha_0^k f_t^0)x_{1t}^i x_{2t}^j, \quad \sum_{t=1}^{n}(\alpha^k f_t - \alpha_0^k f_t^0)x_{1t}^i x_{2t}^j u_t = \mathrm{O_p}(n^{(2j+1)/4+\delta})$$

uniformly in $\theta \in \Theta_n$, for all $i, j, k \geqslant 0$.

Proof of Lemma A1. For part (a), we let $a_0 = \|\beta_0\|$ and $b_0 = v_0$, and define

$$f_\varepsilon(x) = \sup_{|a-a_0| \leqslant \varepsilon} \sup_{|b-b_0| \leqslant \varepsilon} |f(ax+b)|$$

for any $\varepsilon > 0$ given. It can be shown that f_ε is bounded and integrable if f is, and for any $\varepsilon > 0$

$$|f_t| \leqslant f_\varepsilon(x_{1t}) \quad \text{a.s.}$$

for $1 \leqslant t \leqslant n$ as $n \to \infty$. We have

$$\left\|\sum_{t=1}^{n} f_t x_{1t}^i x_{2t}^j\right\| \leqslant \sum_{t=1}^{n} f_\varepsilon(x_{1t})|x_{1t}|^i \|x_{2t}\|^j = \mathrm{O_p}(n^{(j+1)/2})$$

and

$$\left\|\sum_{t=1}^{n} f_t x_{1t}^i x_{2t}^j u_t\right\| \leqslant \sum_{t=1}^{n} f_\varepsilon(x_{1t})|x_{1t}|^i \|x_{2t}\|^j |u_t| = \mathrm{O_p}(n^{(j+1)/2})$$

which prove part (a).

To show part (b), we define $\dot f_\varepsilon$ for $\dot f$ similarly as f_ε for f. Then we have

$$\begin{aligned} |f_t - f_t^0| &\leqslant \dot f_\varepsilon(x_{1t})|(v - v_0) + x_{1t}(\beta_1 - \|\beta_0\|) + x_{2t}'\beta_2| \\ &\leqslant n^{-1/4+\delta}\dot f_\varepsilon(x_{1t})(1 + |x_{1t}|) + n^{-3/4+\delta}\dot f_\varepsilon(x_{1t})\|x_{2t}\| \quad \text{a.s.} \end{aligned}$$

The stated results therefore follow directly from part (a).

It follows immediately from part (b) that

$$\begin{aligned} \sum_{t=1}^{n}(\alpha^k f_t - \alpha_0^k f_t^0)x_{1t}^i x_{2t}^j &= (\alpha^k - \alpha_0^k)\sum_{t=1}^{n} f_t x_{1t}^i x_{2t}^j + \alpha^k \sum_{t=1}^{n}(f_t - f_t^0)x_{1t}^i x_{2t}^j \\ &= \mathrm{O}(n^{-1/2+\delta})\mathrm{O_p}(n^{(j+1)/2}) + \mathrm{O_p}(n^{(2j+1)/4+\delta}) \\ &= \mathrm{O_p}(n^{(2j+1)/4+\delta}). \end{aligned}$$

Similarly, we have

$$\sum_{t=1}^{n}(\alpha^k f_t - \alpha_0^k f_t^0)x_{1t}^i x_{2t}^j u_t = (\alpha^k - \alpha_0^k)\sum_{t=1}^{n} f_t x_{1t}^i x_{2t}^j u_t + \alpha^k \sum_{t=1}^{n}(f_t - f_t^0)x_{1t}^i x_{2t}^j u_t$$
$$= \mathrm{O}(n^{-1/2+\delta})\mathrm{O_p}(n^{(j+1)/2}) + \mathrm{O_p}(n^{(2j+1)/4+\delta})$$
$$= \mathrm{O_p}(n^{(2j+1)/4+\delta})$$

which proves part (c). □

Proof of Theorem 3. For notational brevity, we let $\dot{F} = \dot{F}(x,\theta)$ and $\ddot{F} = \ddot{F}(x,\theta)$. Also, we write $G(v+x'\beta), \dot{G}(v+x'\beta)$ and $\ddot{G}(v+x'\beta)$ respectively as $G, \dot{G}$ and $\ddot{G}$. Then we have

$$\dot{F} = \begin{pmatrix} 1 \\ G \\ \alpha\dot{G} \\ \alpha\dot{G}x \end{pmatrix}, \quad \ddot{F} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \dot{G} & \dot{G}x' \\ 0 & \dot{G} & \alpha\ddot{G} & \alpha\ddot{G}x' \\ 0 & \dot{G}x & \alpha\ddot{G}x & \alpha\ddot{G}xx' \end{pmatrix}$$

and

$$\dot{F}\dot{F}' = \begin{pmatrix} 1 & G & \alpha\dot{G} & \alpha\dot{G}x' \\ G & G^2 & \alpha G\dot{G} & \alpha G\dot{G}x' \\ \alpha\dot{G} & \alpha G\dot{G} & \alpha^2\dot{G}^2 & \alpha^2\dot{G}^2x' \\ \alpha\dot{G}x & \alpha G\dot{G}x & \alpha^2\dot{G}^2x & \alpha^2\dot{G}^2xx' \end{pmatrix}.$$

We let C_n and J be defined as in (15). It follows from the second part of Lemmas 1 and 2 that

$$-C_n^{-1}J'\dot{Q}_n(\theta_0) = C_n^{-1}J'\sum_{t=1}^{n}\dot{F}(x_t,\theta_0)u_t$$
$$\to_d \begin{pmatrix} \int_0^1 \mathrm{N}(r)\,\mathrm{d}U(r) \\ \left(\int_{-\infty}^{\infty}\mathrm{d}s\int_0^1 \mathrm{d}L_1(r,0)M(r,s)M(r,s)'\right)^{1/2} W(1) \end{pmatrix}. \tag{31}$$

Moreover, we have

$$C_n^{-1}J'\sum_{t=1}^{n}\ddot{F}(x_t,\theta_0)u_t J C_n^{-1} \to_p 0$$

because

$$D_n^{-1}H'\sum_{t=1}^{n}\ddot{G}_0(x_{1t})x_tx_t'u_tHD_n^{-1}=\mathrm{O_p}(n^{-1/4})$$

$$n^{-3/4}\sum_{t=1}^{n}\dot{G}_0(x_{1t})u_t,\ \ n^{-3/4}\sum_{t=1}^{n}\dot{G}_0(x_{1t})x_{1t}u_t,\quad n^{-5/4}\sum_{t=1}^{n}\dot{G}_0(x_{1t})x_{2t}u_t=\mathrm{O_p}(n^{-1/2})$$

$$n^{-1/2}\sum_{t=1}^{n}\ddot{G}_0(x_{1t})u_t,\ \ n^{-1/2}\sum_{t=1}^{n}\ddot{G}_0(x_{1t})x_{1t}u_t,\quad n^{-1}\sum_{t=1}^{n}\ddot{G}_0(x_{1t})x_{2t}u_t=\mathrm{O_p}(n^{-1/4})$$

due to the second part of Lemmas 1 and 2, where $\ddot{G}_0$ is defined by $\ddot{G}_0(s)=\ddot{G}(v_0+\|\beta_0\|s)$ similarly as $\dot{G}_0$. Therefore, we have

$$C_n^{-1}J'\ddot{Q}_n(\theta_0)JC_n^{-1}=C_n^{-1}J'\sum_{t=1}^{n}\dot{F}(x_t,\theta_0)\dot{F}(x_t,\theta_0)'JC_n^{-1}+\mathrm{o_p}(1)$$

which converges in distribution to

$$\begin{pmatrix}\int_0^1 \mathrm{N}(r)\mathrm{N}(r)'\mathrm{d}r & 0\\ 0 & \int_{-\infty}^{\infty}\mathrm{d}s\int_0^1 \mathrm{d}L_1(r,0)M(r,s)M(r,s)'\end{pmatrix} \tag{32}$$

by the first part of Lemmas 1 and 2. For the block diagonality of the limiting distribution in (32), note that

$$n^{-3/4}\sum_{t=1}^{n}\dot{G}_0(x_{1t}),\quad n^{-3/4}\sum_{t=1}^{n}G_0(x_{1t})\dot{G}_0(x_{1t})=\mathrm{O_p}(n^{-1/4}),$$

$$n^{-3/4}\sum_{t=1}^{n}\dot{G}_0(x_{1t})x_{1t},\quad n^{-3/4}\sum_{t=1}^{n}G_0(x_{1t})\dot{G}_0(x_{1t})x_{1t}=\mathrm{O_p}(n^{-1/4}),$$

$$n^{-5/4}\sum_{t=1}^{n}\dot{G}_0(x_{1t})x_{2t},\quad n^{-5/4}\sum_{t=1}^{n}G_0(x_{1t})\dot{G}_0(x_{1t})x_{2t}=\mathrm{O_p}(n^{-1/4}),$$

where G_0 is defined by $G_0(s)=G(v_0+\|\beta_0\|s)$ similarly as $\dot{G}_0$. We thus have established (10). It therefore suffices to show (13). The stated results then follow immediately from (31) and (32).

To prove (13), we first write

$$\ddot{Q}_n(\theta)-\ddot{Q}_n(\theta_0)=A_n(\theta)+B_n(\theta)+C_n(\theta), \tag{33}$$

where

$$A_n(\theta) = \sum_{t=1}^{n} \dot{F}(x_t, \theta)\dot{F}(x_t, \theta)' - \sum_{t=1}^{n} \dot{F}(x_t, \theta_0)\dot{F}(x_t, \theta_0)',$$

$$B_n(\theta) = -\sum_{t=1}^{n} (\ddot{F}(x_t, \theta) - \ddot{F}(x_t, \theta_0))u_t,$$

$$C_n(\theta) = \sum_{t=1}^{n} \ddot{F}(x_t, \theta)(F(x_t, \theta) - F(x_t, \theta_0)).$$

Let $0 < \delta < 1/12$. It follows from Lemma A1(b) that

$$J'A_n(\theta)J = \begin{pmatrix} 0 & O_p(n^{1/4+\delta}) & O_p(n^{1/4+\delta}) & O_p(n^{1/4+\delta}) & O_p(n^{3/4+\delta}) \\ O_p(n^{1/4+\delta}) & O_p(n^{1/4+\delta}) & O_p(n^{1/4+\delta}) & O_p(n^{1/4+\delta}) & O_p(n^{3/4+\delta}) \\ O_p(n^{1/4+\delta}) & O_p(n^{1/4+\delta}) & O_p(n^{1/4+\delta}) & O_p(n^{1/4+\delta}) & O_p(n^{3/4+\delta}) \\ O_p(n^{1/4+\delta}) & O_p(n^{1/4+\delta}) & O_p(n^{1/4+\delta}) & O_p(n^{1/4+\delta}) & O_p(n^{3/4+\delta}) \\ O_p(n^{3/4+\delta}) & O_p(n^{3/4+\delta}) & O_p(n^{3/4+\delta}) & O_p(n^{3/4+\delta}) & O_p(n^{5/4+\delta}) \end{pmatrix}$$

and we have

$$C_{n\delta}^{-1}J'A_n(\theta)JC_{n\delta}^{-1} = o_p(1)$$

uniformly in $\theta \in \Theta_n$. Similarly, we have

$$J'B_n(\theta)J = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & O_p(n^{1/4+\delta}) & O_p(n^{1/4+\delta}) & O_p(n^{3/4+\delta}) \\ 0 & O_p(n^{1/4+\delta}) & O_p(n^{1/4+\delta}) & O_p(n^{1/4+\delta}) & O_p(n^{3/4+\delta}) \\ 0 & O_p(n^{1/4+\delta}) & O_p(n^{1/4+\delta}) & O_p(n^{1/4+\delta}) & O_p(n^{3/4+\delta}) \\ 0 & O_p(n^{3/4+\delta}) & O_p(n^{3/4+\delta}) & O_p(n^{3/4+\delta}) & O_p(n^{5/4+\delta}) \end{pmatrix}$$

and

$$C_{n\delta}^{-1}J'B_n(\theta)JC_{n\delta}^{-1} = o_p(1)$$

uniformly in $\theta \in \Theta_n$. Finally, to show that

$$C_{n\delta}^{-1}J'C_n(\theta)JC_{n\delta}^{-1} = o_p(1)$$

we note that $\ddot{F}$ is dominated in modulus by

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & c_1 & c_1 x' \\ 0 & c_1 & \alpha c_2 & c_2 x' \\ 0 & c_1 x & \alpha c_2 x & \alpha c_2 x x' \end{pmatrix},$$

where

$$c_1 = \sup_x |\dot{G}(x)| \quad \text{and} \quad c_2 = \sup_x |\ddot{G}(x)|. \tag{34}$$

Therefore, we may easily deduce from Lemma A1(b) that $J'C_n(\theta)J$ is stochastically at most of the order given by the matrix that we used to bound $J'B_n(\theta)J$. This completes the proof. □

Proof of Theorem 4. As in Proof of Theorem 3, we prove the stated results by showing (10) and (13). Here we have

$$F(x,\theta) = \mu + x'\alpha_1(1 - G(v + x'\beta)) + x'\alpha_2 G(v + x'\beta).$$

Then in the notations introduced in Proof of Theorem 3 we have

$$\dot{F} = \begin{pmatrix} 1 \\ (1-G)x \\ Gx \\ x'(\alpha_2 - \alpha_1)\dot{G} \\ x'(\alpha_2 - \alpha_1)\dot{G}x \end{pmatrix},$$

$$\ddot{F} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\dot{G}x & -\dot{G}xx' \\ 0 & 0 & 0 & \dot{G}x & \dot{G}xx' \\ 0 & -\dot{G}x' & \dot{G}x' & x'(\alpha_2 - \alpha_1)\ddot{G} & x'(\alpha_2 - \alpha_1)\ddot{G}x' \\ 0 & -\dot{G}xx' & \dot{G}xx' & x'(\alpha_2 - \alpha_1)\ddot{G}x & x'(\alpha_2 - \alpha_1)\ddot{G}xx' \end{pmatrix}$$

and $\dot{F}\dot{F}'$ is given by

$$\begin{pmatrix} 1 & (1-G)x' & Gx' & x'(\alpha_2 - \alpha_1)\dot{G} & x'(\alpha_2 - \alpha_1)\dot{G}x' \\ & (1-G)^2xx' & (1-G)Gxx' & (1-G)\dot{G}x'(\alpha_2 - \alpha_1)x & (1-G)\dot{G}x'(\alpha_2 - \alpha_1)xx' \\ & & G^2xx' & G\dot{G}x'(\alpha_2 - \alpha_1)x & G\dot{G}x'(\alpha_2 - \alpha_1)xx' \\ & & & (x'(\alpha_2 - \alpha_1))^2\dot{G}^2 & (x'(\alpha_2 - \alpha_1))^2\dot{G}^2x' \\ & & & & (x'(\alpha_2 - \alpha_1))^2\dot{G}^2xx' \end{pmatrix}.$$

Let C_n and J be given by (19), and let G_0 be defined as in Proof of Theorem 3. Then we have from the second part of Lemmas 1 and 2 that

$$-C_n^{-1}J'\dot{Q}_n(\theta_0) = C_n^{-1}J'\sum_{t=1}^n \dot{F}(x_t,\theta_0)u_t$$

$$= \begin{pmatrix} n^{-1/2}\sum_{t=1}^{n} u_t \\ n^{-1}H'\sum_{t=1}^{n}(1-G_0(x_{1t}))x_t u_t \\ n^{-1}H'\sum_{t=1}^{n}G_0(x_{1t})x_t u_t \\ n^{-3/4}\sum_{t=1}^{n}\dot{G}_0(x_{1t})(\alpha_{20}-\alpha_{10})'x_t u_t \\ D_n^{-1}H'\sum_{t=1}^{n}\dot{G}_0(x_{1t})(\alpha_{20}-\alpha_{10})'x_t x_t u_t \end{pmatrix}$$

$$= \begin{pmatrix} n^{-1/2}\sum_{t=1}^{n} u_t \\ n^{-1}\sum_{t=1}^{n}(1-G_0(x_{1t}))H'x_t u_t \\ n^{-1}\sum_{t=1}^{n}G_0(x_{1t})H'x_t u_t \\ n^{-3/4}\sum_{t=1}^{n}\dot{G}_0(x_{1t})(\alpha_{20}-\alpha_{10})'HH'x_t u_t \\ D_n^{-1}\sum_{t=1}^{n}\dot{G}_0(x_{1t})(\alpha_{20}-\alpha_{10})'HH'x_t H'x_t u_t \end{pmatrix}$$

$$= \begin{pmatrix} n^{-1/2}\sum_{t=1}^{n} u_t \\ n^{-1}\sum_{t=1}^{n}(1-G_0(x_{1t}))\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} u_t \\ n^{-1}\sum_{t=1}^{n}G_0(x_{1t})\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} u_t \\ n^{-3/4}\sum_{t=1}^{n}\dot{G}_0(x_{1t})((\alpha_{20}-\alpha_{10})'h_1 x_{1t}+(\alpha_{20}-\alpha_{10})'H_2 x_{2t})u_t \\ n^{-3/4}\sum_{t=1}^{n}\dot{G}_0(x_{1t})((\alpha_{20}-\alpha_{10})'h_1 x_{1t}+(\alpha_{20}-\alpha_{10})'H_2 x_{2t})x_{1t}u_t \\ n^{-5/4}\sum_{t=1}^{n}\dot{G}_0(x_{1t})((\alpha_{20}-\alpha_{10})'h_1 x_{1t}+(\alpha_{20}-\alpha_{10})'H_2 x_{2t})x_{2t}u_t \end{pmatrix}$$

$$\to_d \begin{pmatrix} \int_0^1 \mathrm{N}(r)\,\mathrm{d}U(r) \\ \left(\int_{-\infty}^{\infty} \mathrm{d}s \int_0^1 \mathrm{d}L_1(r,0)M(r,s)M(r,s)'\right)^{1/2} W(1) \end{pmatrix}, \tag{35}$$

where

$$N(r) = \begin{pmatrix} 1 \\ 1\{V_1(r) < 0\}V(r) \\ 1\{V_1(r) \geqslant 0\}V(r) \end{pmatrix} \quad \text{and} \quad M(r,s) = \begin{pmatrix} \dot{G}_0(s)c'V_2(r) \\ s\dot{G}_0(s)c'V_2(r) \\ \dot{G}_0(s)c'V_2(r)V_2(r) \end{pmatrix}$$

with $c = H_2'(\alpha_{20} - \alpha_{10})$. This is because

$$n^{-3/4}\sum_{t=1}^n \dot{G}_0(x_{1t})x_{1t}u_t, \quad n^{-3/4}\sum_{t=1}^n \dot{G}_0(x_{1t})x_{1t}^2u_t,$$

$$n^{-5/4}\sum_{t=1}^n \dot{G}_0(x_{1t})x_{1t}x_{2t}u_t = \mathrm{O_p}(n^{-1/2})$$

due to the second part of Lemma 1.

We also have

$$C_n^{-1}J'\sum_{t=1}^n \ddot{F}(x_t,\theta_0)u_tJC_n^{-1} \to_p 0$$

since

$$n^{-7/4}\sum_{t=1}^n \dot{G}_0(x_{1t})x_tu_t = \begin{pmatrix} \mathrm{O_p}(n^{-3/2}) \\ \mathrm{O_p}(n^{-1}) \end{pmatrix},$$

$$n^{-1}H'\sum_{t=1}^n \dot{G}_0(x_{1t})x_tx_t'u_tHD_n^{-1} = \begin{pmatrix} \mathrm{O_p}(n^{-3/2}) & \mathrm{O_p}(n^{-3/2}) \\ \mathrm{O_p}(n^{-1}) & \mathrm{O_p}(n^{-1}) \end{pmatrix},$$

$$n^{-3/2}\sum_{t=1}^n \ddot{G}_0(x_{1t})(\alpha_{20}-\alpha_{10})'x_tu_t = \mathrm{O_p}(n^{-3/4}),$$

$$n^{-3/4}\sum_{t=1}^n \ddot{G}_0(x_{1t})(\alpha_{20}-\alpha_{10})'x_tx_t'u_tHD_n^{-1} = (\mathrm{O_p}(n^{-3/4}), \mathrm{O_p}(n^{-3/4})),$$

$$D_n^{-1}H'\sum_{t=1}^n (\alpha_{20}-\alpha_{10})'x_tx_tx_t'u_tHD_n^{-1} = \begin{pmatrix} \mathrm{O_p}(n^{-3/4}) & \mathrm{O_p}(n^{-3/4}) \\ \mathrm{O_p}(n^{-3/4}) & \mathrm{O_p}(n^{-3/4}) \end{pmatrix}.$$

Then it follows that

$$\begin{aligned}
&C_n^{-1}J'\ddot{Q}_n(\theta_0)JC_n \\
&=C_n^{-1}J'\sum_{t=1}^{n}\dot{F}(x_t,\theta_0)\dot{F}(x_t,\theta_0)'JC_n^{-1}+\mathrm{o_p}(1) \\
&\to_d \begin{pmatrix} \int_0^1 \mathrm{N}(r)\mathrm{N}(r)'\mathrm{d}r & 0 \\ 0 & \int_{-\infty}^{\infty}\mathrm{d}s\int_0^1 \mathrm{d}L_1(r,0)M(r,s)M(r,s)' \end{pmatrix}
\end{aligned} \tag{36}$$

by the first part of Lemma 1 and the second part of Lemma 2. The block diagonality above holds since

$$n^{-5/4}\sum_{t=1}^{n}x_t'(\alpha_{20}-\alpha_{10})\dot{G}_0(x_{1t})=\mathrm{O_p}(n^{-1/4}),$$

$$n^{-1/2}\sum_{t=1}^{n}x_t'(\alpha_{20}-\alpha_{10})\dot{G}_0(x_{1t})x_t'HD_n^{-1}=(\mathrm{O_p}(n^{-1/4}),\mathrm{O_p}(n^{-1/4})),$$

$$n^{-7/4}H'\sum_{t=1}^{n}(1-G_0(x_{1t}))\dot{G}_0(x_{1t})x_tx_t'(\alpha_{20}-\alpha_{10})=\begin{pmatrix}\mathrm{O_p}(n^{-3/4})\\ \mathrm{O_p}(n^{-1/4})\end{pmatrix},$$

$$\begin{aligned}
&n^{-1}H'\sum_{t=1}^{n}(1-G_0(x_{1t}))\dot{G}_0(x_{1t})x_t'(\alpha_{20}-\alpha_{10})x_tx_t'HD_n^{-1} \\
&=\begin{pmatrix}\mathrm{O_p}(n^{-3/4}) & \mathrm{O_p}(n^{-3/4})\\ \mathrm{O_p}(n^{-1/4}) & \mathrm{O_p}(n^{-1/4})\end{pmatrix},
\end{aligned}$$

$$n^{-7/4}H'\sum_{t=1}^{n}G_0(x_{1t})\dot{G}_0(x_{1t})x_tx_t'(\alpha_{20}-\alpha_{10})=\begin{pmatrix}\mathrm{O_p}(n^{-3/4})\\ \mathrm{O_p}(n^{-1/4})\end{pmatrix},$$

$$n^{-1}\sum_{t=1}^{n}G_0(x_{1t})\dot{G}_0(x_{1t})x_t'(\alpha_{20}-\alpha_{10})x_tx_t'HD_n^{-1}=\begin{pmatrix}\mathrm{O_p}(n^{-3/4}) & \mathrm{O_p}(n^{-3/4})\\ \mathrm{O_p}(n^{-1/4}) & \mathrm{O_p}(n^{-1/4})\end{pmatrix}.$$

By (35) and (36), we have established (10) for the model (16).

Now we may show (13) just as in Proof of Theorem 3, using the decomposition given in (33). Let $0 < \delta < 1/12$. Then, due to Lemma A1(b), we can write $J'A_n(\theta)J$ as

$$\begin{pmatrix}
0 & \mathrm{O_p}(n^{\frac{1}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{1}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{5}{4}+\delta}) \\
 & \mathrm{O_p}(n^{\frac{1}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{1}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{5}{4}+\delta}) \\
 & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{5}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{5}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{5}{4}+\delta}) & \mathrm{O_p}(n^{\frac{7}{4}+\delta}) \\
 & & & \mathrm{O_p}(n^{\frac{1}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{5}{4}+\delta}) \\
 & & & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{5}{4}+\delta}) & \mathrm{O_p}(n^{\frac{5}{4}+\delta}) & \mathrm{O_p}(n^{\frac{5}{4}+\delta}) & \mathrm{O_p}(n^{\frac{7}{4}+\delta}) \\
 & & & & & \mathrm{O_p}(n^{\frac{5}{4}+\delta}) & \mathrm{O_p}(n^{\frac{5}{4}+\delta}) & \mathrm{O_p}(n^{\frac{7}{4}+\delta}) \\
 & & & & & & \mathrm{O_p}(n^{\frac{5}{4}+\delta}) & \mathrm{O_p}(n^{\frac{7}{4}+\delta}) \\
 & & & & & & \mathrm{O_p}(n^{\frac{7}{4}+\delta}) & \mathrm{O_p}(n^{\frac{9}{4}+\delta})
\end{pmatrix}$$

giving

$$C_{n\delta}^{-1}J'A_n(\theta)JC_{n\delta}^{-1} = \mathrm{o_p}(1)$$

uniformly in $\theta \in \Theta_n$. Similarly, we write $J'B_n(\theta)J$ as

$$\begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \mathrm{O_p}(n^{\frac{1}{4}+\delta}) & \mathrm{O_p}(n^{\frac{1}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) \\
0 & 0 & 0 & 0 & 0 & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{5}{4}+\delta}) \\
0 & 0 & 0 & 0 & 0 & \mathrm{O_p}(n^{\frac{1}{4}+\delta}) & \mathrm{O_p}(n^{\frac{1}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) \\
0 & 0 & 0 & 0 & 0 & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{5}{4}+\delta}) \\
0 & \mathrm{O_p}(n^{\frac{1}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{1}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{5}{4}+\delta}) \\
0 & \mathrm{O_p}(n^{\frac{1}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{1}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{5}{4}+\delta}) \\
0 & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{5}{4}+\delta}) & \mathrm{O_p}(n^{\frac{3}{4}+\delta}) & \mathrm{O_p}(n^{\frac{5}{4}+\delta}) & \mathrm{O_p}(n^{\frac{5}{4}+\delta}) & \mathrm{O_p}(n^{\frac{5}{4}+\delta}) & \mathrm{O_p}(n^{\frac{7}{4}+\delta})
\end{pmatrix} \tag{37}$$

by Lemma A1(b). Clearly, $C_{n\delta}^{-1}J'B_n(\theta)JC_{n\delta}^{-1} = \mathrm{o_p}(1)$, uniformly in $\theta \in \Theta_n$. Next, we note that $\ddot{F}$ is dominated in modulus by

$$\ddot{F} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & c_1x & c_1xx' \\ 0 & 0 & 0 & c_1x & c_1xx' \\ 0 & c_1x' & c_1x' & c_1x'(\alpha_2-\alpha_1) & c_2x'(\alpha_2-\alpha_1)x' \\ 0 & c_1xx' & c_1xx' & c_2x'(\alpha_2-\alpha_1)x & c_2x'(\alpha_2-\alpha_1)xx' \end{pmatrix},$$

where c_1 and c_2 are defined in (34). It is easy to see from Lemma A1(b) that $J'C_n(\theta)J$ is stochastically at most of the order given by (37) above, and this implies $C_{n\delta}^{-1}J'C_n(\theta)JC_{n\delta}^{-1} = \mathrm{o_p}(1)$. The proof is now complete. □

Proof of Theorem 5. The stated result follows immediately from Chang et al. (2001), upon noting that W introduced in Theorems 3 and 4 is independent of both U and V. □

Proof of Corollary 6. Define

$$\sigma_n^2 = \frac{1}{n}\sum_{t=1}^{n} u_t^2.$$

It follows from Assumption 1 that $\sigma_n^2 \to_p \sigma_u^2$. Furthermore, we have

$$|\hat{\sigma}_n^2 - \sigma_n^2| \leqslant A_n + 2B_n,$$

where

$$A_n = \frac{1}{n}\sum_{t=1}^{n}(F(x_t,\hat{\theta}_n) - F(x_t,\theta_0))^2,$$

$$B_n = \left|\frac{1}{n}\sum_{t=1}^{n}(F(x_t,\hat{\theta}_n) - F(x_t,\theta_0))u_t\right| \leqslant (\sigma_n^2 A_n)^{1/2}.$$

Therefore, it suffices to show that $A_n \to 0$.

Define $\hat{G}_{nt} = G(\hat{v}_n + x_t'\hat{\beta}_n)$ and $G_{0t} = G(v_0 + x_t'\beta_0)$. For the SNNM (14), we have

$$F(x_t,\hat{\theta}_n) - F(x_t,\theta_0) = (\hat{\mu}_n - \mu_0) + (\hat{\alpha}_n - \alpha_0)\hat{G}_{nt} + \alpha_0(\hat{G}_{nt} - G_{0t}),$$

where

$$\hat{\mu}_n - \mu_0 = \mathrm{O_p}(n^{-1/2}), \quad \hat{\alpha}_n - \alpha_0 = \mathrm{O_p}(n^{-1/2})$$

from Theorem 3 and

$$\sum_{t=1}^{n}|\hat{G}_{nt} - G_{0t}| = \mathrm{O_p}(n^{1/4}) \tag{38}$$

as shown in Lemma A1(b). Then it follows that

$$A_n = \frac{1}{n}\sum_{t=1}^{n}(\hat\mu_n-\mu_0)^2 + \frac{1}{n}\sum_{t=1}^{n}(\hat\alpha_n-\alpha_0)^2\hat G_{nt}^2 + \frac{2}{n}\sum_{t=1}^{n}(\hat\mu_n-\mu_0)(\hat\alpha_n-\alpha_0)\hat G_{nt}$$
$$+\frac{2\alpha_0}{n}\sum_{t=1}^{n}(\hat\mu_n-\mu_0)(\hat G_{nt}-G_{0t}) + \frac{2\alpha_0}{n}\sum_{t=1}^{n}(\hat\alpha_n-\alpha_0)\hat G_{nt}(\hat G_{nt}-G_{0t})$$
$$= \mathrm{O_p}(n^{-1}) + \mathrm{O_p}(n^{-1}) + \mathrm{O_p}(n^{-3/4}) + \mathrm{O_p}(n^{-1}) + \mathrm{O_p}(n^{-5/4}) + \mathrm{O_p}(n^{-1/2}).$$

Clearly, $A_n = \mathrm{O_p}(n^{-1/2}) = \mathrm{o_p}(1)$.

On the other hand, we have for the STR in (16)

$$F(x_t,\hat\theta_n) - F(x_t,\theta_0) = (\hat\mu_n-\mu_0) + x_t'(\hat\alpha_{1n}-\alpha_{10}) + x_t'((\hat\alpha_{1n}-\alpha_{10})$$
$$+(\hat\alpha_{2n}-\alpha_{20}))\hat G_{nt} + x_t'(\alpha_{20}-\alpha_{10})(\hat G_{nt}-G_{0t}),$$

where

$$\hat\mu_n-\mu_0 = \mathrm{O_p}(n^{-1/2}), \quad \hat\alpha_{1n}-\alpha_{10} = \mathrm{O_p}(n^{-1}), \quad \hat\alpha_{2n}-\alpha_{20} = \mathrm{O_p}(n^{-1})$$

as shown in Theorem 4. Now we may easily deduce from this and (38) that

$$A_n = \frac{1}{n}\sum_{t=1}^{n}((\hat\mu_n-\mu_0)^2 + (x_t'(\hat\alpha_{1n}-\alpha_{10}))^2$$
$$+(x_t'((\hat\alpha_{1n}-\alpha_{10})+(\hat\alpha_{2n}-\alpha_{20})))^2\hat G_{nt}^2 + (x_t'(\alpha_{20}-\alpha_{10}))^2(\hat G_{nt}-G_{0t})^2$$
$$+2(\hat\mu_n-\mu_0)x_t'(\hat\alpha_{1n}-\alpha_{10}) + 2(\hat\mu_n-\mu_0)x_t'((\hat\alpha_{1n}-\alpha_{10})-(\hat\alpha_{2n}-\alpha_{20}))\hat G_{nt}$$
$$+2(\hat\mu_n-\mu_0)x_t'(\hat\alpha_{2n}-\alpha_{20})(\hat G_{nt}-G_{0t})$$
$$+2x_t'(\hat\alpha_{1n}-\alpha_{10})x_t'((\hat\alpha_{1n}-\alpha_{10})-(\hat\alpha_{2n}-\alpha_{20}))\hat G_{nt}$$
$$+2x_t'(\hat\alpha_{1n}-\alpha_{10})x_t'(\hat\alpha_{2n}-\alpha_{20})(\hat G_{nt}-G_{0t})$$
$$+2x_t'((\hat\alpha_{1n}-\alpha_{10})-(\hat\alpha_{2n}-\alpha_{20}))x_t'(\hat\alpha_{2n}-\alpha_{20})\hat G_{nt}(\hat G_{nt}-G_{0t}))$$
$$= \mathrm{O_p}(n^{-1}) + \mathrm{O_p}(n^{-1}) + \mathrm{O_p}(n^{-1}) + \mathrm{O_p}(n^{-7/4}) + \mathrm{O_p}(n^{-1})$$
$$+\mathrm{O_p}(n^{-1}) + \mathrm{O_p}(n^{-7/4}) + \mathrm{O_p}(n^{-7/4}) + \mathrm{O_p}(n^{-1}) + \mathrm{O_p}(n^{-1}).$$

Hence $A_n = \mathrm{O_p}(n^{-1}) = \mathrm{o_p}(1)$ also for the STR model. □

Proof of Theorem 7. Assume that there exists a diagonal matrix D_n such that if we define

$$P_n = D_n\dot R(\hat\theta_n)JC_n^{-1}$$

then

$$P_n \to_d P,$$

where P is a.s. of full row rank. The assumption holds if and only if the restrictions are linearly independent asymptotically. It causes no loss in generality, since we may always formulate the given set of restrictions in such a way that they are not collinear in the limit. For instance, we may want to test $\mu + \nu = 0$ and $\nu = 0$ jointly in the SNNM (14). This set of hypotheses are not asymptotically linearly independent, since

$$\dot{R}(\hat{\theta}_n)JC_n^{-1} = \begin{pmatrix} n^{-1/2} & 0 & n^{-1/4} & 0 & 0 \\ 0 & 0 & n^{-1/4} & 0 & 0 \end{pmatrix}$$

and there is no normalizing matrix D_n for which its rows become linearly independent asymptotically. However, we may reformulate it as $\mu = 0$ and $\nu = 0$. For the reformulated restrictions, we have

$$\dot{R}(\hat{\theta}_n)JC_n^{-1} = \begin{pmatrix} n^{-1/2} & 0 & 0 & 0 & 0 \\ 0 & 0 & n^{-1/4} & 0 & 0 \end{pmatrix}$$

and we may simply let $D_n = \text{diag}(n^{1/2}, n^{1/4})$.

By the mean value theorem, we have

$$R(\hat{\theta}_n^*) = \dot{R}(\theta_n)(\hat{\theta}_n^* - \theta_0),$$

where θ_n lies in the line segment connecting $\hat{\theta}_n^*$ and θ_0. It follows that

$$D_n R(\hat{\theta}_n^*) = D_n \dot{R}(\theta_n)JC_n^{-1}(C_n J'(\hat{\theta}_n^* - \theta_0))$$

and consequently,

$$D_n R(\hat{\theta}_n^*) \to_d W_*(PM^{-1}P'),$$

where W_* and M are given in Theorem 5. The stated result can now be easily deduced upon noticing that the numerator of W_n^* can be written as

$$R(\hat{\theta}_n^*)' D_n \left(D_n \dot{R}(\hat{\theta}_n^*)JC_n^{-1} \left(C_n^{-1} J' \ddot{Q}_n(\hat{\theta}_n^*)JC_n^{-1} \right)^{-1} C_n^{-1} J' \dot{R}(\hat{\theta}_n^*)' D_n \right)^{-1} D_n R(\hat{\theta}_n^*)$$

since

$$C_n^{-1} J' \ddot{Q}(\hat{\theta}_n^*)JC_n^{-1} = C_n^{-1} J' \sum_{t=1}^{n} \dot{F}(x_t, \hat{\theta}_n)\dot{F}(x_t, \hat{\theta}_n)' JC_n^{-1} + o_p(1) \to_d M.$$

The proof is therefore complete. □

Acknowledgements

We are grateful to an Editor, an Associate Editor and anonymous referees for helpful comments. An earlier version of this paper was presented by Chang at the 1998 Econometric Society Summer Meeting at Montreal, Canada. Chang gratefully acknowledges financial support from Rice University through CSIV fund. Park thanks the Department of Economics at Rice University, where he is an Adjunct Professor, for its continuing hospitality, and the Korea Research Foundation for financial support.

References

Andrews, D.W.K., McDermott, C.J., 1995. Nonlinear econometric models with deterministically trending variables. Review of Economic Studies 62, 343–360.

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

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

Jennrich, R.I., 1969. Asymptotic properties of non-linear least squares estimation. Annals of Mathematical Statistics 40, 633–643.

Newey, W.K., McFadden, D., 1994. Large sample estimation and hypothesis testing. In: Engle, R., McFadden, D.L. (Eds.), Handbook of Econometrics, Vol. IV. North-Holland: Amsterdam, pp. 2639–2738.

Park, J.Y., 1992. Canonical cointegrating regressions. Econometrica 60, 119–143.

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

Park, J.Y., Phillips, P.C.B., 2000. Nonstationary binary choice. Econometrica 68, 1249–1280.

Park, J.Y., Phillips, P.C.B., 2001. Nonlinear regressions with integrated time series. Econometrica 69, 117–162.

Phillips, P.C.B., 1991. Optimal inference in cointegrated systems. Econometrica 59, 283–306.

Phillips, P.C.B., Hansen, B.E., 1990. Statistical inference in instrumental variables regressions with I(1) processes. Review of Economic Studies 57, 99–125.

Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. Annals of Statistics 20, 971–1001.

Saikkonnen, P., 1991. Asymptotically efficient estimation of cointegrating regressions. Econometric Theory 7, 1–21.

White, H., 1989. Some asymptotic results for learning in single hidden-layer feedforward network models. Journal of the American Statistical Association 84, 1003–1013.

Wooldridge, J.M., 1994. Estimation and inference for dependent processes. In: Engle, R., McFadden, D.L. (Eds.), Handbook of Econometrics, Vol. IV. North-Holland: Amsterdam, pp. 2639–2738.

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