NONSTATIONARY DISCRETE CHOICE:
A CORRIGENDUM AND ADDENDUM

BY

PETER C. B. PHILIPS, SAINAN JIN and LING HU

COWLES FOUNDATION PAPER NO. 1214

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2007

http://cowles.econ.yale.edu/

Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 141 (2007) 1115–1130
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Nonstationary discrete choice: A corrigendum and addendum

Peter C.B. Phillips[a,b,c,*], Sainan Jin[d], Ling Hu[e]

[a]*Cowles Foundation, Yale University, USA*
[b]*University of Auckland, New Zealand*
[c]*University of York, UK*
[d]*Guanghua School of Management, Peking University, China*
[e]*Department of Economics, Ohio State University, USA*

Available online 26 March 2007

Abstract

We correct the limit theory presented in an earlier paper by Hu and Phillips [2004a. Nonstationary discrete choice. Journal of Econometrics 120, 103–138] for nonstationary time series discrete choice models with multiple choices and thresholds. The new limit theory shows that, in contrast to the binary choice model with nonstationary regressors and a zero threshold where there are dual rates of convergence ($n^{1/4}$ and $n^{3/4}$), all parameters including the thresholds converge at the rate $n^{3/4}$. The presence of nonzero thresholds therefore materially affects rates of convergence. Dual rates of convergence reappear when stationary variables are present in the system. Some simulation evidence is provided, showing how the magnitude of the thresholds affects finite sample performance. A new finding is that predicted probabilities and marginal effect estimates have finite sample distributions that manifest a pile-up, or increasing density, towards the limits of the domain of definition.

JEL classification: C23; C25

Keywords: Brownian motion; Brownian local time; Discrete choices; Integrated processes; Pile-up problem; Threshold parameters

*Corresponding author. Cowles Foundation, Yale University, USA.
E-mail address: peter.phillips@yale.edu (P.C.B. Phillips).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2007.01.017

1. Introduction

This note corrects the limit theory given in Hu and Phillips (2004a, hereafter HP) for discrete choice models with integrated covariates and nonzero thresholds that determine an ordered set of choices. The error occurs in Lemma 1 and Theorem 1 of HP. Those results sought to provide the asymptotic theory for sample moment expressions that appear in the score function and hessian (Eqs. (7)–(9), in HP); and they gave dual rates of convergence ($n^{1/4}$ and $n^{3/4}$) and limit expressions involving the local time of Brownian motion at the origin. Those results turn out to apply only when the threshold parameters are unscaled or zero, and in these cases the results correspond to those in the binary choice model considered in Park and Phillips (2000, hereafter PP). When the threshold parameters are nonzero and are scaled to have the same order of magnitude as the covariates (i.e., by $\sqrt{n}$ for integrated regressors), a single convergence rate of $n^{3/4}$ applies to both parameters and thresholds and the limit theory involves expressions with local time evaluated at the thresholds rather than the origin. The limit theory for the parameter estimates is still mixed normal and usual procedures for statistical inference remain valid, as do the expressions for the arc sine laws and extended arc sine laws given in PP and HP.

As discussed in Hu and Phillips (2004b, hereafter HP_2), practical empirical work on ordered discrete choice models frequently involves explanatory variables that display random wandering characteristics. For instance, HP_2 construct a discrete choice model of the empirical behavior of the Federal Reserve in making discrete adjustments to the federal funds target rate, where the explanatory variables involve economic fundamentals monitored by the Fed such as the inflation rate and unemployment as well as leading indicators like consumer and business confidence. In modeling such intervention decisions where some of the explanatory variables behave like stochastic trends, it seems appropriate for the thresholds in the decision choices to be scaled to have the same order as the regressors so that there are nontrivial effects. This scaling is a theoretical device for developing a more meaningful asymptotic theory. Otherwise, the limit distribution will be degenerate and trivial. When the latent variable y_t^* in the choice model is nonstationary and converges to a continuous stochastic process like Brownian motion after scaling by $\sqrt{n}$, the choices ultimately depend on the behavior of the limiting stochastic process. For example, the observed dependent variable y_t may take on a discrete value such as unity (corresponding to a certain choice) when y_t^* falls in the interval between the scaled thresholds $\sqrt{n}\mu_0^1$ and $\sqrt{n}\mu_0^2$, and for such realizations the limit Brownian motion lies in the interval between μ_0^1 and μ_0^2, and the associated probability will be nonzero when $\mu_0^1 \neq \mu_0^2$. However, if the thresholds were unscaled, the limiting probability of y_t^* falling in the fixed interval between μ_0^1 and μ_0^2 would be zero (since $\mu_0^1/n^{1/2}, \mu_0^2/n^{1/2} \to 0$) and would therefore be trivial. The thresholds could, in fact, be determined by other variables, although this is not explored in HP or the present paper.

In the development that follows, we use the same model and notation as HP. In the interests of brevity, the set-up of HP will not be repeated in detail here and this paper provides a revised version of Lemma 1 and Theorem 1 of HP (given in Lemma R1 and Theorem R2) and the results which depend on them. We also need some supplementary results on convergence to functionals of Brownian local time at spatial points away from the origin, which are of independent interest. These are provided, together with proofs of the main results, in Appendices I and II. Readers are referred to the full length version of this paper (Phillips et al., 2005, hereafter PJH) available on the authors' websites for complete details. Further empirical illustrations of Brownian local time are given in Phillips (1998/2005, 2001).

The results of some simulation experiments are summarized, again with details in PJH. These reveal that the finite sample distributions of the regression coefficient and threshold estimates are generally well approximated by the mixture normal limit theory. A new finding is that predicted probabilities and marginal effect estimates have finite sample distributions in which the density increases towards the limits of the domain of definition. This pile-up problem is shown to occur also in the stationary discrete choice model.

2. Revised notation and assumptions

The set-up here follows HP and PP with some differences and extensions. In particular, we consider the regression model given by

$$y_t^* = x_t'\beta_0 - \varepsilon_t \quad \text{for } t = 1, \ldots, n, \tag{1}$$

where x_t is a $(m \times 1)$ vector of explanatory variables and ε_t is an error with cdf F. The dependent variable y_t^* is unobserved. Instead, what is observed is the indicator y_t, which takes the following possible $(J+1)$ values:

$$\begin{aligned} y_t &= 0 \quad \text{if } y_t^* \in (-\infty, \sqrt{n}\mu_0^1] \\ &= 1 \quad \text{if } y_t^* \in (\sqrt{n}\mu_0^1, \sqrt{n}\mu_0^2] \\ &\;\;\vdots \\ &= J-1 \quad \text{if } y_t^* \in (\sqrt{n}\mu_0^{J-1}, \sqrt{n}\mu_0^J] \\ &= J \text{ if } \quad y_t^* \in (\sqrt{n}\mu_0^J, \infty). \end{aligned} \tag{2}$$

In (1) x_t is predetermined and is an integrated process satisfying Assumption 1 of HP and for which $n^{-1/2}x_{[n\cdot]} \Rightarrow V(\cdot)$, Brownian motion with variance matrix Σ. The conditions are also sufficient to ensure that Skorohod embedding arguments may be used. The parameters are assembled in the vector θ, whose true value $\theta_0 = (\beta_0', \mu_0')'$ is an interior point of a subset of R^{m+J} which is compact and convex. As in HP, the regressor space is rotated using an orthogonal matrix $H = (h_1, H_2)$ with $h_1 = \beta_0/(\beta_0'\beta_0)^{1/2}$ to isolate the effects of the nonlinearities. The process V is correspondingly transformed as $V_1 = h_1'V$, $V_2 = H_2'V$, $L_{V_1}(t,s)$ is the local time of V_1 at the spatial point s over the time interval $[0,t]$, and $L_1(t,s) = (1/\sigma_{11})L_{V_1}(t,s)$, where σ_{11} is the variance of V_1. Under rotation by H, (1) becomes

$$y_t^* = x_t'\beta_0 - \varepsilon_t = x_t'HH'\beta_0 - \varepsilon_t = x_{1t}\alpha_0^1 + x_{2t}'\alpha_0^2 - \varepsilon_t,$$

where $x_{1t} = h_1'x_t$, $x_{2t} = H_2'x_t$, $\alpha_0^1 = h_1'\beta_0 = (\beta_0'\beta_0)^{1/2}$, $\alpha_0^2 = H_2'\beta_0 = 0$, and $\alpha_0 = H'\beta_0$ with $\alpha_0 = (\alpha_0^1, \alpha_0^{2'})'$. Denote $\underline{\theta}_0 = (\alpha_0', \mu_0')'$. The conditional probabilities of y_t are written as $\mathrm{P}(y_t = j|\mathscr{F}_{t-1}) = P_j(x_t; \theta_0)$.

The log likelihood function is $\log L_n(\theta) = \sum_{t=1}^n \sum_{j=0}^J \Lambda(t,j) \log P_j(x_t;\theta)$, where $\Lambda(t,j) = 1\{y_t = j\}$, and the score function is $S_n(\theta) = (S_n(\beta)', S_n(\mu)')' = (\partial \log L_n/\partial\beta', \partial \log L_n/\partial\mu')'$ with elements

$$\frac{\partial \log L_n}{\partial \beta} = \sum_{t=1}^n \sum_{j=0}^J \frac{\Lambda(t,j)}{P_j(x_t;\theta)} p_j(x_t;\theta)x_t, \tag{3}$$

$$\frac{\partial \log L_n}{\partial \mu^j} = \sqrt{n}\sum_{t=1}^{n}\left(\frac{\Lambda(t,j-1)}{P_{j-1}(x_t;\theta)} - \frac{\Lambda(t,j)}{P_j(x_t;\theta)}\right) f(x_t'\beta - \sqrt{n}\mu^j), \tag{4}$$

where

$$p_0(x_t;\theta) = -f(x_t'\beta - \sqrt{n}\mu^1),$$
$$p_j(x_t;\theta) = f(x_t'\beta - \sqrt{n}\mu^j) - f(x_t'\beta - \sqrt{n}\mu^{j+1}) \quad \text{for } j = 1,\ldots,J-1,$$
$$p_J(x_t;\theta) = f(x_t'\beta - \sqrt{n}\mu^J).$$

The first and second derivatives of F are written as f and $\dot{f}$.

The following assumption about the distribution function F and density f of ε_t extends Assumption 2 of HP by placing some additional explicit component functions in the classes and placing uniform tail conditions on F and f. Both probit and logit functions satisfy conditions (a)–(c) of Assumption R2 (as discussed in PP and HP) and (5), as is easily verified. As in HP, we use the following classifications for nonlinear functions: $g : \mathbf{R} \to \mathbf{R}$ is *regular* if it is bounded, integrable, and differentiable with bounded derivative; $\mathbf{F}_\mathrm{R}$ denotes the class of regular functions; $\mathbf{F}_\mathrm{I}$ is the class of bounded and integrable functions; and $\mathbf{F}_0$ the class of functions that are bounded and vanish at infinity. The notation $\dot{g}$ and $\ddot{g}$ is used to denote the first and second derivatives of g. The definitions for $\eta_{kl}, A_k, B_l, C_k, \tau_{klpq}$ are given in PJH.

Assumption R2. (Updates Assumption 2 of HP). *F* is three times differentiable with bounded derivatives and satisfies

$$\sup_{|x|<M}\frac{F(x - M^{1+\eta}\mu)}{F(x)} = \mathrm{o}(1), \quad \sup_{|x|\leqslant M}\frac{1 - F(x + M^{1+\eta}\mu))}{1 - F(x)} = \mathrm{o}(1),$$
$$\sup_{|x|<M}\frac{f(x \pm M^{1+\eta}\mu)}{f(x)} = \mathrm{o}(1), \tag{5}$$

as $M \to \infty$ for any $\eta, \mu > 0$. Further, for $k, l = 1,\ldots,J$:

(a) $\eta_{kl}A_kB_l, \eta_{kl}A_kA_l, \eta_{kl}B_kB_l \in F_\mathrm{R}$;
(b) $\eta_{kk}A_k, \eta_{kk}B_k, (\eta_{kl}\dot{A}_kB_l), (\eta_{kl}\dot{A}_kA_l), (\eta_{kl}\dot{B}_kB_l), \eta_{kk}^{1/2}\dot{C}_k \in F_\mathrm{I}$;
(c) $\tau_{klpq}A_kA_lA_pA_q, \tau_{klpq}A_kA_lB_pB_q, \tau_{klpq}B_kB_lB_pB_q, C_kC_l\eta_{kl} \in F_0$.

3. Correction to Lemma 1 of HP

Lemma R0 gives some limit results for partial sum expressions that are needed in analyzing the asymptotic behavior of the score and hessian functions. Lemma R1 corrects Lemma 1 of HP. Proofs and complementary results are given in the Appendix and PJH.

Lemma R0. *Let f and P be the density and probability distribution functions defined above, let Assumption* 1 *in HP and Assumption* R2 *hold, and let* $\mu_0^j \neq 0$ *and* $\kappa_1 \geqslant 0$. *Then, as* $n \to \infty$,

(a) $$\frac{1}{n^{1/2(1+\kappa_1)}}\sum_{t=1}^{n}\frac{f^2(x_{1t}\alpha_0^1 - \sqrt{n}\mu_0^j)}{P_j}x_{1t}^{\kappa_1} \Rightarrow \frac{(\mu_0^j)^{\kappa_1}}{(\alpha_0^1)^{\kappa_1+1}}L_1\left(1, \frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty}\frac{f^2(s)}{F(s)}\,\mathrm{d}s,$$

(b) $$\frac{1}{n^{1/2(1+\kappa_1)}}\sum_{t=1}^{n}\frac{f^2(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^j)}{P_{j-1}}x_{1t}^{\kappa_1}\Rightarrow\frac{(\mu_0^j)^{\kappa_1}}{(\alpha_0^1)^{\kappa_1+1}}L_1\left(1,\frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty}\frac{f^2(s)}{1-F(s)}\,\mathrm{d}s,$$

(c) $$\frac{1}{n^{3/2}}\sum_{t=1}^{n}\frac{f^2(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^j)}{P_j}x_{1t}x_{2t}\Rightarrow\frac{\mu_0^j}{(\alpha_0^1)^2}\int_0^1 V_2(r)\,\mathrm{d}L_1\left(r,\frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty}\frac{f^2(s)}{F(s)}\,\mathrm{d}s,$$

(d) $$\frac{1}{n^{3/2}}\sum_{t=1}^{n}\frac{f^2(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^j)}{P_{j-1}}x_{1t}x_{2t}\Rightarrow\frac{\mu_0^j}{(\alpha_0^1)^2}\int_0^1 V_2(r)\,\mathrm{d}L_1\left(r,\frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty}\frac{f^2(s)}{1-F(s)}\,\mathrm{d}s,$$

(e) $$\frac{1}{n^{3/2}}\sum_{t=1}^{n}\frac{f^2(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^j)}{P_j}x_{2t}x_{2t}'\Rightarrow\frac{1}{\alpha_0^1}\int_0^1 V_2(r)V_2(r)'\,\mathrm{d}L_1\left(r,\frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty}\frac{f^2(s)}{F(s)}\,\mathrm{d}s,$$

(f) $$\frac{1}{n^{3/2}}\sum_{t=1}^{n}\frac{f^2(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^j)}{P_{j-1}}x_{2t}x_{2t}'\Rightarrow\frac{1}{\alpha_0^1}\int_0^1 V_2(r)V_2(r)'\,\mathrm{d}L_1\left(r,\frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty}\frac{f^2(s)}{1-F(s)}\,\mathrm{d}s,$$

(g) $$n^{-1/2}\sum_{t=1}^{n}\frac{f(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^j)f(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^{j-1})}{P_{j-1}}\rightarrow_{\mathrm{p}}0,$$

(h) $$n^{-1/2}\sum_{t=1}^{n}\frac{f(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^j)f(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^{j+1})}{P_j}\rightarrow_{\mathrm{p}}0.$$

Remark. In a similar fashion to part (a) when $\kappa_1=2$ (as occurs in the hessian expression considered below), we obtain the limit

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n}f(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^j)x_{1t}^2\Rightarrow\frac{(\mu_0^j)^2}{(\alpha_0^1)^3}L_1\left(1,\frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty}f(s)\,\mathrm{d}s, \tag{6}$$

whereas, when $\mu_0^j=0$, we have (e.g. from Lemma 2 part (a) of PP)

$$\frac{1}{n^{1/2}}\sum_{t=1}^{n}f(x_{1t}\alpha_0^1)x_{1t}^2\Rightarrow\frac{1}{(\alpha_0^1)^3}L_1(1,0)\int_{-\infty}^{\infty}f(s)s^2\,\mathrm{d}s. \tag{7}$$

Thus, a major effect of the nonzero threshold $\mu_0^j\neq 0$ is to change the rate of convergence (or standardization) from $1/\sqrt{n}$ in (7) to $1/n^{3/2}$. Another effect is that the limit random variable involves Brownian local time at μ_0^j/α_0^1 instead of the origin. Finally, the scale effect arising from the spatial integral changes from $\int_{-\infty}^{\infty}f(s)s^2\,\mathrm{d}s$ in (7) to $\mu_0^2\int_{-\infty}^{\infty}f(s)\,\mathrm{d}s$ in (6). Each of these effects arises from the fact that the principal contribution to the partial sum comes when x_{1t} is around $\sqrt{n}\mu_0^j/\alpha_0^1$. These are the changes in the limit theory for the nonzero threshold case that lead to the corrections needed for HP.

Lemma R1. (Corrects Lemma 1 of HP). *Let Assumption* 1 *in HP hold, and write* $A_k(x_{1t};\underline{\theta}_0)=A_k$, $B_k(x_t;j,\underline{\theta}_0)=B_k$. *Assume for* $k,l,=1,\ldots,J$, *that* $A_kA_l\eta_{kl}$, $A_kB_l\eta_{kl}$,

$B_kB_l\eta_{kl} \in F_{\mathrm{R}}$, $A_k\eta_{kk}$, $B_k\eta_{kk} \in F_{\mathrm{I}}$, and $\tau_{kkkk}A_k^4$, $\tau_{kkkk}B_k^4 \in F_0$ for A_k, $B_k : R \to R$. Then

$$\begin{pmatrix} n^{-3/4}\sum_{t=1}^{n}\sum_{k=1}^{J} A_k z_{kt} x_{1t} \\ n^{-3/4}\sum_{t=1}^{n}\sum_{k=1}^{J} A_k z_{kt} x_{2t} \\ n^{-1/4}\sum_{t=1}^{n}\sum_{k=1}^{J} B_k z_{kt} \end{pmatrix} \Rightarrow M^{1/2}W(1), \tag{8}$$

where $M = ([M_{ij}])$ is partitioned conformably with component submatrices

$$M_{11} = \sum_{j=1}^{J}\left\{\frac{(\mu_0^j)^2}{(\alpha_0^1)^3} L_1\left(1, \frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty}\frac{f^2(s)}{F(s)(1-F(s))}\,\mathrm{d}s\right\}, \tag{9}$$

$$M_{12} = \sum_{j=1}^{J}\left\{\frac{\mu_0^j}{(\alpha_0^1)^2}\int_0^1 V_2(r)\,\mathrm{d}L_1\left(r, \frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty}\frac{f^2(s)}{F(s)(1-F(s))}\,\mathrm{d}s\right\}, \tag{10}$$

$$M_{22} = \sum_{j=1}^{J}\left\{\frac{1}{\alpha_0^1}\int_0^1 V_2(r)V_2(r)'\mathrm{d}L_1\left(r, \frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty}\frac{f^2(s)}{F(s)(1-F(s))}\,\mathrm{d}s\right\}, \tag{11}$$

$$M_{13} = \frac{\mu_0^j}{(\alpha_0^1)^2} L_1\left(1, \frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty}\frac{f^2(s)}{F(s)(1-F(s))}\,\mathrm{d}s, \tag{12}$$

$$M_{23} = \frac{1}{\alpha_0^1}\int_0^1 \mathrm{d}L_1\left(r, \frac{\mu_0^j}{\alpha_0^1}\right)V_2(r)'\int_{-\infty}^{\infty}\frac{f^2(s)}{F(s)(1-F(s))}\,\mathrm{d}s, \tag{13}$$

$$M_{33} = \frac{1}{\alpha_0^1} L_1\left(1, \frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty}\frac{f^2(s)}{F(s)(1-F(s))}\,\mathrm{d}s, \tag{14}$$

and W is m-dimensional Brownian motion with covariance matrix I, which is independent of V.

Remarks. (1) The main correction that Lemma R1 makes to Lemma 1 of HP is to include the component $n^{-3/4}\sum_{t=1}^{n}\sum_{k=1}^{J} A_k z_{kt} x_{1t}$, which has the same rate of convergence ($n^{-3/4}$) as the element $n^{-3/4}\sum_{t=1}^{n}\sum_{k=1}^{J} A_k z_{kt} x_{2t}$ involving the factor x_{2t}. The corrections, notably that the limit functional involves Brownian local time at spatial points $\{\mu_0^j/\alpha_0^1 : j = 1, \ldots, J\}$ away from the origin, are discussed in the Remark above.

(2) It is pointed out in PP that if x_{2t} were replaced by a stationary variate (as it would in some directions were x_{2t} to be cointegrated), then the norming would be different. Thus, suppose x_{3t} is a stationary ($m_3 \times 1$) vector with coefficient γ_0, satisfies the same conditions as v_t in Assumption 1 of HP and is independent of u_t. Then we have:

$$\frac{1}{n^{1/2}}\sum_{t=1}^{n} f(x_{3t}'\gamma_0 + x_{1t}\alpha_0^1 - \sqrt{n}\mu_0^j)x_{3t}x_{3t}' \Rightarrow \frac{1}{\alpha_0^1} L_1\left(1, \frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty} f(s)\,\mathrm{d}s\Sigma_{33},$$

where $\Sigma_{33} = \mathrm{E}(x_{3t}x_{3t}')$, and

$$\frac{1}{n^{1/4}}\sum_{t=1}^{n}\sum_{k=1}^{J} A_k z_{kt} x_{3t} \Rightarrow MN\left(0, \sum_{j=1}^{J}\left\{\frac{1}{\alpha_0^1} L_1\left(1, \frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty}\frac{f^2(s)}{F(s)(1-F(s))}\,\mathrm{d}s\right\}\Sigma_{33}\right).$$

4. Correction to the main results

The maximum likelihood estimator $\widehat{\theta}_n = (\widehat{\beta}_n', \widehat{\mu}_n')'$ of $\theta_0 = (\beta_0', \mu_0')'$ satisfies the expansion

$$0 = S_n(\widehat{\theta}_n) = S_n(\theta_0) + J_n(\widetilde{\theta})(\widehat{\theta}_n - \theta_0), \tag{15}$$

where $\widetilde{\theta}$ is on the line segment between $\widehat{\theta}_n$ and θ_0, which differs from row to row of the hessian matrix $J_n(\widetilde{\theta})$. Corresponding to the rotation in the regressor space, define

$$G = \begin{pmatrix} H & 0 \\ 0 & I_J \end{pmatrix},$$

and let $\underline{\theta} = (\alpha', \mu')'$. Then, the score function and hessian matrix for the new parameters are based on $S_n(\underline{\theta}) = G' S_n(\theta)$ and $J_n(\underline{\theta}) = G' J_n(\theta) G$, and

$$0 = S_n(\hat{\underline{\theta}}_n) = S_n(\underline{\theta}_0) + J_n(\tilde{\underline{\theta}}_n)(\hat{\underline{\theta}}_n - \underline{\theta}_0). \tag{16}$$

Using Lemma R1, we obtain the following limit theory for the score function and the hessian, which corrects Theorem 1 of HP.

Theorem R2. *Let Assumption* 1 *in HP and Assumption* R2 *hold. Then*

$$n^{-3/4} S_n(\underline{\theta}_0) \Rightarrow Q^{1/2} W(1) \quad \text{and} \quad n^{-3/2} J_n(\underline{\theta}_0) \Rightarrow -Q$$

jointly, where Q is the symmetric matrix partitioned as

$$Q = \begin{pmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{pmatrix} \tag{17}$$

conformably with $(\alpha_0^1, \alpha_0^{2'}, \mu_0')'$, *and where*

$$q_{11} = \sum_{j=1}^{J} \left\{ \frac{(\mu_0^j)^2}{(\alpha_0^1)^3} L_1\left(1, \frac{\mu_0^j}{\alpha_0^1}\right) \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1-F(s))}\,\mathrm{d}s \right\},$$

$$q_{12} = \sum_{j=1}^{J} \left\{ \frac{\mu_0^j}{(\alpha_0^1)^2} \int_0^1 \mathrm{d}L_1\left(r, \frac{\mu_0^j}{\alpha_0^1}\right) V_2(r)' \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1-F(s))}\,\mathrm{d}s \right\},$$

$$q_{13}(j) = \frac{\mu_0^j}{(\alpha_0^1)^2} L_1\left(1, \frac{\mu_0^j}{\alpha_0^1}\right) \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1-F(s))}\,\mathrm{d}s,$$

$$q_{22} = \sum_{j=1}^{J} \left\{ \frac{1}{\alpha_0^1} \int_0^1 V_2(r) V_2(r)' \mathrm{d}L_1\left(r, \frac{\mu_0^j}{\alpha_0^1}\right) \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1-F(s))}\,\mathrm{d}s \right\},$$

$$q_{23}(j) = \frac{1}{\alpha_0^1} \int_0^1 \mathrm{d}L_1\left(r, \frac{\mu_0^j}{\alpha_0^1}\right) V_2(r)' \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1-F(s))}\,\mathrm{d}s,$$

$$q_{33}(j,j) = \frac{1}{\alpha_0^1} L_1\left(1, \frac{\mu_0^j}{\alpha_0^1}\right) \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1-F(s))}\,\mathrm{d}s,$$

$$q_{33}(j,i) = 0 \quad \text{for } i \neq j.$$

and W is defined as in Lemma R1.

Remarks. (1) Notice that with threshold parameters in the model, even if ε_t has a symmetric distribution, as in the probit and logit models, q_{12}, q_{13}, q_{21} and q_{31} are not zero and Q does not reduce to a block diagonal matrix, which differs from the result in PP.

(2) When stationary m_3-dimensional variables x_{3t} are present in the model, we get multiple convergence rates. Suppose x_{3t} is an m_3-vector of zero mean, stationary time series with coefficient γ_0 defined as above. Let $\rho = (\gamma', \theta')'$, $\underline{\rho} = (\gamma', \underline{\theta}')'$, and

$$G_2 = \begin{pmatrix} I_{m_3} & 0 & 0 \\ 0 & H & 0 \\ 0 & 0 & I_J \end{pmatrix},$$

$$D_n = Diag(n^{1/4} I_{m_3}, n^{3/4} I_{m+J}).$$

Following similar steps as those in the proof of Theorem R2, and using Remark 2 after Lemma R1, we obtain the following limit theory:

$$D_n S_n(\underline{\rho}_0) \Rightarrow \Xi^{1/2} W(1) \quad \text{and} \quad D_n^{-1} J_n(\underline{\rho}_0) D_n^{-1} \Rightarrow -\Xi,$$

where

$$\Xi = \begin{pmatrix} \Xi_{11} & 0 \\ 0 & Q \end{pmatrix},$$

with

$$\Xi_{11} = \sum_{j=1}^{J} \left\{ \frac{1}{\alpha_0^1} L_1 \left(1, \frac{\mu_0^j}{\alpha_0^1} \right) \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1 - F(s))} \, \mathrm{d}s \right\} \Sigma_{33},$$

and Q is defined as in Theorem R2, and $\Sigma_{33} = \mathrm{E}(x_{3t} x_{3t}')$.

The asymptotic results for $S_n(\underline{\theta}_0)$ and $J_n(\underline{\theta}_0)$ in Theorem R2 lead to the limit distribution of $\widehat{\underline{\theta}}_n$. From expansion (16), the normed and centered estimator satisfies

$$n^{3/4}(\widehat{\underline{\theta}}_n - \underline{\theta}_0) = -(n^{-3/2} J_n(\underline{\theta}_0))^{-1} n^{-3/4} S_n(\underline{\theta}_0) + o_p(1), \tag{18}$$

a result that is established in the proof of Theorem R3, which corrects Theorem 2 of HP.

Theorem R3. *Let Assumption* 1 *in HP and Assumption* R2 *hold. Then there exists a sequence of ML estimators for which as* $\widehat{\underline{\theta}}_n \to_p \underline{\theta}_0$, *and* $n^{3/4}(\widehat{\underline{\theta}}_n - \underline{\theta}_0) \Rightarrow Q^{-1/2} W(1)$, *in the notation introduced in Theorem* R2.

Remarks. (1) From the above, we get

$$n^{3/4} G'(\widehat{\theta}_n - \theta_0) \Rightarrow Q^{-1/2} W(1), \tag{19}$$

and therefore $n^{3/4}(\widehat{\theta}_n - \theta_0) \Rightarrow G Q^{-1/2} W(1) = MN(0, G Q^{-1} G')$,

(2) Following arguments similar to those in Theorem 3 and using Remark 2, when there are stationary variables in the model, we have

$$D_n(\widehat{\underline{\rho}}_n - \underline{\rho}_0) \Rightarrow \Xi^{-1/2} W(1)$$

and

$$D_n G_2'(\widehat{\rho}_n - \rho_0) \Rightarrow \Xi^{-1/2} W(1)$$

or

$$n^{1/4}(\widehat{\gamma}_n - \gamma_0) \Rightarrow \Xi_{11}^{-1/2} W(1),$$
$$n^{3/4} G'(\widehat{\theta}_n - \theta_0) \Rightarrow Q^{-1/2} W(1).$$

Thus

$$n^{3/4}(\widehat{\theta}_n - \theta_0) \Rightarrow GQ^{-1/2} W(1) = MN(0, GQ^{-1}G'),$$
$$n^{1/4}(\widehat{\gamma}_n - \gamma_0) \Rightarrow \Xi_{11}^{-1/2} W(1),$$

which we formalize in the Corollary that follows, which replaces Corollary 1 of HP.

Corollary R4. *Under Assumption* 1 *in HP and Assumption* R2, *as* $n \to \infty$

$$\begin{pmatrix} n^{3/4}(\widehat{\beta}_n - \beta_0) \\ n^{3/4}(\widehat{\mu}_n - \mu_0) \end{pmatrix} \Rightarrow MN(0, GQ^{-1}G'). \tag{20}$$

When there are stationary variables in the system with coefficients γ_0, $n^{1/4}(\widehat{\gamma}_n - \gamma_0) \Rightarrow MN(0, \Xi_{11}^{-1})$ and is independent of (20), so that

$$\begin{pmatrix} n^{1/4}(\widehat{\gamma}_n - \gamma_0) \\ n^{3/4}(\widehat{\beta}_n - \beta_0) \\ n^{3/4}(\widehat{\mu}_n - \mu_0) \end{pmatrix} \Rightarrow MN(0, G_2 \Xi^{-1} G_2'),$$

in which case the convergence rates for the parameter estimates differ, with a slower $n^{1/4}$ rate for the parameters of stationary variables, and a faster $n^{3/4}$ convergence rate for the other parameter estimates.

The conditional covariance matrix of $\hat{\theta}_n$ can be estimated by the hessian inverse $-J_n(\hat{\theta}_n)^{-1}$, or the more commonly used alternative $-\underline{J}_n(\hat{\theta}_n)^{-1}$, where $\underline{J}_{n,ij}$ excludes the terms in $J_{n,ij}$ that involve martingale differences (see HP and PJH for details). The following result replaces Theorem 3 of HP.

Theorem R5. *Under Assumption* 1 *in HP and Assumption* R2, $-[n^{-3/2} J_n(\widehat{\theta}_n)]^{-1} \Rightarrow GQ^{-1}G'$ *as* $n \to \infty$, *with the same limit holding for* $-[n^{-3/2}\underline{J}_n(\widehat{\theta}_n)]^{-1}$.

Again, when we have stationary variables, $-[n^{-1/2} J_n(\widehat{\gamma}_n)]^{-1} \Rightarrow \Xi_{11}^{-1}$, and $-[n^{-3/2} J_n(\widehat{\theta}_n)]^{-1} \Rightarrow GQ^{-1}G'$ as $n \to \infty$.

5. Predicted probabilities and marginal effects

5.1. Predicted probability

Next consider $\hat{P}_{j,x} = \hat{P}_j(x_t; \hat{\theta}_n)$, the predicted probability of the choice $y_t = j$, and $\widehat{\upsilon}_{j,x} = \hat{p}_j(x_t; \hat{\theta}_n)\hat{\beta}_n$, the estimated marginal effect of x_t on $\hat{P}_j(x_t; \hat{\theta}_n)$ both evaluated for some $x_t = x$. To achieve comparability between $x'\beta_0$ and the thresholds, and thereby assist in simulating the finite sample and asymptotic distributions of the predicted probabilities, we write the scaled thresholds in the comparable form $z_n \mu_0^j$ (in place of $\sqrt{n}\mu_0^j$) and suppose $z_n > 0$ is a realization of some (independent) unit root time series so that $z_n = O_p(\sqrt{n})$, and the ordering on the thresholds is positively scaled and therefore not reversed. This scaling is analogous to the $\sqrt{n}\mu_0^j$ scaling of the thresholds used in previous sections and serves as a device for developing the asymptotic theory in a convenient way. The probabilities P_j are then evaluated at $x_t = x$ and $z_n = z$ for some specific values x and z. The probabilities satisfy

$$P_0(x_t;\theta_0) = 1 - F(x'\beta_0 - z\mu_0^1),$$
$$P_j(x_t;\theta_0) = F(x'\beta_0 - z\mu_0^j) - F(x'\beta_0 - z\mu_0^{j+1}) \quad \text{for } j = 1,\ldots,J-1,$$
$$P_J(x_t;\theta_0) = F(x'\beta_0 - z\mu_0^J).$$

To analyze these quantities, we define a matrix $R(0) = \text{Diag}(I_m, \iota_1')$ where ι_j is a vector of length J with the jth element 1 and other elements zero. Similarly, $R(J) = \text{Diag}(I_m, \iota_J')$ and for $1 \leqslant j \leqslant J-1$, $R(j) = \text{Diag}(I_m, (\iota_j, \iota_{j+1})')$. Accordingly, we may write

$$\begin{pmatrix} \hat{\beta}_n - \beta_0 \\ \hat{\mu}_n^1 - \mu_0^1 \end{pmatrix} = R(0)\begin{pmatrix} \hat{\beta}_n - \beta_0 \\ \hat{\mu}_n - \mu_0 \end{pmatrix} \quad \begin{pmatrix} \hat{\beta}_n - \beta_0 \\ \hat{\mu}_n^J - \mu_0^J \end{pmatrix} = R(J)\begin{pmatrix} \hat{\beta}_n - \beta_0 \\ \hat{\mu}_n - \mu_0 \end{pmatrix},$$

and for $1 \leqslant j \leqslant J-1$,

$$\begin{pmatrix} \hat{\beta}_n - \beta_0 \\ \hat{\mu}_n^j - \mu_0^j \\ \hat{\mu}_n^{j+1} - \mu_0^{j+1} \end{pmatrix} = R(j)\begin{pmatrix} \hat{\beta}_n - \beta_0 \\ \hat{\mu}_n - \mu_0 \end{pmatrix}.$$

Corollary R6. *Let Assumption* 1 *in HP and Assumption* R2 *hold. Given* $x_t = x$, $z_n = z$, *for* $j = 0,\ldots,J$, *the predicted probabilities of* $y_t = j$ $(j = 0,\ldots,J)$ *satisfy*

$$n^{3/4}(\widehat{P}_{j,x} - P_{j,x}) \Rightarrow MN(0, \Upsilon(j)GQ^{-1}G'\Upsilon(j)').$$

The above expressions use the following notation:

$$P_{j,x} = P_j(x;\theta_0) \quad \textit{for } j = 0,1,\ldots,J,$$
$$\Upsilon(0) = f(x'\beta_0 - z\mu_0^1)\begin{pmatrix} -x \\ z \end{pmatrix}' R(0),$$
$$\Upsilon(J) = f(x'\beta_0 - z\mu_0^J)\begin{pmatrix} x \\ -z \end{pmatrix}' R(J),$$
$$\Upsilon(j) = \begin{pmatrix} [f(x'\beta_0 - z\mu_0^j) - f(x'\beta_0 - z\mu_0^{j+1})]x \\ -f(x'\beta_0 - z\mu_0^j)z \\ f(x'\beta_0 - z\mu_0^{j+1})z \end{pmatrix}' R(j) \quad \text{for } j = 1,\ldots,J-1.$$

When we have stationary variables, given $x_t = x$, $z_n = z$, $x_{3t} = x_3$ the limit theory becomes

$$n^{1/4}(\widehat{P}_{j,x} - P_{j,x}) \Rightarrow MN(0, f(x_3'\gamma_0 + x'\beta_0 - z\mu_0^j)^2 x_3'\Xi_{11}^{-1}x_3) \quad \text{for } j = 0, J,$$
$$n^{1/4}(\widehat{P}_{j,x} - P_{j,x}) \Rightarrow MN(0, [f(x'\beta_0 - z\mu_0^j) - f(x'\beta_0 - z\mu_0^{j+1})]^2 x_3'\Xi_{11}^{-1}x_3) \quad \text{for } j = J,\ldots,J-1.$$

Therefore, the limit theory when stationary variables are present is dominated by the stationary coefficients and the convergence rate is $n^{1/4}$, just as in PP.

5.2. Marginal effects

For the marginal effects, we have the following limit theory.

Corollary R7. *Let Assumption* 1 *in HP and Assumption* R2 *hold. Given* $x_t = x$, $z_n = z$, *for* $j = 0, \ldots, J$, *the estimated marginal effects* $\widehat{\upsilon}_{j,x}$ *have the following asymptotic distributions as* $n \to \infty$

$$n^{3/4}(\widehat{\upsilon}_{j,x} - \upsilon_{j,x}) \Rightarrow MN(0, \Pi(j) G Q^{-1} G' \Pi(j)').$$

These expressions use the notation:

$$\upsilon_{j,x} = \upsilon_j(x;\theta_0) = p_j(x;\theta_0)\beta_0 \quad \text{for } j = 0, 1, \ldots, J,$$

$$\Pi(0) = \begin{pmatrix} -\dot{f}((x'\beta_0 - z\mu_0^1))x\beta_0' - f((x'\beta_0 - z\mu_0^1))I_m \\ \dot{f}((x'\beta_0 - z\mu_0^1))z\beta_0 \end{pmatrix}' R(0),$$

$$\Pi(J) = \begin{pmatrix} \dot{f}((x'\beta_0 - z\mu_0^J))x\beta_0' + f((x'\beta_0 - z\mu_0^J))I_m \\ -\dot{f}((x'\beta_0 - z\mu_0^J))z\beta_0 \end{pmatrix}' R(J),$$

$$\Pi(j) = \begin{pmatrix} [\dot{f}((x'\beta_0 - z\mu_0^j)) - \dot{f}((x'\beta_0 - z\mu_0^{j+1}))]x\beta_0' + p_j(x;\theta_0)I_m \\ -\dot{f}((x'\beta_0 - z\mu_0^j))z\beta_0 \\ \dot{f}((x'\beta_0 - z\mu_0^{j+1}))z\beta_0 \end{pmatrix}' R(j)$$

for $j = 1, \ldots, J-1$.

When stationary variables are present, given $x_t = x$, $z_n = z, x_{3t} = x_3$, the estimated marginal effects $\widehat{\upsilon}_{j,x}$ have the following asymptotic distributions as $n \to \infty$

$$n^{1/4}(\widehat{\upsilon}_{j,x} - \upsilon_{j,x}) \Rightarrow MN(0, \Lambda(j)\Xi_{11}^{-1}\Lambda(j)'),$$

where

$$\Lambda(0) = -\dot{f}((x_3'\gamma + x'\beta_0 - z\mu_0^1))\rho x_3' - f((x_3'\gamma + x'\beta_0 - z\mu_0^1))I_{m_3},$$
$$\Lambda(j) = [\dot{f}((x_3'\gamma + x'\beta_0 - z\mu_0^j)) - \dot{f}((x_3'\gamma + x'\beta_0 - z\mu_0^{j+1}))]\rho x_3' + p_j(x, x_3;\rho_0)I_{m_3},$$
$$\Lambda(J) = \dot{f}((x_3'\gamma + x'\beta_0 - z\mu_0^J))\rho x_3' + f((x_3'\gamma + x'\beta_0 - z\mu_0^1))I_{m_3}.$$

Therefore, the limit theory when stationary variables are present is dominated by the stationary coefficients and the convergence rate is $n^{1/4}$, just as in PP.

6. Simulation experiments

Some extensive simulations were conducted to examine the finite sample performance of ML estimation, predicted probabilities, and marginal effects in a polychotomous choice model under nonstationarity. This section briefly summarizes some of the findings and readers are referred to PJH for details and further discussion.

The experimental design was based on a model with $m = 2$ explanatory variables and $J = 2$, giving a triple-choice dependent variable y_t. The DGP for the exogeneous data is the system

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} \rho_1 & 0 \\ 0 & \rho_2 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}, \tag{21}$$

with $v_t = (v_{1t}, v_{2t})' = iid\ \mathrm{N}(0, I_2)$, and $\rho_1 = \rho_2 = \rho = 1$. The coefficient parameter vector was set at $\beta_0 = (1, 0)'$, so $x_t'\beta_0 = \beta_0^1 x_{1t} = x_{1t}$ and the direction orthogonal to β_0 is $(0, 1)$, giving the coefficient $\beta_0^2 = 0$ of x_{2t}. This set-up is analogous to that of the simulation study of PP. The number of replications was 50 000, and sample sizes ranging from $n = 100$ to 1000 were used. The main conclusions are as follows:

1. As the magnitude of the threshold parameters increased (from 0.1 to 1.5), the convergence rates of the coefficient estimates in the two directions showed evidence of equalizing, thereby corroborating the limit theory of Theorem R3 in contrast to the zero threshold case of PP, where the convergence rates differ.
2. The distributions of both parameters and threshold estimates generally appear to approach symmetric distributions corresponding to the mixed normal limit theory. However, there is some evidence that, as the magnitude of the thresholds μ_0^j increase, the distributions of the estimates become biased. The reason for the bias appears to be related to the behavior of the choice probabilities in such cases, which quickly go to zero or unity when the arguments are large. This bias is also found to occur in the stationary case (for values of $\rho_i \geqslant 0.95$ in (21)) when the thresholds are large.
3. Fig. 1 shows kernel estimates of the sampling distributions of the (scaled and centered) choice probability when $j = 0$ for sample sizes $n = 100, 250, 500, 1000$. Different choices of μ_0, β_0, z, and x do not change the results in a material way provided the parameter

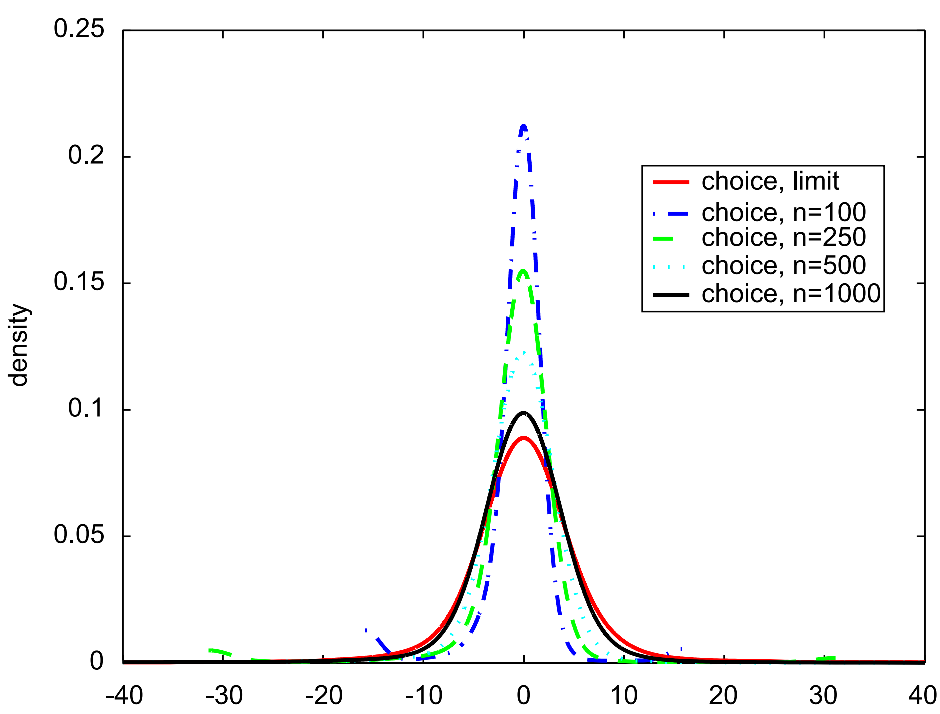

Fig. 1. Density of choice probability for $j = 0$.

settings are small, but when they are large the choice probabilities can quickly go to zero or unity and this appears to bias the distributions, as mentioned above.

4. The finite sample distribution of the choice probability has finite support and reveals a pile-up problem where the density increases towards the limits of the domain of definition, as is apparent in Fig. 1. This pile-up problem, which to our knowledge has not before been noticed in the discrete choice literature, also occurs in the stationary case—see Fig. 2, where $\rho_1 = \rho_2 = \rho = 0.95$, and Fig. 3 where $\rho_1 = \rho_2 = \rho = 0.99$, with sample sizes $n = 100, 500, 1000, 5000$. The figures show that as n passes to infinity the pile-up problem steadily dissipates. For $n = 5000$ the upper and lower bounds are close to the extremes of the support where the limit distribution is nonnegligible. Thus, the problem of pile-up is not confined to the nonstationary discrete choice problem but is a more generic problem. In effect, the asymptotic approximations (such as those given in Corollary R6) are valid in an immediate interval around the true values. Outside that interval, behavior is rather different because of the fact that $\widehat{P}_{0,x}$ goes to zero or unity depending on the sign of its argument, resulting in a pile-up of the distribution in finite samples. It might therefore be argued that the true finite sample distribution would be better approximated by a mixture of three distributions, one of which is the local asymptotic result given above and the other two are based on pile-ups around $\widehat{P}_{0,x}{\sim}0$, and $\widehat{P}_{0,x}{\sim}1$. Developing such a mixture approximation clearly involves further complications and is left for the future research.
5. Figs. 4 and 5, show kernel estimates of the sampling distributions of the marginal effects $\widehat{\upsilon}_{j,x} = \widehat{p}_j(x_t;\widehat{\theta}_n)\widehat{\beta}_n = \widehat{\upsilon}_{j,x} = \widehat{p}_j(x_t;\widehat{\theta}_n)(\widehat{\beta}^1_n, \widehat{\beta}^2_n)'$when $j = 0$ for sample sizes $n = 100, 250, 500,$ 1000. In the graphs, we use ME1 to denote $\widehat{p}_j(x_t;\widehat{\theta}_n)\widehat{\beta}^1_n$, and ME2 to denote $\widehat{p}_j(x_t;\widehat{\theta}_n)\widehat{\beta}^2_n$. The graphs show that in large samples the distributions of scaled marginal effects appear to approach the asymptotic distributions derived in the paper. Again, there

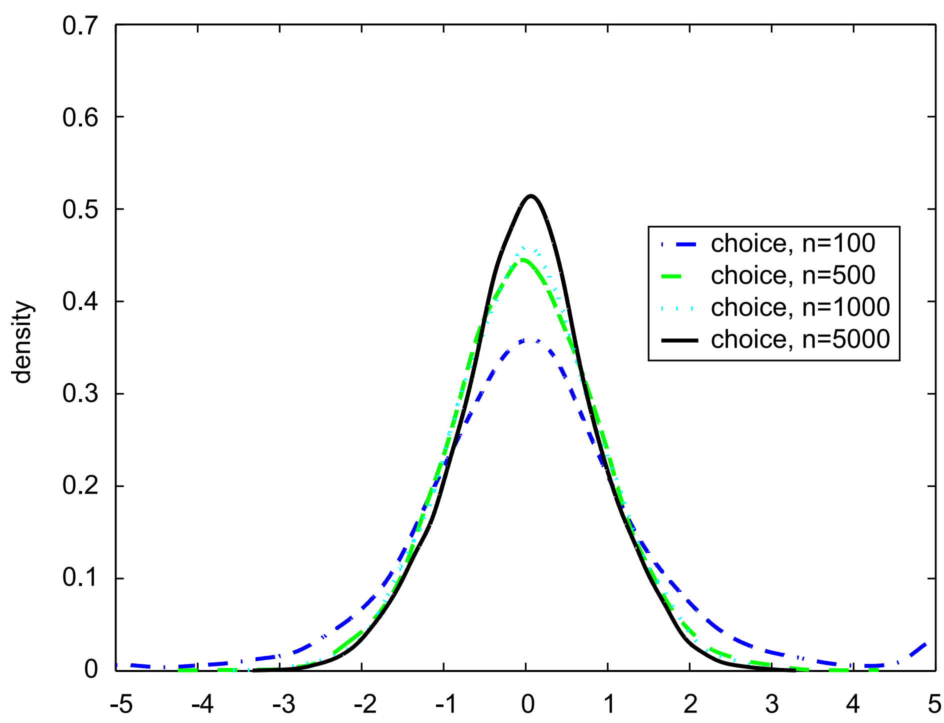

Fig. 2. Stationary models with $\rho = 0.95$.

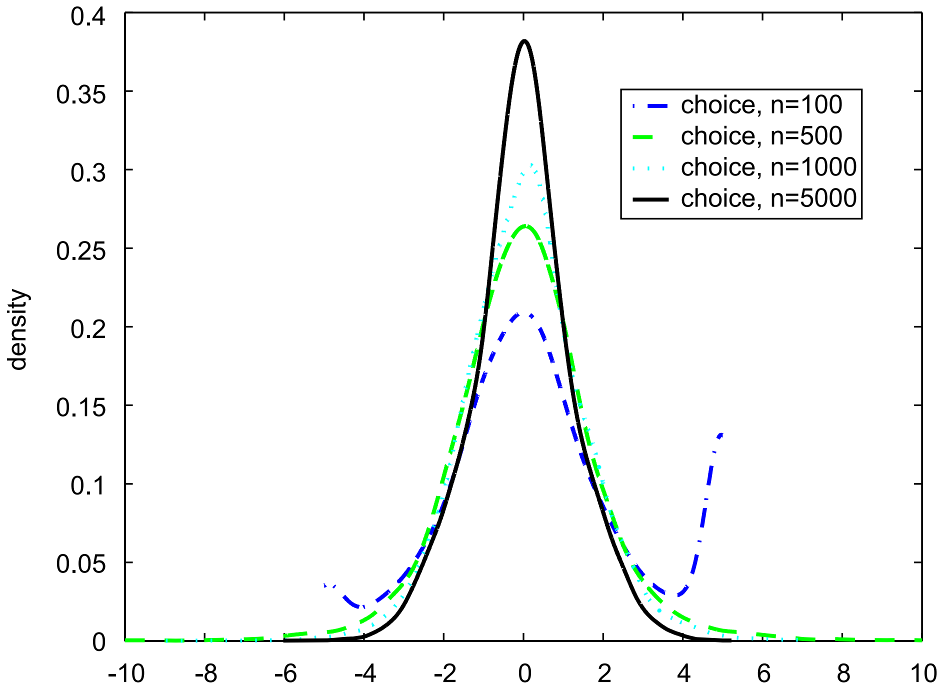

Fig. 3. Stationary models with $\rho = 0.99$.

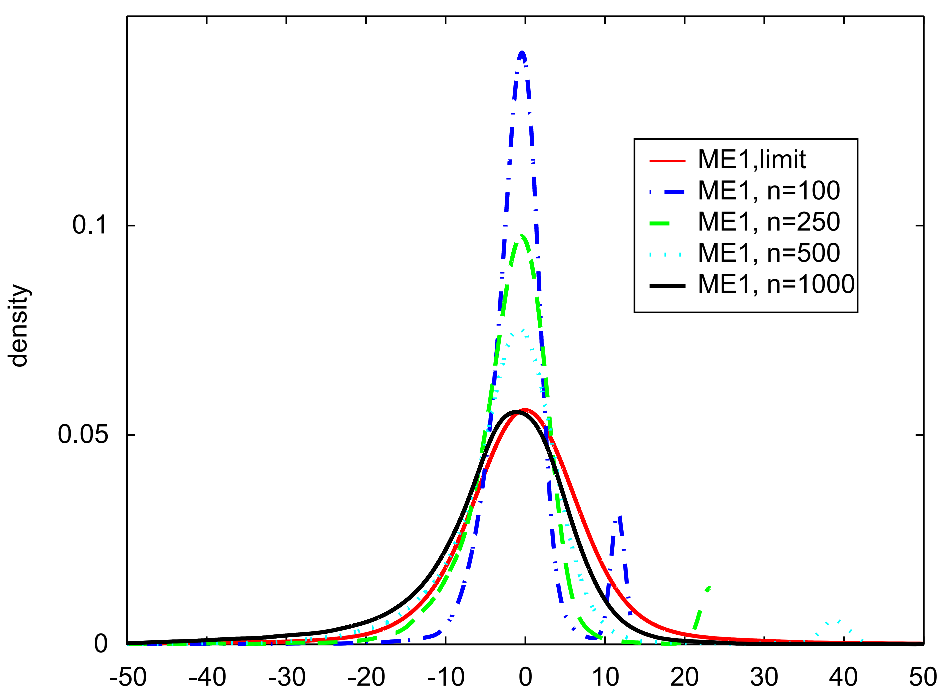

Fig. 4. Density of marginal effects when $j = 0$.

appears to be a pile-up problem towards the limits of the domain of definition particularly in the case of ME1. Investigation shows that this problem also occurs in the stationary case for large values of the autoregressive coefficient. As for the predicted probabilities, this phenomenon deserves further study.

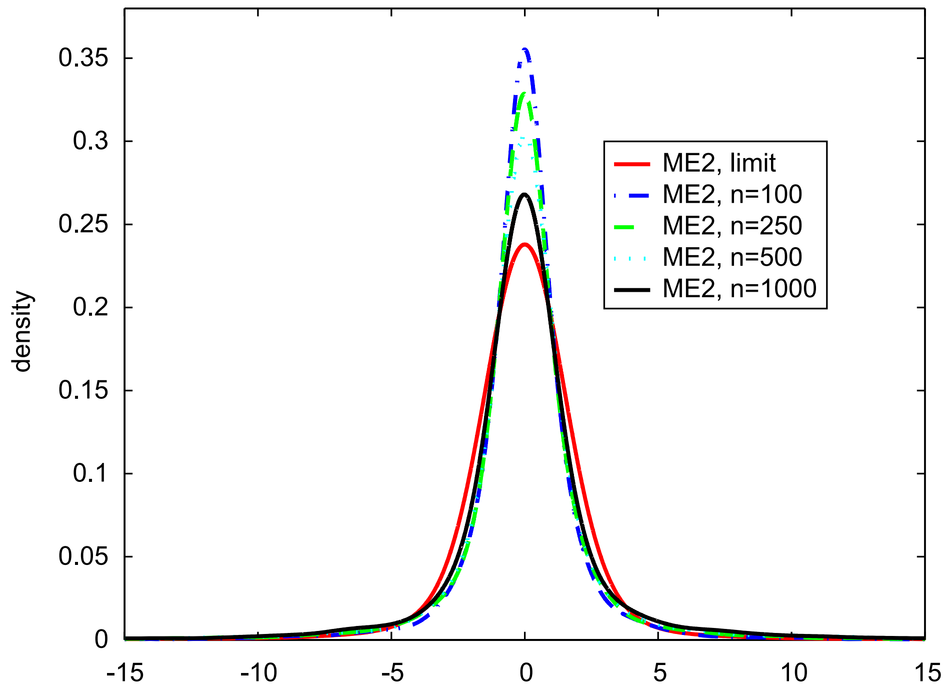

Fig. 5. Density of marginal effects when $j = 0$.

Acknowledgments

We thank the referees and Co-Editor for helpful comments and advice. A full length version of this paper (Phillips et al., 2005) is available on the authors' websites. Phillips gratefully acknowledges research support from a Kelly Fellowship at the School of Business, University of Auckland, and the NSF under Grant No. SES 04-142254. Jin thanks the Cowles Foundation for support under a Cowles Fellowship.

Appendix. Useful lemmas and proofs

The corresponding appendix of PJH contains some useful lemmas, proofs of those lemmas, and proofs of the main results in the paper, which update and extend those in HP, PP and Park and Phillips (1999, 2001). The updating takes into account the explicit form of the dependence of functions on the threshold. The reader is referred to PJH for details. We provide here only the statement of the key lemma that is used directly in the derivation of the main results, showing the effects of nonzero thresholds. Some related work, which gives a version of part (a) of the lemma, is contained in the recent paper by Jeganathan (2004).

Lemma E (Extends Lemma 2 of PP to local time away from the origin). *Let Assumption* 1 *in HP hold*, $f : R \to R$ *be regular*, *and* $\mu \neq 0$. *Then we have*:

(a) $(1/\sqrt{n})\sum_{t=1}^{n} f(x_{1t} - \sqrt{n}\mu) \Rightarrow L_1(1,\mu)\int_{-\infty}^{\infty} f(s)\,\mathrm{d}s$,
(b) $(1/n)\sum_{t=1}^{n} f(x_{1t} - \sqrt{n}\mu)x_{2t} \Rightarrow \int_0^1 V_2(r)\,\mathrm{d}L_1(r,\mu)\int_{-\infty}^{\infty} f(s)\,\mathrm{d}s$,
(c) $(1/n^{3/2})\sum_{t=1}^{n} f(x_{1t} - \sqrt{n}\mu)x_{2t}x_{2t}' \Rightarrow \int_0^1 V_2(r)V_2(r)'\mathrm{d}L_1(r,\mu)\int_{-\infty}^{\infty} f(s)\,\mathrm{d}s$.

References

Hu, L., Phillips, P.C.B., 2004a. Nonstationary discrete choice. Journal of Econometrics 120, 103–138.

Hu, L., Phillips, P.C.B., 2004b. Dynamics of the federal funds target rate: a nonstationary discrete choice approach. Journal of Applied Econometrics 19, 851–867.

Jeganathan, J., 2004. Convergence of functionals of sums of random variables to local times of fractional stable motions. Annals of Probability 32, 1771–1795.

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 regression with integrated processes. Econometrica 69, 117–161.

Phillips, P.C.B., 1998/2005. Econometric analysis of the Fisher equation. Cowles Foundation Discussion Paper No. 1180. Published in American Journal of Economics and Sociology 64(1).

Phillips, P.C.B., 2001. Descriptive econometrics for non-stationary time series with empirical applications. Journal of Applied Econometrics 16, 389–413.

Phillips, P.C.B., Jin, S., Hu, L., 2005. Nonstationary discrete choice: a corrigendum and addendum. Cowles Foundation Discussion Paper No. 1516, Yale University.