NONSTATIONARY DISCRETE CHOICE
A CORRIGENDUM AND ADDENDUM

By

Peter C.B. Phillips, Sainan Jin and Ling Hu

June 2005

COWLES FOUNDATION DISCUSSION PAPER NO. 1516

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

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

Nonstationary Discrete Choice: A Corrigendum and Addendum*

Peter C. B. Phillips
Cowles Foundation, Yale University,
University of Auckland & University of York

Sainan Jin
Guanghua School of Management
Peking University

and

Ling Hu
Department of Economics
Ohio State University

May 26, 2005

*This is the full length version of a paper with the same title. Phillips gratefully acknowledges research support from a Kelly Fellowship and the NSF under Grant No. SES 04-142254. Jin thanks the Cowles Foundation for support under a Cowles Fellowship.

ABSTRACT

We correct the limit theory presented in an earlier paper by Hu and Phillips (*Journal of Econometrics*, 2004) 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 non-zero 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.

Key words and Phrases: Brownian motion, Brownian local time, Discrete choices, Integrated processes, Pile-up problem, Threshold parameters.

JEL Classification: C23, C25

1 Introduction

This note corrects the limit theory given in Hu and Phillips (2004, hereafter HP) for discrete choice models with integrated covariates and non-zero 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 (equations (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 non-zero 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 (2004, 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 generally be non zero. 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 therefore 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. To correct the error in the limit theory, we provide a revised version of lemma 1 and theorem 1 of HP (given in Lemma R1 and Theorem R2 below) and the results which depend on it. We also need some supplementary results on convergence to functionals of Brownian local time at spatial points away from the origin. These are provided, together with proofs of the main results, in Appendices I and II.

The results of some simulation experiments are reported. 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 Model, 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

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

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

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

The threshold parameters in (2) are scaled by $\sqrt{n}$ so that the thresholds have the same order of magnitude as the dependent variable y_t^* in (1) when the covariates x_t are integrated time series. This avoids trivial results and means, in effect, that the threshold levels adjust according to the sample size of the data. This seems realistic in a model where the covariates are allowed to be recurrent time series like integrated processes.

We assume that x_t is predetermined, i.e., x_{t+1} is adapted to some filtration $(\mathcal{F}_t)$ with respect to which ε_t is measurable. The theory of the discrete choice model in (1) and (2) when x_t is a stationary and ergodic process and when the thresholds are fixed is obtained by standard methods. In this paper, x_t is taken to be an integrated time series with integration order unity. The error process ε_t is assumed to be iid conditionally on $\mathcal{F}_{t-1}$ with marginal distribution F, which is assumed to be known and standardized, like a standard normal (leading to the probit model) or the standard logistic (leading to the logit model). Thus, the model given by (1) and (2) is taken as correctly specified. 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 we assume to be compact and convex.

In the general discrete choice model with error distribution F, the conditional probability distribution of y_t, $P(y_t = j|\mathcal{F}_{t-1}) = P_j(x_t;\theta_0)$ is given by

$$
\begin{aligned}
P_0(x_t;\theta_0) &= 1 - F(x_t'\beta_0 - \sqrt{n}\mu_0^1) \\
P_j(x_t;\theta_0) &= F(x_t'\beta_0 - \sqrt{n}\mu_0^j) - F(x_t'\beta_0 - \sqrt{n}\mu_0^{j+1}) \quad \text{for} \quad j = 1, \ldots, J-1 \\
P_J(x_t;\theta_0) &= F(x_t'\beta_0 - \sqrt{n}\mu_0^J)
\end{aligned}
$$

The corresponding conditional expectation of y_t is

$$
\begin{aligned}
m(x_t;\theta_0) &= \sum_{j=0}^{J} j \cdot P_j(x_t;\theta_0) \\
&= \sum_{j=1}^{J} F(x_t'\beta_0 - \sqrt{n}\mu_0^j)
\end{aligned}
$$

If u_t is defined as the residual in the equation

$$
y_t = m_t + u_t = \sum_{j=1}^{J} F(x_t'\beta_0 - \sqrt{n}\mu_0^j) + u_t \tag{3}
$$

then $(u_t, \mathcal{F}_t)$ is a martingale difference with conditional moments:

$$
\begin{aligned}
\sigma_k(x_t;\theta_0) &= E(u_t^k|\mathcal{F}_{t-1}) \\
&= \sum_{j=0}^{J} (j - m_t)^k \cdot P_j(x_t;\theta_0) = \sigma_{kt}, \text{ say.}
\end{aligned}
$$

Define z_{kt} as $z_k(x_t;\theta_0) = u_t^k - \sigma_{k,t}$, its conditional second moments $\eta_{kl,t}$ as $\eta_{kl}(x_t;\theta_0) = E(z_{kt} \cdot z_{lt}|\mathcal{F}_{t-1})$, and $a_{kl,t}$ as $a_{kl}(x_t;\theta_0) = z_{kt}z_{lt} - \eta_{kl,t}$. Then $(z_{kt}, \mathcal{F}_t)$, $(a_{kl,t}, \mathcal{F}_t)$ are also martingale difference. Obviously, $\sigma_{1t} = 0$ and $z_{1t} = u_t$. Further, define $\tau_{klpq,t} = E\{a_{kl,t} \cdot a_{pq,t}|\mathcal{F}_{t-1}\}$, giving fourth conditional moments for $z_{kt.}$

These moment conditions and Assumption 1 below make available the use of embedding arguments that allow for a stochastic process representation of key partial sum processes. For example, from PP (Lemma1), there exists a probability space (Ω, F, P) supporting sequences of random variables U_{nt} and V_{nt} for which

$$
(U_{nt}, V_{nt}) =_d \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{t} u_i, \frac{1}{\sqrt{n}} \sum_{i=1}^{t} v_i \right), \quad \text{for all } t \leq n, \tag{4}
$$

and for which: (i) $U_{nt} = U\left(\frac{T_{nt}}{n}\right)$, for a standard Brownian motion U and certain time changes T_{nt} in (Ω, F, P); and (ii) the process

$$V_n(r) = \sum_{t=1}^{n} V_{nt}\, 1\left\{\frac{t-1}{n} \le r < \frac{t}{n}\right\} \tag{5}$$

is such that $V_n \to_{a.s.} V$, vector Brownian motion in (Ω, F, P) with variance matrix Σ. Embeddings of this type are used in subsequent arguments, in particular in the proof of Lemma R1. We note here that $\to_{a.s.}$ reverts to weak convergence $(\Rightarrow)$ in the original space.

Assumption 1 *Let $x_t = x_{t-1} + v_t$ with $x_0 = 0$ and where*

$$v_t = \Pi(L)e_t = \sum_{i=1}^{\infty} \Pi_i e_{t-i},$$

with $\Pi(1)$ nonsingular and $\sum_{i=0}^{\infty} i\|\Pi_i\| < \infty$. The innovations e_t are iid with mean zero and $E\|e_t\|^r < \infty$ for some $r > 8$, have a distribution that is absolutely continuous with respect to Lebesgue measure and have characteristic function $\varphi(t)$ which satisfies $\lim_{\|t\|\to\infty} \|t\|^{\kappa}\varphi(t) = 0$ for some $\kappa > 0$.

As in PP, we rotate the regressor space to help isolate the effects of the nonlinearities. In particular, we assume that $\beta_0 \neq 0$ and rotate the regressor space using an orthogonal matrix $H = (h_1, H_2)$ with $h_1 = \beta_0/(\beta_0'\beta_0)^{1/2}$. Let $(\alpha_0^1, \alpha_0^{2\prime})' = \alpha_0 = H'\beta_0$. Then we can write (1) as:

$$\begin{aligned} y_t^* &= x_t'\beta_0 - \varepsilon_t \\ &= x_t'HH'\beta_0 - \varepsilon_t \\ &= (H'x_t)'H'\beta_0 - \varepsilon_t \\ &= x_{1t}\alpha_0^1 + x_{2t}'\alpha_0^2 - \varepsilon_t \end{aligned}$$

where

$$x_{1t} = h_1'x_t \quad \text{and} \quad x_{2t} = H_2'x_t,$$
$$\alpha_0^1 = h_1'\beta_0 = (\beta_0'\beta_0)^{1/2} \quad \text{and} \quad \alpha_0^2 = H_2'\beta_0 = 0.$$

Accordingly, we now define

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

which are Brownian motions of dimensions 1 and $(m-1)$, respectively. Our subsequent theory involves the local time of the scalar process V_1, which we denote by $L_{V_1}(t, s)$, where t and s are the temporal and spatial parameters. $L_{V_1}(t, s)$ is a stochastic process in time (t) and space (s) and represents the sojourn density of the process V_1 around the spatial point s over the time interval $[0, t]$. The reader is referred to Revuz and Yor (1994) for an introduction to the properties of local time and to Phillips (1998, 2001), Phillips and Park (1998), Park and Phillips (1999, 2001) for discussions and applications of this process in econometrics. In our analysis, it is more convenient to use the scaled local time of V_1 given by

$$L_1(t,s) = (1/\sigma_{11}) L_{V_1}(t,s)$$

where σ_{11} is the variance of V_1.

Now we come back to the estimation of the multiple choice model. Let

$$\Lambda(t,j) = \frac{\prod_{i=0,\ldots,J \ \& \ i\neq j}(y_t - i)}{\prod_{i=0,\ldots,J \ \& \ i\neq j}(j-i)}. \tag{6}$$

It is easy to see that $\Lambda(t,j) = 1\{y_t = j\}$. The log likelihood function can be written as:

$$\log L_n(\theta) = \sum_{t=1}^{n}\sum_{j=0}^{J}\Lambda(t,j)\log P_j(x_t;\theta).$$

Let the first derivative of F be denoted f and the second derivative be denoted $\dot{f}$. The elements of the score function $S_n(\theta) = (S_n(\beta)', S_n(\mu)')' = \left(\frac{\partial \log L_n}{\partial \beta'}, \frac{\partial \log L_n}{\partial \mu'}\right)'$ are

$$\frac{\partial \log L_n}{\partial \beta} = \sum_{t=1}^{n}\sum_{j=0}^{J}\frac{\Lambda(t,j)}{P_j(x_t;\theta)}p_j(x_t;\theta)x_t, \tag{7}$$

$$\frac{\partial \log L_n}{\partial \mu^j} = \sqrt{n}\sum_{t=1}^{n}\left(\frac{\Lambda(t,j-1)}{P_{j-1}(x_t;\theta)} - \frac{\Lambda(t,j)}{P_j(x_t;\theta)}\right) f(x_t'\beta - \sqrt{n}\mu^j), \tag{8}$$

where

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

Note that the ratio $\Lambda(t,j)/P_j$ appears in both (7) and (8). Since $E(\Lambda(t,j)|\mathcal{F}_{t-1}) = P_j(x_t;\theta_0)$, the expected value of the ratio $\Lambda(t,j)/P_j$ is 1. The ratio can be written as a sum of martingale differences, as is clear from the following calculation:

$$\begin{aligned}
\frac{\Lambda(t,j)}{P_j(x_t;\theta_0)} &= \frac{1}{P_j(x_t;\theta_0)}\frac{\prod_{i=0,\ldots,J \ \& \ i\neq j}(y_t - i)}{\prod_{i=0,\ldots,J \ \& \ i\neq j}(j-i)} \\
&= \frac{1}{P_j(x_t;\theta_0)}\frac{\prod_{i=0,\ldots,J \ \& \ i\neq j}(m_t + u_t - i)}{\prod_{i=0,\ldots,J \ \& \ i\neq j}(j-i)} \\
&= \sum_{k=1}^{J} g_k(x_t;j,\theta_0))(u_t^k - \sigma_{kt}(x_t;\theta_0)) + 1 \\
&= \sum_{k=1}^{J} g_k(x_t;j,\theta_0))z_{kt} + 1, \qquad (9)
\end{aligned}$$

where $g_k(j)$ is defined to be the coefficient associated with z_{kt} for a given j and where $z_{kt} = u_t^k - E(u_t^k|\mathcal{F}_t - 1)$, which is a martingale difference. The binary choice case is much simpler. Here, $J = 1$ and we have either $y_t = 0$, with probability $P_0(x_t;\theta_0) = 1 - F(x_t'\beta_0 - \sqrt{n}\mu_0^1)$ or $y_t = 1$, with probability $P_1(x_t;\theta_0) = F(x_t'\beta_0 - \sqrt{n}\mu_0^1)$. The indicator functions are $\Lambda(t,0) = 1 - y_t$ and $\Lambda(t,1) = y_t$. The ratio of $\Lambda(t,j)/P_j$ is then simply

$$\begin{aligned}
\frac{\Lambda(t,0)}{P_0(x_t;\theta_0)} &= \frac{1 - (0 \cdot P_0(x_t;\theta_0) + 1 \cdot P_1(x_t;\theta_0) + u_t)}{P_0(x_t;\theta_0)} \\
&= -\frac{1}{1 - F(x_t'\beta_0 - \sqrt{n}\mu_0^1)} z_{1t} + 1, \\
\frac{\Lambda(t,1)}{P_1(x_t;\theta_0)} &= \frac{0 \cdot P_0(x_t;\theta_0) + 1 \cdot P_1(x_t;\theta_0) + u_t}{P_1(x_t;\theta_0)} \\
&= \frac{1}{F(x_t'\beta_0 - \sqrt{n}\mu_0^1)} z_{1t} + 1.
\end{aligned}$$

Therefore, in a binary choice case, $g_1(x_t;0,\theta_0) = -1/(1-F)$ and $g_1(x_t;1,\theta_0) = 1/F$. Using the above results, rewrite the score functions (7) and (8) as

$$\frac{\partial \log L_n}{\partial \beta} = \sum_{t=1}^{n} \sum_{k=1}^{J} A_k(x_t;\theta) z_k(x_t;\theta) x_t, \tag{10}$$

$$\frac{\partial \log L_n}{\partial \mu^j} = \sqrt{n} \sum_{t=1}^{n} \sum_{k=1}^{J} B_k(x_t;j,\theta) z_k(x_t;\theta), \tag{11}$$

where

$$\begin{aligned}
A_k(x_t;\theta) &= \sum_{j=0}^{J} g_k(x_t;j,\theta) p_j(x_t;\theta) \\
&= \sum_{j=1}^{J} f(x_t'\beta - \sqrt{n}\mu^j) \left[g_k(x_t;j,\theta) - g_k(x_t;j-1,\theta) \right], \qquad (12)
\end{aligned}$$

and

$$B_k(x_t;j,\theta) = f(x_t'\beta - \sqrt{n}\mu^j)[g_k(x_t;j-1,\theta) - g_k(x_t;j,\theta)] \tag{13}$$

Again, in the binary choice example, it is easy to see that $A(x_t;\theta) = f/(F(1-F))$ and $B(x_t;1,\theta) = -f/(F(1-F))$. Taking second derivatives of the log likelihood function with respect to β and μ gives the hessian matrix $J_n(\theta)$. To present the elements of this matrix, we let $M(i,j)$ denote the (i,j)'th element of the matrix M and let $M(j)$ denote its j'th column. Then

$$J_n(\theta) = \begin{pmatrix} J_{n,11}(\theta) & J_{n,12}(\theta) \\ J_{n,21}(\theta) & J_{n,22}(\theta) \end{pmatrix} \tag{14}$$

where

$$
\begin{aligned}
& J_{n,11}(\theta) = \frac{\partial \log L_n}{\partial \beta \partial \beta'} \\
= \quad & -\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} A_k A_l z_k z_l x_t x_t' + \sum_{t=1}^{n}\sum_{k=1}^{J} C_{\beta\beta,k} z_k x_t x_t', \\
& J_{n,12}(\theta)(j) = \frac{\partial \log L_n}{\partial \beta \partial \mu^j} \\
= \quad & -\sqrt{n}\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} A_k B_l(j) z_k z_l x_t' + \sqrt{n}\sum_{t=1}^{n}\sum_{k=1}^{J} C_{\beta\mu^j,k} z_k x_t', \\
& J_{n,22}(\theta)(i,i) = \frac{\partial^2 \log L_n}{\partial^2 \mu^i} \\
= \quad & -n\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} B_k(i) B_l(i) z_k z_l - n\sum_{t=1}^{n}\sum_{k=1}^{J} C_{\mu^i\mu^i,k} z_k, \\
& J_{n,22}(\theta)(i,i-1) = \frac{\partial \log L_n}{\partial \mu^i \partial \mu^{i-1}} \\
= \quad & -n\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} B_k(i) B_l(i-1) z_k z_l \quad \text{for} \quad i = 2, \ldots, J \\
& J_{n,22}(\theta)(i,i+1) = \frac{\partial \log L_n}{\partial \mu^i \partial \mu^{i+1}} \\
= \quad & -n\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} B_k(i) B_l(i+1) z_k z_l \quad \text{for} \quad i = 1, \ldots, J-1 \\
& J_{n,22}(\theta)(i,j) = 0 \quad \text{for} \quad j > i+1 \quad \text{and} \quad j < i-1
\end{aligned}
$$

where we omit the arguments $(x_t;\theta)$ in the functions A, B, C and z for simplicity and where

$$
\begin{aligned}
C_{\beta\beta,k}(x_t;\theta) &= \sum_{j=0}^{J} g_k(x_t;j,\theta)\dot{p}_j(x_t;\theta), \\
C_{\beta\mu^j,k}(x_t;\theta) &= g_k(x_t;j,\theta)\dot{p}_j(x_t;\theta), \\
C_{\mu^i\mu^i,k}(x_t;\theta) &= (g_k(x_t;i-1,\theta) - g_k(x_t;i,\theta))\dot{f}(x_t'\beta - \sqrt{n}\mu^i).
\end{aligned}
$$

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 (15), as is easily verified. As in HP, we use the following classifications for nonlinear functions: $g : R \to R$ is regular if it is bounded, integrable, and differentiable with bounded derivative; F_R denotes the class of regular functions; F_I is the class of bounded and integrable functions;

and 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.

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

$$\begin{aligned}\sup_{|x|<M}\frac{F\left(x-M^{1+\eta}\mu\right)}{F\left(x\right)} &= o\left(1\right), \quad \sup_{|x|\leq M}\frac{1-F\left(x+M^{1+\eta}\mu\right)\right)}{1-F(x)}=o(1),\\ \sup_{|x|<M}\frac{f\left(x\pm M^{1+\eta}\mu\right)}{f\left(x\right)} &= o\left(1\right), \qquad (15)\end{aligned}$$

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

(a) $\eta_{kl}A_kB_l, \eta_{kl}A_kA_l, \eta_{kl}B_kB_l \in F_R$;
(b) $\eta_{kk}A_k, \eta_{kk}B_k, (\eta_{kl}\dot{A}_kB_l), (\eta_{kl}\dot{A}_kA_l), (\eta_{kl}\dot{B}_kB_l), \eta_{kk}^{1/2}\dot{C}_k \in F_I$;
(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 below corrects lemma 1 of HP. Proofs and complementary results are given in the Appendix.

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

(a)

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

(b)

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

(c)

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n}\frac{f^2(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^j)}{P_j}x_{1t}x_{2t}\Rightarrow\frac{\mu_0^j}{(\alpha_0^1)^2}\int_0^1V_2(r)dL_1\left(r,\frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty}\frac{f^2(s)}{F(s)}ds,$$

(d)

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n}\frac{f^2(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^j)}{P_{j-1}}x_{1t}x_{2t}\Rightarrow\frac{\mu_0^j}{(\alpha_0^1)^2}\int_0^1V_2(r)dL_1\left(r,\frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty}\frac{f^2(s)}{1-F(s)}ds,$$

(e)

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n}\frac{f^2(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^j)}{P_j}x_{2t}x_{2t}' \Rightarrow \frac{1}{\alpha_0^1}\int_0^1 V_2(r)V_2(r)'dL_1\left(r,\frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty}\frac{f^2(s)}{F(s)}ds,$$

(f)

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n}\frac{f^2(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^j)}{P_{j-1}}x_{2t}x_{2t}' \Rightarrow \frac{1}{\alpha_0^1}\int_0^1 V_2(r)V_2(r)'dL_1\left(r,\frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty}\frac{f^2(s)}{1-F(s)}ds,$$

(g)

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

(h)

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

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

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n}f(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^j)x_{1t}^2 \Rightarrow \frac{(\mu_0^j)^2}{(\alpha_0^1)^3}L_1\left(1,\frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty}f(s)ds, \tag{16}$$

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

$$\frac{1}{n^{1/2}}\sum_{t=1}^{n}f(x_{1t}\alpha_0^1)x_{1t}^2 \Rightarrow \frac{1}{(\alpha_0^1)^3}L_1(1,0)\int_{-\infty}^{\infty}f(s)s^2ds. \tag{17}$$

Thus, a major effect of the non-zero threshold $\mu_0^j \neq 0$ is to change the rate of convergence (or standardization) from $1/\sqrt{n}$ in (17) 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 $\int_{-\infty}^{\infty}f(s)s^2ds$ in (17) to $\mu_0^2\int_{-\infty}^{\infty}f(s)ds$ in (16). 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 non-zero threshold case that lead to the corrections needed for HP.

Lemma R1 (corrects Lemma 1 of HP) : *Let Assumption 1 hold, and write* $A_k(x_{1t};\underline{\theta}_0) = A_k$, $B_k(x_t;j,\underline{\theta}_0) = B_k$. *Assume for* $k,l,=1,..,J$, *that* $A_kA_l\eta_{kl}$, $A_kB_l\eta_{kl}$, $B_kB_l\eta_{kl} \in F_R$, $A_k\eta_{kk}$, $B_k\eta_{kk} \in F_I$, *and* $\tau_{kkkk}A_k^4$, $\tau_{kkkk}B_k^4 \in F_0$ *for* A_k, $B_k : R \to R$. *Then*

$$\begin{pmatrix} n^{-3/4}\sum_{t=1}^{n}\sum_{k=1}^{J}A_kz_{kt}x_{1t} \\ n^{-3/4}\sum_{t=1}^{n}\sum_{k=1}^{J}A_kz_{kt}x_{2t} \\ n^{-1/4}\sum_{t=1}^{n}\sum_{k=1}^{J}B_kz_{kt}\end{pmatrix} \Rightarrow M^{1/2}W(1), \tag{18}$$

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

$$M_{11} = \sum_{j=1}^{J} \left\{ \frac{(\mu_0^j)^2}{(\alpha_0^1)^3} L_1\left(1, \frac{\mu_0^j}{\alpha_0^1}\right) \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1-F(s))} ds \right\}, \tag{19}$$

$$M_{12} = \sum_{j=1}^{J} \left\{ \frac{\mu_0^j}{(\alpha_0^1)^2} \int_0^1 V_2(r) dL_1\left(r, \frac{\mu_0^j}{\alpha_0^1}\right) \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1-F(s))} ds \right\}, \tag{20}$$

$$M_{22} = \sum_{j=1}^{J} \left\{ \frac{1}{\alpha_0^1} \int_0^1 V_2(r) V_2(r)' dL_1\left(r, \frac{\mu_0^j}{\alpha_0^1}\right) \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1-F(s))} ds \right\}, \tag{21}$$

$$M_{13} = \frac{\mu_0^j}{(\alpha_0^1)^2} L_1\left(1, \frac{\mu_0^j}{\alpha_0^1}\right) \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1-F(s))} ds, \tag{22}$$

$$M_{23} = \frac{1}{\alpha_0^1} \int_0^1 dL_1\left(r, \frac{\mu_0^j}{\alpha_0^1}\right) V_2(r)' \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1-F(s))} ds, \tag{23}$$

$$M_{33} = \frac{1}{\alpha_0^1} L_1\left(1, \frac{\mu_0^j}{\alpha_0^1}\right) \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1-F(s))} ds, \tag{24}$$

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 other corrections, notably that the limit functional involves Brownian local time at spatial points $\{\mu_0^j/\alpha_0^1 : j = 1, ..., J\}$ 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 and is independent of u_t. Then we have:

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

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

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

4 Correction to the Main Results

Let $\widehat{\theta}_n = (\widehat{\beta}_n', \widehat{\mu}_n')'$ be the maximum likelihood estimator of $\theta_0 = (\beta_0', \mu_0')'$. As in PP, the asymptotic distribution of $\widehat{\theta}_n$ is obtained from the expansion

$$0 = S_n(\widehat{\theta}_n) = S_n(\theta_0) + J_n(\widetilde{\theta})(\widehat{\theta}_n - \theta_0), \tag{25}$$

which in partitioned form is

$$0 = \begin{pmatrix} S_n(\widehat{\beta}_n) \\ S_n(\widehat{\mu}_n) \end{pmatrix} = \begin{pmatrix} S_n(\beta_0) \\ S_n(\mu_0) \end{pmatrix} + \begin{pmatrix} J_{n,11}(\widetilde{\theta}) & J_{n,12}(\widetilde{\theta}) \\ J_{n,21}(\widetilde{\theta}) & J_{n,22}(\widetilde{\theta}) \end{pmatrix} \begin{pmatrix} \widehat{\beta}_n - \beta_0 \\ \widehat{\mu}_n - \mu_0 \end{pmatrix},$$

where $\widetilde{\theta}$ is on the line segment between $\widehat{\theta}_n$ and θ_0, which differs from row to row of 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 obtained from $S_n(\underline{\theta}) = G'S_n(\theta)$ and $J_n(\underline{\theta}) = G'J_n(\theta)G$. Pre-multiplying (25) by G', we have

$$0 = S_n(\underline{\hat{\theta}}_n) = S_n(\underline{\theta}_0) + J_n(\underline{\tilde{\theta}}_n)(\underline{\hat{\theta}}_n - \underline{\theta}_0). \tag{26}$$

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

Theorem R2 *Let Assumptions 1 and R2 hold. Then*

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

jointly, where Q is the symmetric matrix partitioned as

$$Q = \begin{pmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{pmatrix} \tag{27}$$

with

$$\begin{aligned}
q_{11} &= \sum_{j=1}^{J} \left\{ \frac{(\mu_0^j)^2}{(\alpha_0^1)^3} L_1\left(1, \frac{\mu_0^j}{\alpha_0^1}\right) \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1-F(s))} ds \right\}, \\
q_{12} &= \sum_{j=1}^{J} \left\{ \frac{\mu_0^j}{(\alpha_0^1)^2} \int_0^1 dL_1\left(r, \frac{\mu_0^j}{\alpha_0^1}\right) V_2(r)' \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1-F(s))} ds \right\}, \\
q_{13}(j) &= \frac{\mu_0^j}{(\alpha_0^1)^2} L_1\left(1, \frac{\mu_0^j}{\alpha_0^1}\right) \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1-F(s))} ds,
\end{aligned}$$

$$
\begin{aligned}
q_{22} &= \sum_{j=1}^{J}\left\{\frac{1}{\alpha_0^1}\int_0^1 V_2(r)V_2(r)'dL_1\left(r,\frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty}\frac{f^2(s)}{F(s)(1-F(s))}ds\right\},\\
q_{23}(j) &= \frac{1}{\alpha_0^1}\int_0^1 dL_1\left(r,\frac{\mu_0^j}{\alpha_0^1}\right)V_2(r)'\int_{-\infty}^{\infty}\frac{f^2(s)}{F(s)(1-F(s))}ds,\\
q_{33}(j,j) &= \frac{1}{\alpha_0^1}L_1\left(1,\frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty}\frac{f^2(s)}{F(s)(1-F(s))}ds,\\
q_{33}(j,i) &= 0 \text{ for } i\neq j.
\end{aligned}
$$

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:

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

where

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

with

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

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

The asymptotic results for $S_n(\underline{\theta}_0)$ and $J_n(\underline{\theta}_0)$ in Theorem R2 help deliver the limit distribution of $\widehat{\underline{\theta}}_n$. From the expansion (26), we expect that the normed and centered estimator satisfies

$$n^{3/4}(\widehat{\underline{\theta}}_n - \underline{\theta}_0) = -(n^{-3/2}J_n(\underline{\theta}_0))^{-1}n^{-3/4}S_n(\underline{\theta}_0) + o_p(1), \tag{28}$$

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

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

Remarks

1. From the above, we get

$$n^{3/4}G'\left(\widehat{\theta}_n - \theta_0\right) \Rightarrow Q^{-1/2}W(1), \tag{29}$$

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

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

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

   and

$$D_nG_2'\left(\widehat{\rho}_n - \rho_0\right) \Rightarrow \Xi^{-1/2}W(1),$$

   or

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

   thus

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

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

Corollary R4 *Under Assumptions 1 and R2, as $n \to \infty$*

$$\begin{pmatrix} n^{3/4}(\widehat{\beta}_n - \beta_0) \\ n^{3/4}(\widehat{\mu}_n - \mu_0) \end{pmatrix} \Rightarrow MN(0, GQ^{-1}G'). \tag{30}$$

When there are stationary variables in the system with coefficients γ_0, $n^{1/4}\left(\widehat{\gamma}_n - \gamma_0\right) \Rightarrow MN(0, \Xi_{11}^{-1})$ and is independent of (30), so that

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

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

$$\underline{J}_n(\hat{\theta}_n) = \begin{pmatrix} \underline{J}_{n11}(\hat{\theta}_n) & \underline{J}_{n12}(\hat{\theta}_n) \\ \underline{J}_{n21}(\hat{\theta}_n) & \underline{J}_{n22}(\hat{\theta}_n) \end{pmatrix},$$

where $\underline{J}_{n,ij}$ excludes the term in $J_{n,ij}$ that involves martingale differences, i.e.

$$\begin{aligned}
\underline{J}_{n,11}(\theta) &= -\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} A_k A_l z_k z_l x_t x_t', \\
\underline{J}_{n,12}(\theta)(i) &= -\sqrt{n}\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} A_k B_l z_k z_l x_t', \\
\underline{J}_{n,22}(\theta)(i,i) &= -n\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} B_k B_l z_k z_l,
\end{aligned}$$

and other terms in $\underline{J}$ are the same as in J. This leads to the following result, which replaces Theorem 3 of HP.

Theorem R5 *Under Assumptions 1 and 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 Simulation Experiments

5.1 Simulation Evidence of the Effects of Nonstationarity

This section provides some simulation evidence highlighting the effects of non zero thresholds on the finite sample performance of ML estimation in a poly-

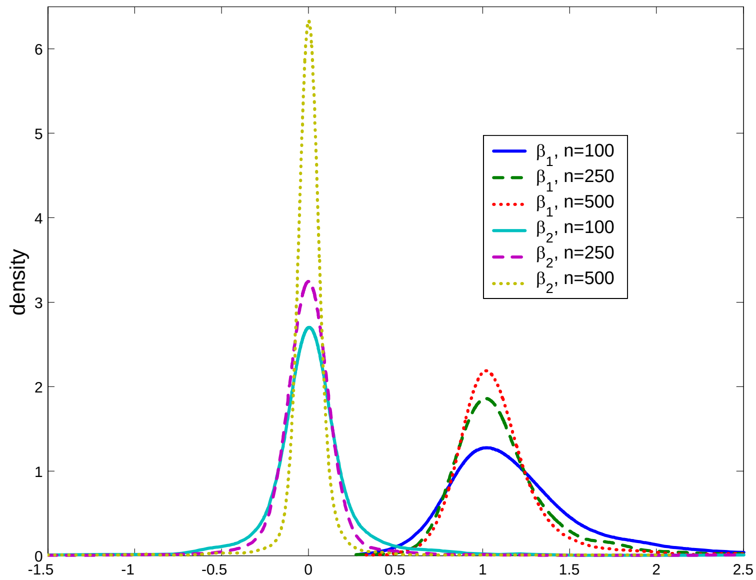

Figure 1: Densities of estimators of $\beta_0^1 = 1$, $\beta_0^2 = 0$, $\mu_0 = (-0.1, 0.1)'$.

chotomous choice model under nonstationarity. We take a model with $m = 2$ explanatory variables and $J = 2$, giving a triple-choice dependent variable y_t. The DGP for the exogenous data is the system

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

with $v_t = (v_{1t}, v_{2t})' = iid\ N(0, I_2)$, and $\rho_1 = \rho_2 = \rho = 1$. The coefficient parameter vector was set at $\beta_0 = (1, 0)'$ and $\mu_0 = (\mu_0^1, \mu_0^2)' = (-0.1, 0.1)'$ and $(-1.5, 1.5)'$ respectively. Thus $x_t'\beta_0 = \beta_0^1 x_{1t} = x_{1t}$ and the direction orthogonal to β_0 is $(0, 1)$, giving the coefficient $\beta_0^2 = 0$ of x_{2t}, so that this set up is analogous to that of the simulation study of PP. The number of replications is 50000.

Figs. 1 and 2 show that kernel estimates of the sampling distributions of the probit estimates of the coefficients β_0^1 and β_0^2 in the unit root case for sample sizes $n = 100, 250, 500$. It is obvious that when μ_0 is close to zero, the estimate of β_0^2 is more concentrated than the estimate of β_0^1, which is similar to the binary choice results found in PP. As the magnitude of the threshold parameters increase, however, it is evident that the two estimates have the same convergence rate, corroborating the limit theory of the previous section.

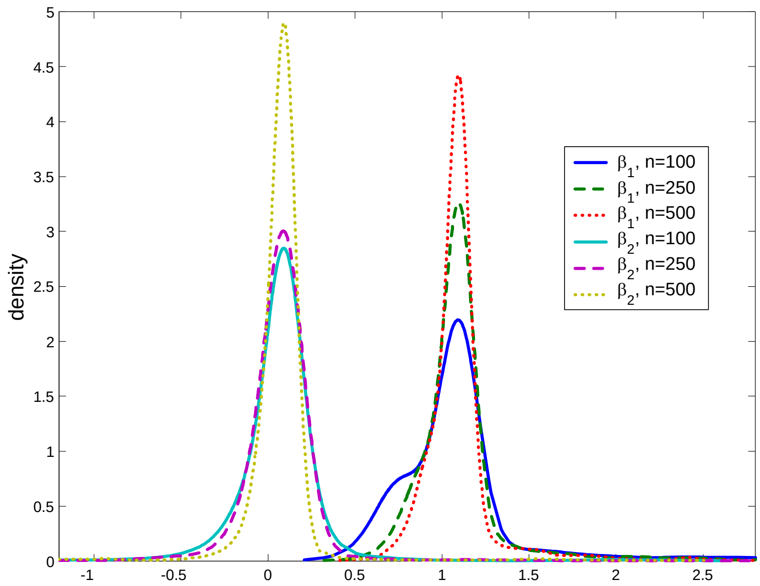

Figure 2: Densities of estimators of $\beta_0^1 = 1$, $\beta_0^2 = 0$, $\mu = (-1.5, 1.5)'$.

5.2 Simulation Experiments for the Main Results

First, we consider the distribution of the regression coefficients and thresholds $\left(\widehat{\beta}_n, \widehat{\mu}_n\right)$ and their convergence rates. We take the triple-choice case, based on the DGP

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

with $v_t = (v_{1t}, v_{2t})' = iid\ N(0, I_2)$, and $\rho_1 = \rho_2 = \rho = 1$. The coefficient parameter vector was set at $\beta_0 = (1, 0)'$ and $\mu_0 = (\mu_0^1, \mu_0^2)' = (-0.2, 0.2)'$. The number of replications is $50,000$.

Figs. 3-6 show kernel estimates of the sampling distributions of the centred and scaled probit estimates of the coefficients β_0^1, β_0^2, μ_0^1, and μ_0^2 for sample sizes $n = 100, 250, 500, 1000$. The distributions of the parameters and threshold estimates both appear to approach the asymptotic mixed normal distributions derived above. Similar results were obtained for different values of $\beta_0$and μ_0 provided the elements of μ_0 are small. When the magnitude of the μ_0^j increase, the distributions of the estimates are biased (to save the space, the graphs are not shown here). The reason for the bias appears related to the behavior of the choice probabilities, which quickly go to zero or unity when the arguments are large. This behavior also leads to a pile-up problem in the predicted choice probability distributions which we discuss below. Further, the bias is found to occur in the stationary case as well when the thresholds are large (Graphs are

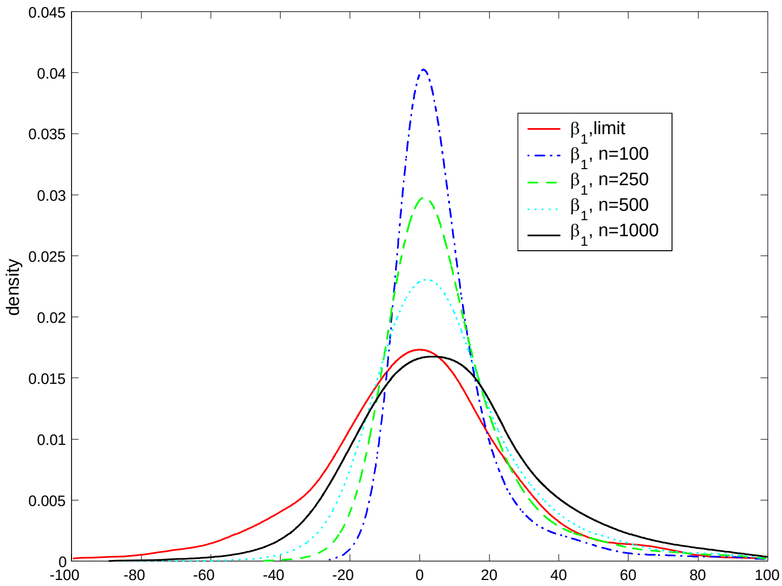

Figure 3: Density of Scaled Estimator of β_0^1 when $\mu_0 = (-0.2, 0.2)'$

not reported here to save the space).

6 Predicted Probability and Marginal Effects

6.1 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 $\widehat{v}_{j,x} = \hat{p}_j(x_t; \hat{\theta}_n)\hat{\beta}_n$, 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

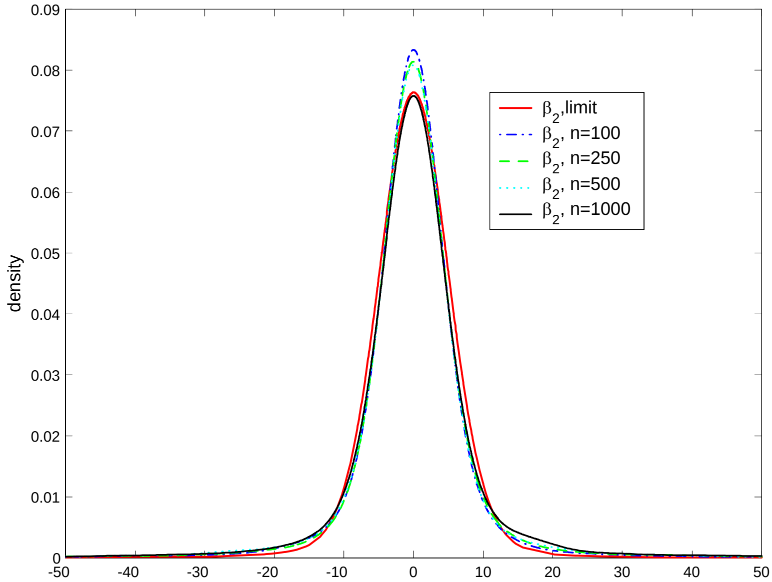

Figure 4: Density of Scaled Estimator of β_0^2 when $\mu_0 = (-0.2, 0.2)'$

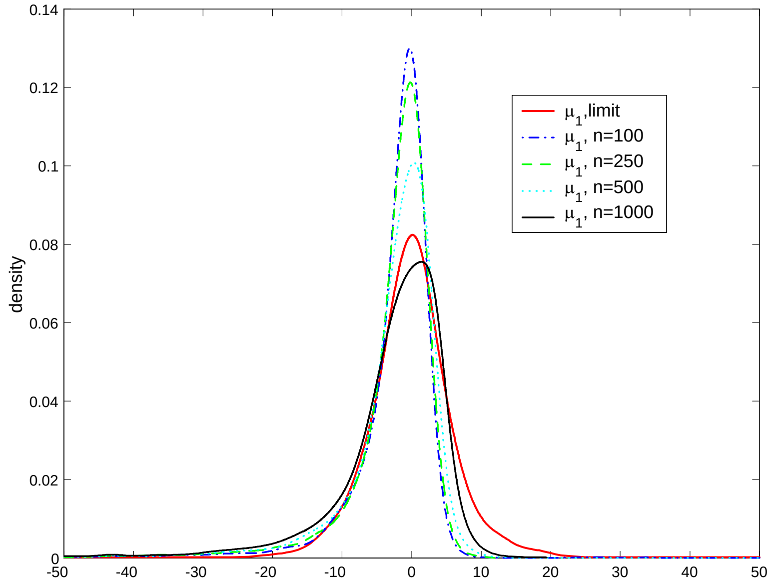

Figure 5: Density of Scaled Estimator of μ_0^1 when $\mu_0 = (-0.2, 0.2)'$

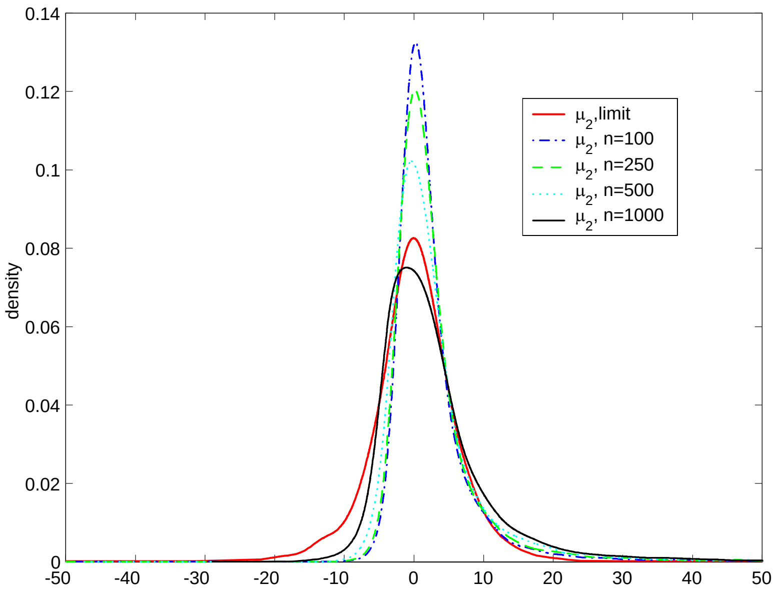

Figure 6: Density of Scaled Estimator of μ_0^2 when $\mu = (-0.2, 0.2)'$

values x and z. These probabilities now satisfy

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

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

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

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

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

The above expressions use the following notation:

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

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

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

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.

6.1.1 Simulation Experiments for R6

We use the same DGP as in the previous section, and set $\mu_0 = (\mu_0^1, \mu_0^2)' = (-0.2, 0.2)'$, $z = 1$. The number of replications is 50000. For $j = 0$, we set $x_t = x = (-0.2, 0)'$. Fig. 7 shows kernel estimates of the sampling distributions of the (scaled and centred) 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 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 discussed earlier.

The distributions of the scaled choice probabilities approach the asymptotic distributions given in the paper as n increases. However, the finite sample distribution has finite support and the figures reveal an interesting pile-up problem where the density increases towards the limits of the domain of definition. 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. 8, where $\rho_1 = \rho_2 = \rho = 0.95$, and Fig. 9 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

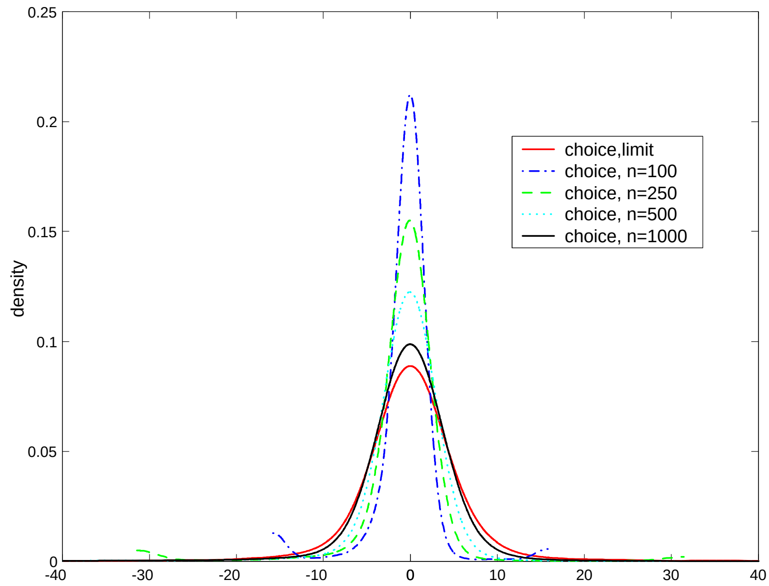

Figure 7: Density of Choice Probability for $j = 0$

are close to the extremes of the support where the limit distribution is non negligible. Thus, the problem of pile-up is not in any way special to the non-stationary 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.

The simulation results for the choice probabilities with $j = 1$, and $j = 2$ are similar (see Figs. 10 and 11).

6.2 Marginal Effects

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

Corollary R7 *Let Assumptions 1 and R2 hold. Given $x_t = x$, $z_n = z$, for $j = 0, ..., J$, the estimated marginal effects $\widehat{\upsilon}_{j,x}$ have the following asymptotic*

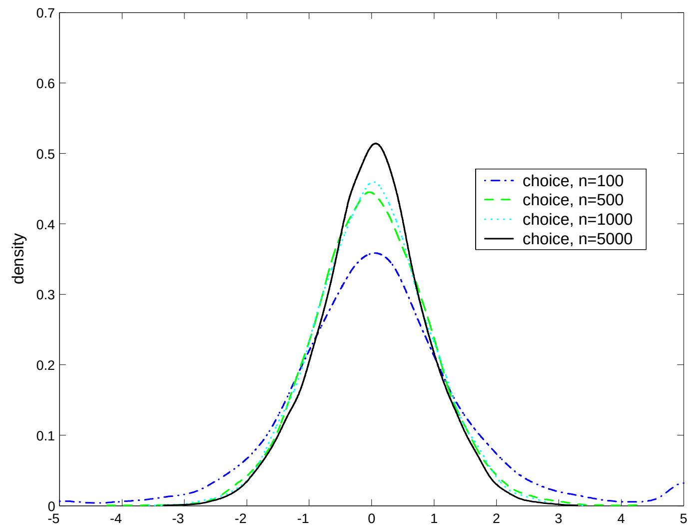

Figure 8: Stationary Models with $\rho = 0.95$

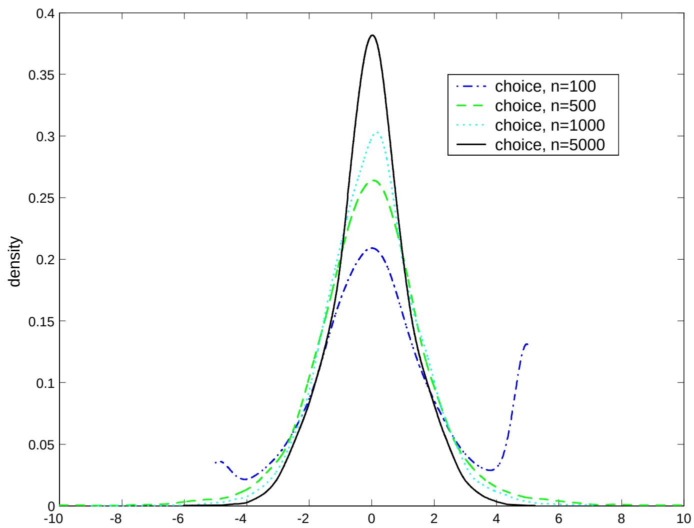

Figure 9: Stationary Models with $\rho = 0.99$

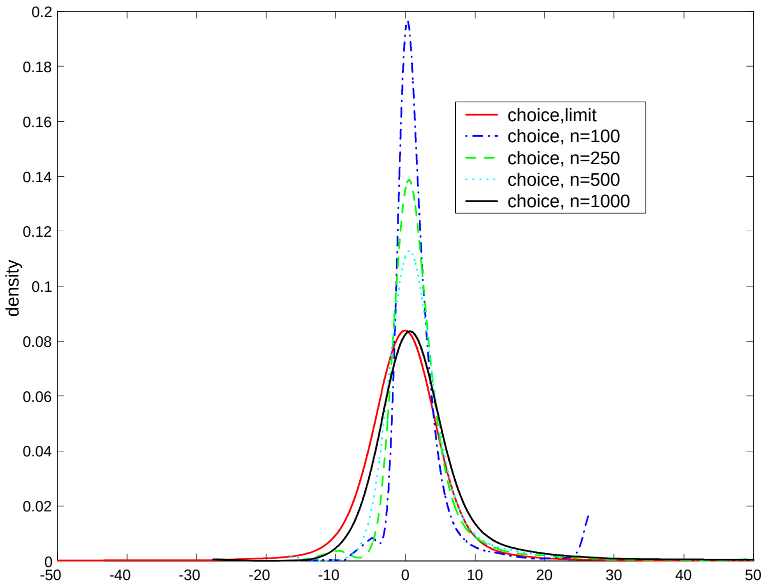

Figure 10: Density of Choice Probability for $j = 1$

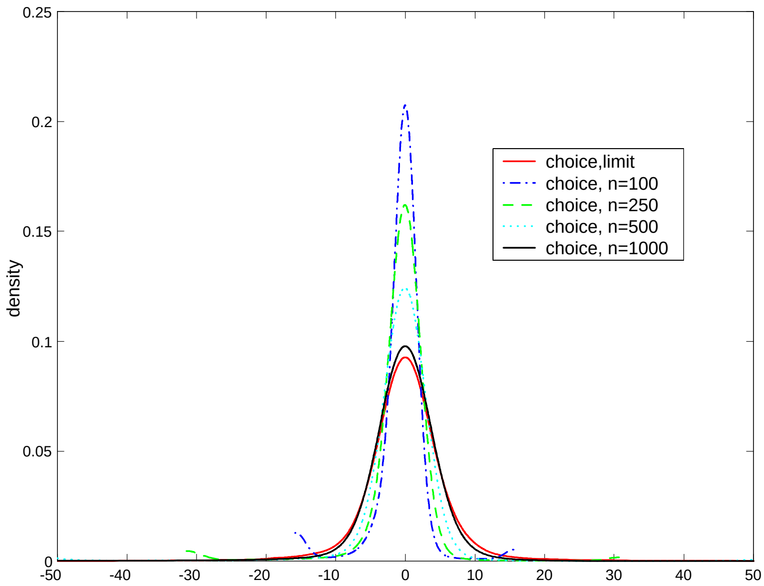

Figure 11: Density of Choice Probability for $j = 2$

distributions as $n \to \infty$

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

These expressions use the notation:

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

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

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

$$\begin{aligned} \Pi(j) &= \begin{pmatrix} \left[\dot{f}((x'\beta_0 - z\mu_0^j)) - \dot{f}((x'\beta_0 - z\mu_0^{j+1}))\right] x\beta_0' + p_j(x;\theta_0)I_m \\ -\dot{f}((x'\beta_0 - z\mu_0^j))z\beta_0 \\ \dot{f}((x'\beta_0 - z\mu_0^{j+1}))z\beta_0 \end{pmatrix}' R(j), \\ \text{for } j &= 1,...,J-1. \end{aligned}$$

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 $n \to \infty$

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

where

$$\begin{aligned} \Lambda(0) &= -\dot{f}((x_3'\gamma + x'\beta_0 - z\mu_0^1))\rho x_3' - f((x_3'\gamma + x'\beta_0 - z\mu_0^1))I_{m_3}, \\ \Lambda(j) &= \left[\dot{f}((x_3'\gamma + x'\beta_0 - z\mu_0^j)) - \dot{f}((x_3'\gamma + x'\beta_0 - z\mu_0^{j+1}))\right]\rho x_3' + p_j(x, x_3;\gamma,\theta_0)I_{m_3}, \\ \Lambda(J) &= \dot{f}((x_3'\gamma + x'\beta_0 - z\mu_0^J))\rho x_3' + f((x_3'\gamma + x'\beta_0 - z\mu_0^1))I_{m_3}. \end{aligned}$$

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.

6.2.1 Simulation Experiments for R7

We use the same DGP as in the previous section, and set $\mu_0 = (\mu_0^1, \mu_0^2)' = (-0.2, 0.2)'$, $z = 1$. The number of replications is again $50,000$. For $j = 0$, we set $x_t = x = (-0.6, 0)'$. Figs. 12-17 show kernel estimates of the sampling distributions of the marginal effects $\widehat{\upsilon}_{j,x} = \widehat{p}_j(x_t;\widehat{\theta}_n)\widehat{\beta}_n = \widehat{\upsilon}_{j,x} = \widehat{p}_j(x_t;\widehat{\theta}_n)\left(\widehat{\beta}_n^1, \widehat{\beta}_n^2\right)'$ when $j = 0, 1$, and 2 for sample sizes $n = 100, 250, 500, 1000$. In the graphs, we use ME1 to denote $\widehat{p}_j(x_t;\widehat{\theta}_n)\widehat{\beta}_n^1$, and ME2 to denote $\widehat{p}_j(x_t;\widehat{\theta}_n)\widehat{\beta}_n^2$. 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 appears

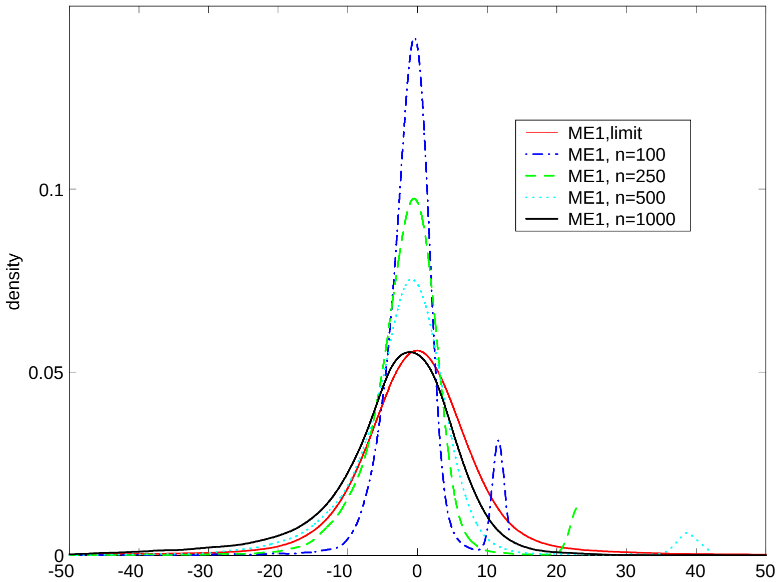

Figure 12: Density of Marginal Effects when $j = 0$

to be a pile-up problem towards the limits of the domain of definition. 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.

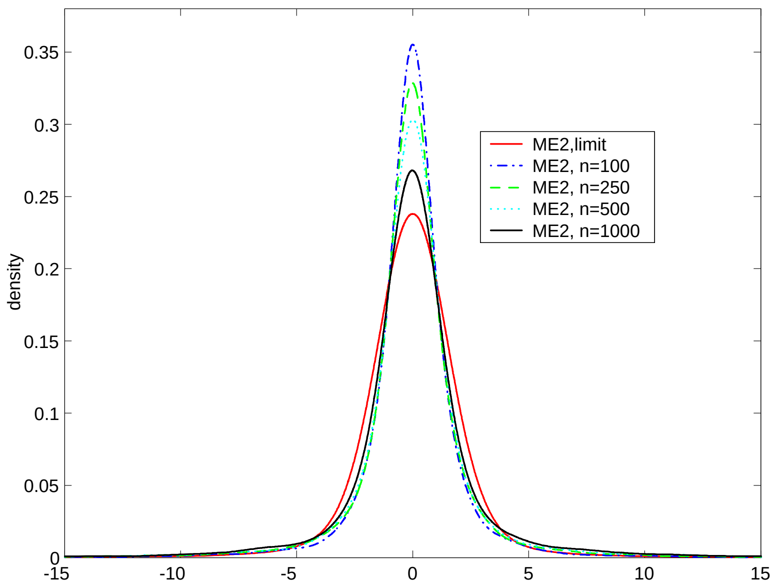

Figure 13: Density of Marginal Effects when $j = 0$

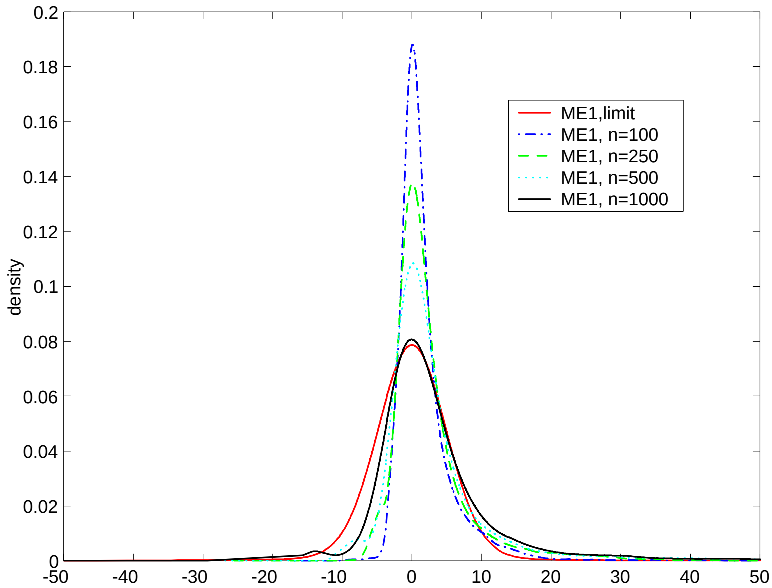

Figure 14: Density of Marginal Effects when $j = 1$

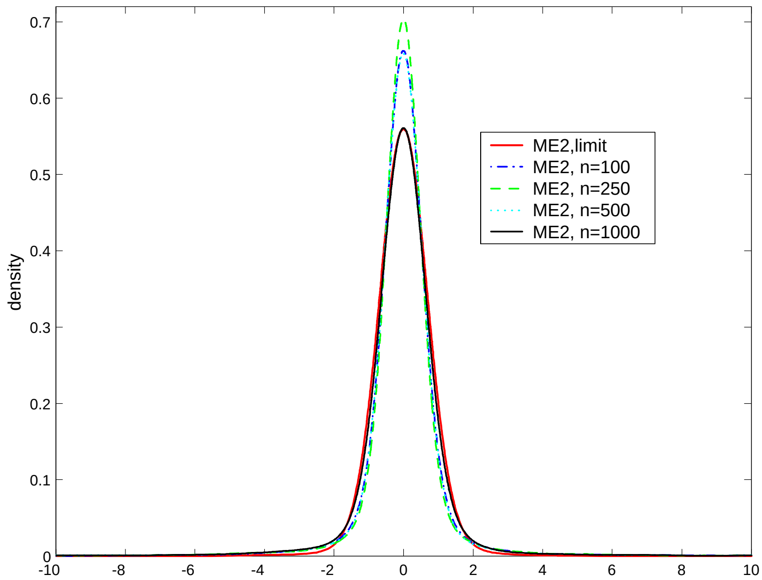

Figure 15: Density of Marginal Effects when $j = 1$

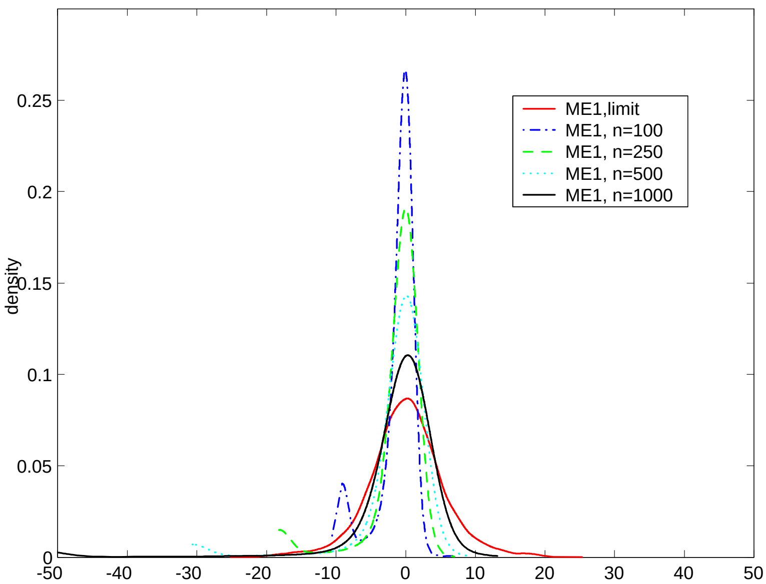

Figure 16: Density of Marginal Effects when $j = 2$

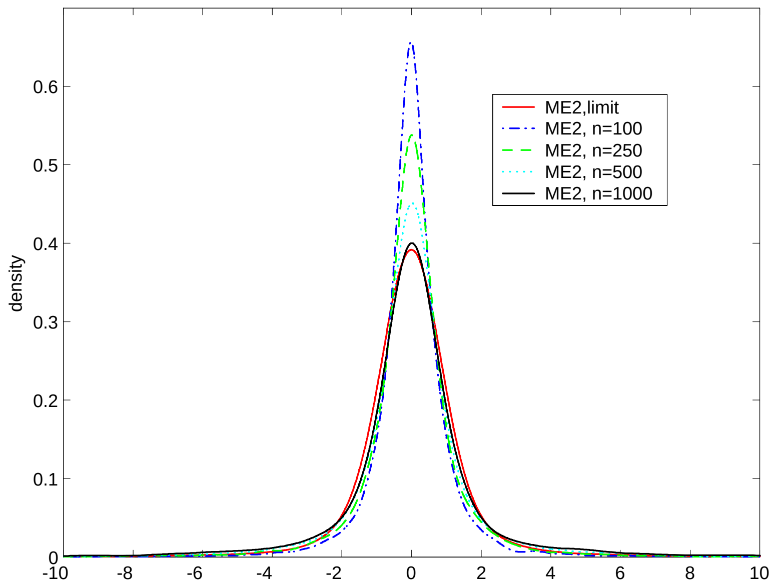

Figure 17: Density of Marginal Effects when $j = 2$

7 Appendix I: Useful Lemmas and Proofs

The following lemmas update and extend as needed here some preliminary results used in HP and proved in PP. The updating takes into account the explicit form of the dependence of functions on the threshold.

Lemma A (updates Lemma A.3 in HP) *Let Assumption 1 hold, and $f, f_k : R \to R$. Denote $x_{1t}^{\kappa_1}$ the κ_1-times tensor product of x_{1t} with itself, and $x_{2t}^{\kappa_2}$ the κ_2-times tensor product of x_{2t} with itself. Define:*

$$\begin{aligned} {}_1M_n^{\kappa_1,\kappa_2} &= \sum_{t=1}^{n} f(x_{1t};\underline{\theta}_0)x_{1t}^{\kappa_1}x_{2t}^{\kappa_2},\ {}_2M_n^{\kappa_1,\kappa_2} = \sum_{t=1}^{n}\sum_{k=1}^{J} f_k(x_{1t};\underline{\theta}_0)x_{1t}^{\kappa_1}x_{2t}^{\kappa_2}z_{kt}, \\ {}_3M_n^{\kappa_1,\kappa_2} &= \sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} f_k f_l(x_{1t};\underline{\theta}_0)x_{1t}^{\kappa_1}x_{2t}^{\kappa_2}a_{kl,t} \end{aligned}$$

(a) for $f \in F_0$, then ${}_1M_n^{\kappa_1,\kappa_2} = o_p(n^{1+\kappa_1/2+\kappa_2/2})$. Moreover, if $f \in F_I$, then ${}_1M_n^{\kappa_1,\kappa_2} = O_p(n^{(1+\kappa_1+\kappa_2)/2})$.

(b) If $\eta_{kl}f_kf_l \in F_0$, then ${}_2M_n^{\kappa_1,\kappa_2} = o_p(n^{(1+\kappa_1+\kappa_2)/2})$.

(c) If $\tau_{klpq}f_kf_lf_pf_q \in F_0$,then ${}_3M_n^{\kappa_1,\kappa_2} = o_p(n^{(1+\kappa_1+\kappa_2)/2})$.

In applying this lemma, we note the following: for part (a) where $f \in F_I$,

f could be $\eta_{kk}A_k, \eta_{kk}B_k, (\eta_{kl}\dot{A}_kB_l), (\eta_{kl}\dot{A}_kA_l), (\eta_{kl}\dot{B}_kB_l)$, or $\eta_{kk}^{1/2}\dot{C}_k$; for part (b), f_k could be C_k; and for part (c), f_k could be A_k or B_k.

Lemma B (updates Lemma A3 of PP) *Let Assumption 1 hold. Assume $\eta_{kl}f_kf_l \in F_R$, and $\tau_{klpq}f_kf_lf_pf_q \in F_0$ for some $f_k : R \to R$. Define*

$$ {}_1P_{nt}^2 = \frac{1}{n^{3/2}}\sum_{k=1}^{J}\sum_{l=1}^{J} f_kf_l(x_{1t};\underline{\theta}_0)x_{1t}^2z_{kt}z_{lt},\ {}_2P_{nt}^2 = \frac{1}{n^{3/2}}\sum_{k=1}^{J}\sum_{l=1}^{J} f_kf_l(x_{1t};\underline{\theta}_0)x_{2t}x_{2t}'z_{kt}z_{lt} $$

and ${}_sQ_n^2 = E\left({}_sP_{nt}^2|\mathcal{F}_{t-1}\right)$ Then for $s = 1,2$ we have, $\sup_{1\le t\le n} \parallel \sum_{j=1}^{t}({}_sP_{nj}^2) - \sum_{j=1}^{t}({}_sQ_{nj}^2) \parallel \to_p 0$,*as $n \to \infty$.*

In applying this lemma, f_k could be A_k or B_k.

Lemma C (updates Lemma A4 of PP and Lemma A.4 of HP) *Let Assumption 1 hold. Assume $\eta_{kk}f_k \in F_I$, and $\tau_{kkkk}f_k^2 \in F_0$ for some $f_k : R \to R$. Define*

$$ {}_1N_{nt}^2 = \frac{1}{n^{5/4}}f_k(x_{1t};\underline{\theta}_0)x_{1t}z_{kt}^2,\ {}_2N_{nt}^2 = \frac{1}{n^{5/4}}f_k(x_{1t};\underline{\theta}_0)x_{2t}z_{kt}^2 $$

Then, for $i = 1,2$ we have, as $n \to \infty$, $\sup_{1\le t\le n} \parallel \sum_{s=1}^{t}({}_iN_{ns}^2) \parallel \to_p 0$.

Again, in applying this lemma, f_k could be A_k or B_k.

Lemma D (updates Lemma A1 of PP) *Let $f \in F_I$, $\varepsilon > 0$ and define*

$$ f_{\sup}^{\varepsilon}(x,y) = \sup_{|a|\le\varepsilon}\sup_{|b|\le\varepsilon} |f(x+a;y+b)|\,. $$

Then $f_{\sup}^{\varepsilon} \in F_I$.

Proofs of Lemmas A-D The proofs follow arguments similar to those given in the proofs of Lemmas A1-A4 of PP.

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

(a) $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} f\left(x_{1t} - \sqrt{n}\mu\right) \Rightarrow L_1(1,\mu)\int_{-\infty}^{\infty} f(s)\,ds,$

(b) $\frac{1}{n}\sum_{t=1}^{n} f\left(x_{1t} - \sqrt{n}\mu\right)x_{2t} \Rightarrow \int_0^1 V_2(r)\,dL_1(r,\mu)\int_{-\infty}^{\infty} f(s)\,ds,$

(c) $\frac{1}{n^{3/2}}\sum_{t=1}^{n} f\left(x_{1t} - \sqrt{n}\mu\right)x_{2t}x_{2t}' \Rightarrow \int_0^1 V_2(r)\,V_2(r)'\,dL_1(r,\mu)\int_{-\infty}^{\infty} f(s)\,ds.$

Proof of Lemma E The proof follows the same line of argument as that used in the proof of lemma 2 of PP. The difference arises from the fact that the main contribution to the sums in each case come from the neighborhood of the threshold μ rather than the origin. More specifically, it follows by setting $a = \sqrt{n}\mu$ in theorem 4 of Akonom (1993) that

$$\frac{\sqrt{n}}{\delta_n}\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{\left[\sqrt{n}\mu \le x_{1t} \le \sqrt{n}\mu + \delta_n\right]} \to_p L_1\left(1, \mu\right), \tag{31}$$

where $\delta_n = n^{-\delta}$ for some $\delta \in (0, 1/8)$. Then, just as in the proof of theorem 5.1 of Park and Phillips (1999) we get part (a). An independent proof of a version of part (a) has also been given recently by Jeganathan (2004). In particular, theorem 2 of Jegananthan (2004) establishes that for the partial sum process $S_t = \sum_{s=1}^{t} v_s$ of a linear process v_s (that satisfies Assumption 1 of HP)

$$\frac{\beta_n}{n}\sum_{t=1}^{n} f\left(\beta_n\left(\gamma_n^{-1}S_t - \mu\right)\right) \Rightarrow L_1\left(1, \mu\right)\int_{-\infty}^{\infty} f\left(s\right) ds, \tag{32}$$

where, in the present case of domain of attraction to a normal law, $\beta_n = \gamma_n = \sqrt{n}$. Jeganathan's result is more general than part (a) because it covers cases where the component innovations in the process determining v_t may belong to the domain of attraction of a stable law and the standardized partial sums $\gamma_n^{-1}S_{[nr]}$ converge to fractional stable motion and L_1 in (32) is the local time of the limit stable process. Convergence in probability occurs in (31) because the convergence is taken to apply in a suitably augmented probability space that includes the limit processes and has random elements distributionally equivalent to the original random elements.

Parts (b) and (c) follow in the same way as parts (b) and (c) of the proof of lemma 2 in PP. We demonstrate part (b). In particular, define $f_n(x) = \sum_{k=-\kappa_n}^{\kappa_n} f(k\delta_n)\mathbf{1}_{[k\delta_n \le x < (k+1)\delta_n]}$, 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, $\kappa_n \to \infty$, $\delta_n \to 0$ and $\kappa_n\delta_n = n^d \to \infty$ with $d \in (1/p, 1/4)$ and $p > 4$. Including the nonzero threshold $\mu \neq 0$ in the development, equation (28) of PP (which represents the sum $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} f\left(x_{1t} - \sqrt{n}\mu\right) x_{2t}$ in integral form) and the lines that follow now become

$$\begin{aligned}
& \sqrt{n}\int_0^1 f(\sqrt{n}V_{1n}(r) - \sqrt{n}\mu)V_{2n}(r)\,dr \\
= \; & \sqrt{n}\int_0^1 f_n(\sqrt{n}V_{1n}(r) - \sqrt{n}\mu)V_{2n}(r)\,dr + o_p(1) \\
= \; & \sqrt{n}\sum_{k=-\kappa_n}^{\kappa_n} f(k\delta_n)\int_0^1 \mathbf{1}_{nk}(r;\mu)V_{2n}(r)\,dr + o_p(1) \\
= \; & \left(\int_{-\infty}^{\infty} f_n(s)ds\right)\frac{\sqrt{n}}{\delta_n}\int_0^1 \mathbf{1}_n(r;\mu)V_{2n}(r)\,dr + o_p(1),
\end{aligned} \tag{33}$$

where $V_{1n}(r) = h_1' V_n(r)$ and $V_{2n}(r) = H_2' V_n(r)$, $V_n(r)$ is defined in (5), and

$$\begin{aligned}
\mathbf{1}_{nk}(r;\mu) &= \mathbf{1}_{\left[k\delta_n \le \sqrt{n}\{V_{1n}(r)-\mu\}<(k+1)\delta_n\right]}, \\
\mathbf{1}_{n}(r;\mu) &= \mathbf{1}_{\left[0 \le \sqrt{n}\{V_{1n}(r)-\mu\}<\delta_n\right]}, \\
\mathbf{1}^{n}(r;\mu) &= \mathbf{1}_{\left[0 \le \sqrt{n}\{V_{1}(r)-\mu\}<\delta_n\right]}.
\end{aligned}$$

Thus, the argument in PP changes by the replacement of $V_{1n}(r)$ with $V_{1n}(r)-\mu$. We have

$$\int_{-\infty}^{\infty} f_n(s)ds \to \int_{-\infty}^{\infty} f(s)\,ds. \tag{34}$$

Moreover, with $\delta_n = n^{-\delta}$, $0 < \delta < 1/8$, and $\pi_n = \delta_n/\sqrt{n}$, we have $n^{5/8}\pi_n \to \infty$ and then as in the proof of lemma 2 of PP

$$\frac{1}{\pi_n}\int_0^1 \mathbf{1}_n(r;\mu)\,V_{2n}(r)\,dr = \frac{1}{\pi_n}\int_0^1 \mathbf{1}^n(r;\mu)\,V_2(r)\,dr + o_p(1). \tag{35}$$

Using the extended occupation times formula [e.g., Revuz and Yor (1994, Exercise 1.15, p222)], we have

$$\begin{aligned}
\frac{1}{\pi_n}\int_0^1 \mathbf{1}^n(r;\mu)\,V_2(r)\,dr &= \frac{1}{\pi_n}\int_{-\infty}^{\infty}\int_0^1 \mathbf{1}_{\left[0\le\sqrt{n}\{s-\mu\}<\delta_n\right]} V_2(r)\,dL_1(r,s)ds \\
&= \frac{1}{\pi_n}\int_{\mu}^{\mu+\delta_n/\sqrt{n}}\int_0^1 V_2(r)\,dL_1(r,s)ds \\
&= \frac{1}{\pi_n}\int_0^{\pi_n}\int_0^1 V_2(r)\,dL_1(r,\mu+q)dq \\
&\to \int_0^1 V_2(r)\,dL_1(r,\mu),
\end{aligned} \tag{36}$$

the last line following from the uniform continuity of the local time process. Result (b) now follows from (33) - (36). Part (c) follows in the same way.

7.1 Appendix II: Proofs of the Main Results

Proof of Lemma R0: We use the same approach as in the proof of Lemma E. We give the arguments here for part (a) and parts (b) - (f) follow in a similar way. Define

$$g(x) = \frac{f^2(x)}{F(x) - F\left(x+\sqrt{n}(\mu_0^j - \mu_0^{j+1})\right)} = \frac{f^2(x)}{F(x)}\left\{1 - \frac{F\left(x+\sqrt{n}(\mu_0^j-\mu_0^{j+1})\right)}{F(x)}\right\}^{-1}, \tag{37}$$

and set $g_n(x) = \sum_{k=-\kappa_n}^{\kappa_n} g(k\delta_n)\mathbf{1}_{[k\delta_n \le x < (k+1)\delta_n]}$. Using the embedding (4)-(5) and the same argument as in the proof of PP (theorem 5.1), we write

$$\begin{aligned}
&\frac{1}{n^{1/2(1+\kappa_1)}}\sum_{t=1}^{n}\frac{f^2(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^j)}{P_j}x_{1t}^{\kappa_1}\\
=\;&\frac{1}{n^{1/2}}\sum_{t=1}^{n}\frac{f^2\left(\sqrt{n}(\frac{x_{1t}}{\sqrt{n}}\alpha_0^1-\mu_0^j)\right)}{F(\sqrt{n}(\frac{x_{1t}}{\sqrt{n}}\alpha_0^1-\mu_0^j))-F(\sqrt{n}(\frac{x_{1t}}{\sqrt{n}}\alpha_0^1-\mu_0^{j+1})}\left(\frac{x_{1t}}{\sqrt{n}}\right)^{\kappa_1}\\
=_d&\sqrt{n}\int_0^1 g(\sqrt{n}\left(\alpha_0^1V_{1n}(r)-\mu_0^j\right)V_{1n}(r)^{\kappa_1}\,dr\,\{1+o_p\,(1)\}\\
=\;&\sqrt{n}\sum_{k=-\kappa_n}^{\kappa_n} g(k\delta_n)\int_0^1 \mathbf{1}_{nk}(r;\frac{\mu_0^j}{\alpha_0^1})V_{1n}(r)^{\kappa_1}dr+o_p(1)\\
=\;&\left(\int_{-\kappa_n\delta_n}^{\kappa_n\delta_n} g_n(s)ds\right)\frac{\sqrt{n}}{\delta_n}\int_0^1 \mathbf{1}_n(r;\frac{\mu_0^j}{\alpha_0^1})V_{1n}(r)^{\kappa_1}dr+o_p(1). \qquad (38)
\end{aligned}$$

Next, as in the proof of Lemma E and lemma 2 of PP, we have for $\pi_n = \delta_n/\sqrt{n}$,

$$\begin{aligned}
&\frac{1}{\pi_n}\int_0^1 \mathbf{1}_n\left(r;\frac{\mu_0^j}{\alpha_0^1}\right)V_{1n}(r)^{\kappa_1}\,dr\\
=\;&\frac{1}{\pi_n}\int_0^1 \mathbf{1}^n\left(r;\frac{\mu_0^j}{\alpha_0^1}\right)V_1(r)^{\kappa_1}dr+o_p(1)\\
=\;&\frac{1}{\pi_n}\int_{-\infty}^{\infty}\int_0^1 \mathbf{1}_{\left[0\le\sqrt{n}\left\{\alpha_0^1 s-\mu_0^j\right\}<\delta_n\right]}s^{\kappa_1}dL_1(r,s)ds\\
=\;&\frac{1}{\alpha_0^1\pi_n}\int_0^{\pi_n}\left(\frac{\mu_0^j}{\alpha_0^1}+\frac{c}{\alpha_0^1}\right)^{\kappa_1}L_1\left(1,\frac{\mu_0^j}{\alpha_0^1}+\frac{c}{\alpha_0^1}\right)dc\\
\to\;&\frac{1}{\alpha_0^1}\left(\frac{\mu_0^j}{\alpha_0^1}\right)^{\kappa_1}L_1\left(1,\frac{\mu_0^j}{\alpha_0^1}\right), \qquad (39)
\end{aligned}$$

by the uniform continuity of the local time function.

Since $\mu_0^j < \mu_0^{j+1}$, and $\kappa_n\delta_n = n^d \to \infty$ with $d \in (1/p, 1/4)$, we have in view of (15)

$$\sup_k\left|\frac{F\left(k\delta_n+\sqrt{n}(\mu_0^j-\mu_0^{j+1})\right)}{F\left(k\delta_n\right)}\right| \le \sup_{|x|\le n^d}\left|\frac{F\left(x+\sqrt{n}(\mu_0^j-\mu_0^{j+1})\right)}{F\left(x\right)}\right| = o\,(1)\,,$$

and then

$$\begin{aligned}
\int_{-\kappa_n\delta_n}^{\kappa_n\delta_n} g_n(s)ds \;&=\; \int_{-\kappa_n\delta_n}^{\kappa_n\delta_n}\frac{f^2s)}{F(s)}ds\,\{1+o\,(1)\}\\
&=\;\int_{-\infty}^{\infty}\frac{f^2s)}{F(s)}ds+o\,(1)\,, \qquad (40)
\end{aligned}$$

as $n \to \infty$ by the integrability of $f^2(s)/F(s)$. Combining (38) - (40) we get

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

as required for part (a). Parts (b) - (f) of the lemma are proved in the same manner.

The proofs of parts (g) and (h) are also similar. To show part (g), for instance, we set

$$\begin{aligned} g(x) &= \frac{f(x) f\left(x + \sqrt{n}(\mu_0^j - \mu_0^{j-1})\right)}{F\left(x + \sqrt{n}(\mu_0^j - \mu_0^{j-1})\right) - F(x)} = \frac{f(x) f\left(x + \sqrt{n}(\mu_0^j - \mu_0^{j-1})\right)}{1 - F(x) - \left\{1 - F\left(x + \sqrt{n}(\mu_0^j - \mu_0^{j-1})\right)\right\}} \\ &= \frac{f(x)^2}{1 - F(x)} \frac{f\left(x + \sqrt{n}(\mu_0^j - \mu_0^{j-1})\right) / f(x)}{1 - \left\{1 - F\left(x + \sqrt{n}(\mu_0^j - \mu_0^{j-1})\right)\right\} / \{1 - F(x)\}}, \end{aligned} \tag{41}$$

and define $g_n(x) = \sum_{k=-\kappa_n}^{\kappa_n} g(k\delta_n) \mathbf{1}_{[k\delta_n \leq x < (k+1)\delta_n]}$ as before. Using the embedding (4)-(5) again and the same argument as in PP (theorem 5.1), we have

$$\begin{aligned} &\frac{1}{n^{1/2}} \sum_{t=1}^{n} \frac{f(x_{1t}\alpha_0^1 - \sqrt{n}\mu_0^j) f(x_{1t}\alpha_0^1 - \sqrt{n}\mu_0^j)}{P_{j-1}} \\ &= \frac{1}{n^{1/2}} \sum_{t=1}^{n} \frac{f(x_{1t}\alpha_0^1 - \sqrt{n}\mu_0^j) f(x_{1t}\alpha_0^1 - \sqrt{n}\mu_0^{j-1})}{F(\sqrt{n}(\frac{x_{1t}}{\sqrt{n}}\alpha_0^1 - \mu_0^{j-1})) - F(\sqrt{n}(\frac{x_{1t}}{\sqrt{n}}\alpha_0^1 - \mu_0^j)} \\ &=_d \sqrt{n} \int_0^1 g(\sqrt{n}\left(\alpha_0^1 V_{1n}(r) - \mu_0^j\right) dr \{1 + o_p(1)\} \\ &= \sqrt{n} \sum_{k=-\kappa_n}^{\kappa_n} g(k\delta_n) \int_0^1 \mathbf{1}_{nk}(r; \frac{\mu_0^j}{\alpha_0^1}) dr + o_p(1) \\ &= \left(\int_{-\kappa_n\delta_n}^{\kappa_n\delta_n} g_n(s) ds\right) \frac{\sqrt{n}}{\delta_n} \int_0^1 \mathbf{1}_n(r; \frac{\mu_0^j}{\alpha_0^1}) dr + o_p(1). \end{aligned} \tag{42}$$

Next, since $\mu_0^j > \mu_0^{j-1}$, $\kappa_n\delta_n = n^d \to \infty$, and $d \in (1/p, 1/4)$, it follows from Assumption **R2** and (15) that

$$\sup_k \left| \frac{1 - F\left(k\delta_n + \sqrt{n}(\mu_0^j - \mu_0^{j-1})\right)}{1 - F(k\delta_n)} \right| \leq \sup_{|x| \leq n^d} \left| \frac{1 - F\left(x + \sqrt{n}(\mu_0^j - \mu_0^{j-1})\right)}{1 - F(x)} \right| = o(1),$$

and

$$\sup_k \frac{f\left(k\delta_n + \sqrt{n}(\mu_0^j - \mu_0^{j-1})\right)}{f(k\delta_n)} \leq \sup_{|x| \leq n^d} \frac{f\left(x + \sqrt{n}(\mu_0^j - \mu_0^{j-1})\right)}{f(x)} = o(1),$$

so that

$$\begin{aligned}
g\left(k\delta_n\right) &\leq \frac{f(k\delta_n)^2}{1-F(k\delta_n)} \sup_{|s|\leq\kappa_n\delta_n} \frac{f\left(s+\sqrt{n}(\mu_0^j-\mu_0^{j-1})\right)/f\left(s\right)}{1-\left\{1-F\left(s+\sqrt{n}(\mu_0^j-\mu_0^{j-1})\right)\right\}/\left\{1-F(s)\right\}} \\
&\leq \frac{f(k\delta_n)^2}{1-F(k\delta_n)} \frac{\sup_{|x|\leq\kappa_n\delta_n}\left\{f\left(x+\sqrt{n}(\mu_0^j-\mu_0^{j-1})\right)/f\left(x\right)\right\}}{1-\eta_n} \\
&\leq \frac{f(k\delta_n)^2}{1-F(k\delta_n)} \frac{\varepsilon_n}{1-\eta_n},
\end{aligned}$$

for some sequences $\varepsilon_n, \eta_n \to 0$ as $n \to \infty$. Then, setting $h\left(x\right) = f(x)^2/\left(1-F(x)\right)$, and $h_n(x) = \sum_{k=-\kappa_n}^{\kappa_n} h(k\delta_n)\mathbf{1}_{[k\delta_n\leq x<(k+1)\delta_n]}$, we have

$$\sum_{k=-\kappa_n}^{\kappa_n} g(k\delta_n)\delta_n \leq \sum_{k=-\kappa_n}^{\kappa_n} h(k\delta_n)\delta_n\frac{\varepsilon_n}{1-\eta_n} = \int_{-\kappa_n\delta_n}^{\kappa_n\delta_n} h_n(s)ds\frac{\varepsilon_n}{1-\eta_n}.$$

Since $f(x)^2/\left(1-F(x)\right)$ is integrable it follows that

$$\int_{-\kappa_n\delta_n}^{\kappa_n\delta_n} g_n(s)ds = \sum_{k=-\kappa_n}^{\kappa_n} g(k\delta_n)\delta_n = o\left(1\right). \tag{43}$$

Finally, as in (31), we have

$$\frac{\sqrt{n}}{\delta_n}\int_0^1 \mathbf{1}_n(r;\frac{\mu_0^j}{\alpha_0^1})dr \to_p L_1\left(1,\frac{\mu_0^j}{\alpha_0^1}\right). \tag{44}$$

Hence, from (42), (43) and (44) we deduce that

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

giving part (g). Part (h) follows in the same way.

Proof of Lemma R1: The proof uses the same approach as that given in lemma 3 of PP, but adjusts for the scaling rate and uses the limit results from Lemma R0. It is also necessary to use the Skorohod embedding on a linear combination of the components, rather than individual components as in HP because the embedding is valid only for scalar processes. In particular, setting $m = 2$ without loss of generality and for any $c = (c_1, c_2, c_3) \in \mathbf{R}^3$, we let

$$C_{kn}(x_1,x_2) = c_1 n^{-3/4}A_k x_1 + c_2 n^{-3/4}A_k x_2 + c_3 n^{-1/4}B_k := c' F_{kn}\left(x_1,x_2\right),$$

and define $w_{nt}^c = \sum_{k=1}^J C_{kn}(x_{1t},x_{2t})z_{kt}$, which is a martingale difference sequence by construction. As in the proof of lemma 1 of PP, there exists a probability space supporting sequences of random variables U_{nt} and V_{nt} for which we have the distributional equivalence

$$(U_{nt}^c, V_{nt}) =_d \left(\frac{1}{\sqrt{n}} \sum_{i=1}^t w_{ni}^c, \frac{1}{\sqrt{n}} \sum_{i=1}^t v_i \right), \quad \text{jointly for all } t \le n$$

and for which the following hold. First, there is a representation $U_{nt}^c = U^c\left(\frac{T_{nt}}{n}\right)$ in terms of a standard Brownian motion U^c with time changes T_{nt} in $(\Omega, \mathcal{F}, \mathbf{P})$. Moreover, letting $T_{nt} = \sum_{i=1}^t \tau_{ni}$ and defining $\mathcal{F}_{nt} = \sigma\left((U^c(r))_{r \le T_{nt}/n}, (V_{ns})_{s=1}^{t+1}\right)$, the time changes satisfy

$$\begin{aligned}
\mathbf{E}(\tau_{nt}|\mathcal{F}_{n,t-1}) &= \mathbf{E}(w_{nt}^{c2}|\mathcal{F}_{t-1}) = \sum_{k=1}^J \sum_{l=1}^J C_{kn}(x_{1t}, x_{2t}) C_{ln}(x_{1t}, x_{2t}) E\left(z_{kt} z_{lt}|\mathcal{F}_{n,t-1}\right) \\
&= \sum_{k=1}^J \sum_{l=1}^J C_{kn}(x_{1t}, x_{2t}) C_{ln}(x_{1t}, x_{2t}) \eta_{kl}(x_{1t}),
\end{aligned}$$

and $\mathbf{E}(\tau_{nt}^r|\mathcal{F}_{n,t-1}) \le c_r \mathbf{E}(|w_{nt}^c|^{2r}|\mathcal{F}_{t-1})$ for all $r \ge 1$, where c_r is some constant depending only upon r (c.f. Hall and Heyde, 1980, theorem A1). Finally, defining

$$V_n(r) = \sum_{t=1}^n V_{nt} \, 1\left\{ \frac{t-1}{n} \le r < \frac{t}{n} \right\},$$

then $V_n \to_{a.s.} V$ where V is Brownian motion in (Ω, F, P) with variance matrix Σ.

Define $M_n^c(r) = U^c(r)$ for $\frac{T_{n[nr]}}{n} \le r < \frac{T_{n[nr]+1}}{n}$. Then, $M_n^c(r)$ is a continuous martingale satisfying

$$\sum_{t=1}^{[nr]} \sum_{k=1}^J C_{kn}(x_{1t}, x_{2t}) z_{kt} =_d M_n^c\left(\frac{T_{n[nr]}}{n}\right).$$

Set

$$D_{kl,n}(x_1, x_2) = \eta_{kl}(x_1) C_{kn}(x_1, x_2) C_{ln}(x_1, x_2) = c' F_{kn}(x_1, x_2) \eta_{kl}(x_1) F_{ln}(x_1, x_2)' c,$$

and then the quadratic variation of $M_n^c\left(T_{n[nr]}/n\right)$ is

$$\begin{aligned}
[M_n^c]_r &= \sum_{k=1}^J \sum_{l=1}^J \sum_{t=1}^{[nr]} D_{kl,n}(\sqrt{n} V_{nt}) + o_p(1) \\
&= c' \left\{ \sum_{k=1}^J \sum_{l=1}^J \sum_{t=1}^{[nr]} F_{kn}\left(\sqrt{n} V_{nt}\right) F_{ln}\left(\sqrt{n} V_{nt}\right)' \eta_{kl}(\sqrt{n} V_{1nt}) \right\} c,
\end{aligned}$$

uniformly in $r \in [0, 1]$. Next, as we show below,

$$[M_n^c]_r \to_p c' M(r) c, \tag{45}$$

uniformly in $r \in [0,1]$, where $M(r) = ([M_{ij}(r)])$ with submatrices M_{ij} (in block i,j of a 3×3 partition) given for $r = 1$ by (19) - (21) in the statement of the Lemma. Limit expressions for the M_{ij} are obtained as in Lemma R0. We illustrate the argument for the $(1,1)$ submatrix M_{11} of $M(1)$, which corresponds to the coefficient of c_1^2 in $[M_n^c](1)$. This element has the form

$$n^{-3/2} \sum_{k=1}^{J} \sum_{l=1}^{J} \sum_{t=1}^{n} A_k \left(\sqrt{n} V_{1nt}; \underline{\theta}_0\right) A_l \left(\sqrt{n} V_{1nt}; \underline{\theta}_0\right) \eta_{kl}(\sqrt{n} V_{1nt}; \underline{\theta}_0) x_{1t}^2, \tag{46}$$

where the product $A_k A_l \eta_{kl}$ satisfies Assumption **R2**(a), which is now demonstrated. We have

$$\begin{aligned}
& n^{-3/2} \sum_{k=1}^{J} \sum_{l=1}^{J} \sum_{t=1}^{n} A_k (x_{1t}; \underline{\theta}_0) A_l (x_{1t}; \underline{\theta}_0) \eta_{kl}(x_{1t}; \underline{\theta}_0) x_{1t}^2 \\
= \; & n^{-3/2} \sum_{t=1}^{n} \sum_{j=1}^{J} \sum_{i=1}^{J} \{ f(x_{1t}\alpha_0^1 - \sqrt{n}\mu_0^j) f(x_{1t}\alpha_0^1 - \sqrt{n}\mu_0^i) \\
& \times \sum_{k=1}^{J} \sum_{l=1}^{J} [g_k(j) - g_k(j-1)] [g_l(i) - g_l(i-1)] \eta_{kl,t} x_{1t}^2 \} \\
= \; & n^{-3/2} \sum_{t=1}^{n} \sum_{j=1}^{J} \sum_{i=1, i \neq j}^{J} \{ f(x_{1t}\alpha_0^1 - \sqrt{n}\mu_0^j) f(x_{1t}\alpha_0^1 - \sqrt{n}\mu_0^i) \\
& \times \sum_{k=1}^{J} \sum_{l=1}^{J} [g_k(j)g_l(i) - g_k(j)g_l(i-1) - g_k(j-1)g_l(i) + g_k(j-1)g_l(i-1)] \eta_{kl,t} x_{1t}^2 \} \\
& + n^{-3/2} \sum_{t=1}^{n} \sum_{j=1}^{J} \{ f^2(x_{1t}\alpha_0^1 - \sqrt{n}\mu_0^j) \\
& \times \sum_{k=1}^{J} \sum_{l=1}^{J} [g_k(j)g_l(j) - g_k(j)g_l(j-1) - g_k(j-1)g_l(j) + g_k(j-1)g_l(j-1)] \eta_{kl,t} x_{1t}^2 \}
\end{aligned} \tag{47}$$

From (9), we have

$$\sum_{k=1}^{J} g_k(x_t; j, \underline{\theta}_0) z_{kt} = \frac{\Lambda(t,j)}{P_j(x_t; \underline{\theta}_0)} - 1 \tag{48}$$

and

$$\sum_{k=1}^{J} \sum_{l=1}^{J} g_k(x_t; j, \underline{\theta}_0) g_l(x_t; i, \underline{\theta}_0) z_{kt} z_{lt} = \left(\frac{\Lambda(t,j)}{P_j(x_t; \underline{\theta}_0)} - 1 \right) \left(\frac{\Lambda(t,i)}{P_i(x_t; \underline{\theta}_0)} - 1 \right). \tag{49}$$

Then

$$
\begin{aligned}
&\sum_{k=1}^{J}\sum_{l=1}^{J} g_k(x_t;j,\underline{\theta}_0)g_l(x_t;i,\underline{\theta}_0)\eta_{kl,t} \\
=\;& E\left\{\left(\frac{\Lambda(t,j)}{P_j(x_t;\underline{\theta}_0)}-1\right)\left(\frac{\Lambda(t,i)}{P_i(x_t;\underline{\theta}_0)}-1\right)|\mathcal{F}_{t-1}\right\} \\
=\;& E\left\{\frac{\Lambda(t,j)}{P_j(x_t;\underline{\theta}_0)}\frac{\Lambda(t,i)}{P_i(x_t;\underline{\theta}_0)}-\frac{\Lambda(t,j)}{P_j(x_t;\underline{\theta}_0)}-\frac{\Lambda(t,i)}{P_i(x_t;\underline{\theta}_0)}+1|\mathcal{F}_{t-1}\right\} \\
=\;& -1 \text{ for } i\neq j \qquad (50)\\
=\;& \frac{1}{P_j(x_t;\underline{\theta}_0)}-1 \text{ for } i=j \qquad (51)
\end{aligned}
$$

since $E\left(\Lambda(t,j)|\mathcal{F}_{t-1}\right)=P_j(x_t;\underline{\theta}_0)$.

Next, from (50), the first term of (47) is

$$n^{-3/2}\sum_{t=1}^{n}\sum_{j=1}^{J}\sum_{i=1,i\neq j}^{J}\left\{f(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^j)f(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^i)x_{1t}^2(-1-(-1)-(-1)-1)\right\}=0 \tag{52}$$

From (50) and (51), the second term of (47) is

$$n^{-3/2}\sum_{t=1}^{n}\sum_{j=1}^{J}\left\{f^2(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^j)\left[\frac{1}{P_j}+\frac{1}{P_{j-1}}\right]x_{1t}^2\right\} \tag{53}$$

for $j=1,...,J-1$. Using a Mills ratio argument, as in Park and Phillips (2000), it is apparent that (53) has elements that belong to F_R. Hence, $A_kA_l\eta_{kl}$ in (56) belongs to F_R, thereby satisfying Assumption **R2**; and the other conditions follow upon some further routine calculations.

The required limit result now follows directly from Lemma R0 parts (a) and (b). Thus, the $(1,1)$ submatrix of $M(1)$ converges weakly to

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

The proofs for the other elements of $M(1)$ are similar, and they are obtained in the proof of the hessian asymptotics given below in the proof of Theorem R2.

Next, let σ_{uv}^c be the covariance of U^c and V and define

$$E_n(x_1,x_2)=\sum_{k=1}^{J}C_{kn}(x_1,x_2)\eta_{kk}(x_1).$$

The quadratic covariation process $[M_n^c,V]$ of M_n^c and V is:

$$
\begin{aligned}
[M_n^c,V](r) &= \sqrt{n}\sum_{i=1}^{[nr]}\sum_{k=1}^{J}C_{kn}(\sqrt{n}V_{ni})\left(\frac{T_{n,i}}{n}-\frac{T_{n,i-1}}{n}\right)\sigma_{uv}^c \\
&= \sigma_{uv}\sum_{t=1}^{n}E_n(\sqrt{n}V_{nt})1\left\{r\geq\frac{T_{n,t}}{n}\right\}+o_p(1)\to_p 0,
\end{aligned}
$$

uniformly in $r \in [0,1]$, by Lemma C. It follows, in particular, that

$$[M_n^c, V](\rho_n(r)) \to_p 0, \tag{54}$$

where $\rho_n(r) = \inf\{s \in [0,1] : [M_n^c]_s > r\}$ is a sequence of time changes.

The asymptotic distribution of the martingale M_n^c is completely determined by (45) and (54), as shown in Revuz and Yor (1994, Theorem 2.3). Define

$$W_n(r) = M_n^c(\rho_n(r)).$$

The process W_n is the DDS (or Dambis, Dubins-Schwarz) Brownian motion (see Revuz and Yor, 1994) of the martingale M_n^c. It follows that (V, M_n) converges jointly in distribution to two independent standard linear Brownian motions (V, Y), say. Therefore,

$$M_n^c\left(\frac{T_{kn,n}}{n}\right) \to_d Y(c'Mc),$$

which completes the argument because c is arbitrary.

When $J = 1$ we are back to the binary choice model. If we set the threshold parameters to zero in this case, then (53) reduces to

$$\sum_{t=1}^{n}\left\{f^2(x_{1t})\left[\frac{1}{F} + \frac{1}{1-F}\right]x_{1t}^2\right\} = \sum_{t=1}^{n}\frac{f^2(x_{1t})x_{1t}^2}{F\,(1-F)} = \sum_{t=1}^{n}\ell\,(x_{1t}),$$

a formula that occurs in the calculations in Park and Phillips (2000). Here, since ℓ belongs to F_R and the major contribution to the sum $\sum_{t=1}^{n}\ell\,(x_{1t})$ comes from the vicinity of the origin, we have $n^{-1/2}\sum_{t=1}^{n}\ell\,(x_{1t}) \Rightarrow L_1\,(1,0)\int_{-\infty}^{\infty}\ell\,(s)\,ds$, so that a different rate of convergence and a different limit result hold compared with (19).

Proof of Theorem R2 The results for the score function follow from Lemma R1 and those for the hessian involve similar calculations. Specifically, we partition the hessian $J_n(\underline{\theta}_n) = G'J_n(\underline{\theta}_n)G$ as

$$\begin{pmatrix} J_{n,11}(\underline{\theta}_0) & J_{n,12}(\underline{\theta}_0) & J_{n,13}(\underline{\theta}_0) \\ J_{n,21}(\underline{\theta}_0) & J_{n,22}(\underline{\theta}_0) & J_{n,23}(\underline{\theta}_0) \\ J_{n,31}(\underline{\theta}_0) & J_{n,32}(\underline{\theta}_0) & J_{n,33}(\underline{\theta}_0) \end{pmatrix}.$$

In view of symmetry, we consider only the upper-right triangular block:

$$\begin{aligned} J_{n,11}(\underline{\theta}_0) &= -\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J}A_kA_lz_kz_lx_{1t}^2 \\ &\quad +\sum_{t=1}^{n}\sum_{k=1}^{J}C_{\beta\beta,k}z_kx_{1t}^2, \end{aligned}$$

$$J_{n,12}(\underline{\theta}_0) = -\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} A_k A_l z_k z_l x_{1t} x_{2t}' + \sum_{t=1}^{n}\sum_{k=1}^{J} C_{\beta\beta,k} z_k x_{1t} x_{2t}',$$

$$J_{n,22}(\underline{\theta}_0) = -\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} A_k A_l z_k z_l x_{2t} x_{2t}' + \sum_{t=1}^{n}\sum_{k=1}^{J} C_{\beta\beta,k} z_k x_{2t} x_{2t}',$$

$$J_{n,13}(\underline{\theta}_0)(i) = -\sqrt{n}\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} A_k B_l(x\underline{\theta}) z_k(x_{1t}, i, \underline{\theta}) z_l x_{1t} + \sqrt{n}\sum_{t=1}^{n}\sum_{k=1}^{J} C_{\beta\mu^i,k} z_k x_{1t},$$

$$J_{n,23}(\underline{\theta}_0)(i) = -\sqrt{n}\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} A_k B_l(x\underline{\theta}) z_k(x_{1t}, i, \underline{\theta}) z_l x_{2t}' + \sqrt{n}\sum_{t=1}^{n}\sum_{k=1}^{J} C_{\beta\mu^i,k} z_k x_{2t}',$$

$$J_{n,33}(\underline{\theta}_0) = J_{n,22}(\underline{\theta}_0),$$

where the arguments $(x_{1t}; \underline{\theta}_0)$ in the functions of A, B, C are omitted for simplicity. Observe that all terms involving z_k alone as a factor are $o_p(1)$ by Lemma A (b), that is,

$$n^{-3/2}\sum_{t=1}^{n}\sum_{k=1}^{J} C_{\beta\beta,k} z_k x_{1t}^2, \; n^{-3/2}\sum_{t=1}^{n}\sum_{k=1}^{J} C_{\beta\beta,k} z_k x_{1t} x_{2t}',$$
$$n^{-3/2}\sum_{t=1}^{n}\sum_{k=1}^{J} C_{\beta\beta,k} z_k x_{2t} x_{2t}', \; n^{-1}\sum_{t=1}^{n}\sum_{k=1}^{J} C_{\beta\mu^i,k} z_k x_{1t},$$
$$n^{-1}\sum_{t=1}^{n}\sum_{k=1}^{J} C_{\beta\mu^i,k} z_k x_{2t}', \; n^{-1/2}\sum_{t=1}^{n}\sum_{k=1}^{J} C_{\mu^j\mu^j,k} z_k$$

are all $o_p(1)$.

To get the stated results for $J_n(\underline{\theta}_0)$, we will proceed element by element. First, for $J_{n,11}(\underline{\theta}_0)$, we have

$$\begin{aligned} & n^{-3/2} J_{n,11}(\underline{\theta}_0) \\ = \quad & -n^{-3/2} \sum_{t=1}^{n} \sum_{k=1}^{J} \sum_{l=1}^{J} A_k A_l z_k z_l x_{1t}^2 + o_p(1), \end{aligned} \tag{55}$$

which, by Lemma B can be asymptotically validly approximated by its conditional expectation

$$-n^{-3/2} \sum_{t=1}^{n} \sum_{k=1}^{J} \sum_{l=1}^{J} A_k A_l \eta_{kl} x_{1t}^2. \tag{56}$$

The proof given in Lemma R1 above showed that (56) coverges weakly to

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

and thus

$$n^{-3/2} J_{n,11}(\underline{\theta}_0) \Rightarrow -\sum_{j=1}^{J} \left\{ \frac{(\mu_0^j)^2}{(\alpha_0^1)^3} L_1 \left(1, \frac{\mu_0^j}{\alpha_0^1}\right) \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1-F(s))} ds \right\} \tag{57}$$

The limits for $J_{n,12}(\underline{\theta}_0)$ and $J_{n,22}(\underline{\theta}_0)$ follow similarly from Lemma B and Lemma R0 parts (c)-(f), viz.,

$$\begin{aligned} & n^{-3/2} J_{n,12}(\underline{\theta}_0) \\ = \quad & -n^{-3/2} \sum_{t=1}^{n} \sum_{k=1}^{J} \sum_{l=1}^{J} A_k A_l z_k z_l x_{1t} x_{2t}' + o_p(1) \\ \Rightarrow \quad & -\sum_{j=1}^{J} \left\{ \frac{\mu_0^j}{(\alpha_0^1)^2} \int_0^1 dL_1 \left(r, \frac{\mu_0^j}{\alpha_0^1}\right) V_2(r)' \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1-F(s))} ds \right\}, \end{aligned} \tag{58}$$

$$\begin{aligned} & n^{-3/2} J_{n,22}(\underline{\theta}_0) \\ = \quad & -n^{-3/2} \sum_{t=1}^{n} \sum_{k=1}^{J} \sum_{l=1}^{J} A_k A_l z_k z_l x_{2t} x_{2t}' + o_p(1) \\ \Rightarrow \quad & -\sum_{j=1}^{J} \left\{ \frac{1}{\alpha_0^1} \int_0^1 V_2(r) V_2(r)' dL_1 \left(r, \frac{\mu_0^j}{\alpha_0^1}\right) \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1-F(s))} ds \right\}. \end{aligned} \tag{59}$$

Next, for $J_{n,13}(\underline{\theta}_0)$, we have

$$
\begin{aligned}
& n^{-3/2} J_{n,13}(\underline{\theta}_0)(j) \\
= \quad & -n^{-1} \sum_{t=1}^{n} \sum_{k=1}^{J} \sum_{l=1}^{J} A_k B_l z_k z_l x_{1t} + o_p(1) \\
= \quad & -n^{-1} \sum_{t=1}^{n} \sum_{k=1}^{J} \sum_{l=1}^{J} \sum_{i=1}^{J} \{ f(x_{1t}\alpha_0^1 - \sqrt{n}\mu_0^i) f(x_{1t}\alpha_0^1 - \sqrt{n}\mu_0^j) \\
& \times [g_k(i) - g_k(i-1)][g_l(j-1) - g_l(j)] z_{kt} z_{lt} x_{1t} \} \\
= \quad & -n^{-1} \sum_{t=1}^{n} \sum_{i=1, i \neq j}^{J} \{ f(x_{1t}\alpha_0^1 - \sqrt{n}\mu_0^i) f(x_{1t}\alpha_0^1 - \sqrt{n}\mu_0^j) \\
& \times \sum_{k=1}^{J} \sum_{l=1}^{J} [g_k(i) - g_k(i-1)][g_l(j-1) - g_l(j)] z_{kt} z_{lt} x_{1t} \} \\
& -n^{-1} \sum_{t=1}^{n} \left\{ f^2(x_{1t}\alpha_0^1 - \sqrt{n}\mu_0^j) \sum_{k=1}^{J} \sum_{l=1}^{J} [g_k(j) - g_k(j-1)][g_l(j-1) - g_l(j)] z_{kt} z_{lt} x_{1t} \right\},
\end{aligned} \tag{60}
$$

which can be asymptotically validly approximated by its conditional expectation

$$
n^{-3/2} \sum_{t=1}^{n} \sum_{k=1}^{J} \sum_{l=1}^{J} A_k B_l \eta_{kl} x_{1t},
$$

using Lemma B. Again, the conditional expectation of the first term of (60) is 0 by (50), and the conditional expectation of the second term behaves as

$$
\begin{aligned}
& -n^{-1} \sum_{t=1}^{n} \left\{ f^2(x_{1t}\alpha_0^1 - \sqrt{n}\mu_0^j) \left[\frac{1}{P_j} + \frac{1}{P_{j-1}} \right] x_{1t} \right\} \\
\Rightarrow \quad & -\frac{\mu_0^j}{(\alpha_0^1)^2} L_1 \left(1, \frac{\mu_0^j}{\alpha_0^1} \right) \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1-F(s))} ds,
\end{aligned}
$$

by Lemma R0 parts (a) and (b). Thus,

$$
\begin{aligned}
& n^{-3/2} J_{n,13}(\underline{\theta}_0)(j) \\
\Rightarrow \quad & -\frac{\mu_0^j}{(\alpha_0^1)^2} L_1 \left(1, \frac{\mu_0^j}{\alpha_0^1} \right) \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1-F(s))} ds.
\end{aligned} \tag{61}
$$

Similarly,

$$
\begin{aligned}
& n^{-3/2} J_{n,23}(\underline{\theta}_0)(j) \\
\Rightarrow \quad & \frac{1}{\alpha_0^1} \int_0^1 dL_1 \left(r, \frac{\mu_0^j}{\alpha_0^1} \right) V_2(r)' \int_{-\infty}^{\infty} \frac{f^2(s)}{F(s)(1-F(s))} ds.
\end{aligned} \tag{62}
$$

Finally, for $J_{n,23}(\underline{\theta}_0)(j,j)$

$$
\begin{aligned}
& n^{-3/2}J_{n,33}(\underline{\theta}_0)(j,j) \\
= & -n^{-1/2}\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J}B_k(j)B_l(j)z_kz_l+o_p(1) \\
\Rightarrow & \frac{1}{\alpha_0^1}L_1\left(1,\frac{\mu_0^j}{\alpha_0^1}\right)\int_{-\infty}^{\infty}\frac{f^2(s)}{F(s)(1-F(s))}ds,
\end{aligned} \tag{63}
$$

using Lemma R0 parts (a) and (b); and for $J_{n,23}(\underline{\theta}_0)(j,j-1)$ we have

$$
\begin{aligned}
& n^{-3/2}J_{n,33}(\underline{\theta}_0)(j,j-1) \\
= & -n^{-1/2}\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J}B_k(j)B_l(j-1)z_kz_l+o_p(1),
\end{aligned}
$$

the conditional expectation of which is

$$
\begin{aligned}
& n^{-1/2}\sum_{t=1}^{n}f(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^j)f(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^{j-1})\left(\frac{1}{P_{j-1}}\right) \\
= & n^{-1/2}\sum_{t=1}^{n}\frac{f(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^j)f(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^{j-1})}{F(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^{j-1})-F(x_{1t}\alpha_0^1-\sqrt{n}\mu_0^j)} \\
\Rightarrow & 0,
\end{aligned}
$$

as is shown in Lemma R0 part (g). Thus,

$$
n^{-3/2}J_{n,33}(\underline{\theta}_0)(j,j-1)\Rightarrow 0. \tag{64}
$$

By a similar argument and Lemma R0 part (h), we get

$$
n^{-3/2}J_{n,33}(\underline{\theta}_0)(j,j+1)\Rightarrow 0. \tag{65}
$$

Combining (57), (58), (59), (61), (62), (63), (64) and (65), we get the stated asymptotic results.

Proof of Theorem R3 As in Park and Phillips (2000), we can apply Theorem 10.1 of Wooldridge (1994) to show that (28) holds and thus there is a consistent local solution to the likelihood equation. The likelihood equation for the ML estimator $\hat{\underline{\theta}}_n$ is

$$
S_n(\hat{\underline{\theta}}_n)=0, \tag{66}
$$

which has the expansion

$$
0=S_n(\hat{\underline{\theta}}_n)=S_n(\underline{\theta}_0)+J_n(\underline{\theta}_n)(\hat{\underline{\theta}}_n-\underline{\theta}_0),
$$

or

$$
S_n(\underline{\theta}_0)+J_n(\underline{\theta}_0)(\hat{\underline{\theta}}_n-\underline{\theta}_0)+[J_n(\underline{\theta}_n)-J_n(\underline{\theta}_0)](\hat{\underline{\theta}}_n-\underline{\theta}_0)=0, \tag{67}
$$

where $S_n(\hat{\underline{\theta}}_n)$ and $S_n(\underline{\theta}_0)$ are the scores respectively at $\hat{\underline{\theta}}_n$ and $\underline{\theta}_0$, and $J_n(\underline{\theta}_n)$ is the hessian matrix with rows evaluated at mean values $\underline{\theta}_n$ that lie on the line segment connecting $\hat{\underline{\theta}}_n$ and $\underline{\theta}_0$. Then (67) can be written as

$$\begin{aligned} 0 &= n^{-3/4}S_n(\underline{\theta}_0) + [n^{-3/2}J_n(\underline{\theta}_0)]n^{-3/4}(\hat{\underline{\theta}}_n - \underline{\theta}_0) \\ &\quad + \left(n^{-3/2}[J_n(\underline{\theta}_n) - J_n(\underline{\theta}_0)]\right) n^{-3/4}(\hat{\underline{\theta}}_n - \underline{\theta}_0), \end{aligned}$$

or

$$\begin{aligned} 0 &= n^{-3/4}S_n(\underline{\theta}_0) + [n^{-3/2}J_n(\underline{\theta}_0)]n^{-3/4}(\hat{\underline{\theta}}_n - \underline{\theta}_0) + \\ &\quad + n^{-2\delta}(n^{-3/2+2\delta}[J_n(\underline{\theta}_n) - J_n(\underline{\theta}_0)])n^{-3/4}(\hat{\underline{\theta}}_n - \underline{\theta}_0), \end{aligned} \tag{68}$$

Equation (28) now follows from (68) if the final term of (68) is $o_p(1)$. This will be so, if condition (iii) (b) of Wooldridge's theorem holds. To show this condition holds, we need to establish that

$$\sup_{\left\{\underline{\theta}:\|n^{3/4-\delta}(\underline{\theta}-\underline{\theta}_0)\|\leq 1\right\}} \| n^{-3/2+2\delta}[J_n(\underline{\theta}) - J_n(\underline{\theta}_0)] \| = o_p(1). \tag{69}$$

Our proof involves looking at the components of the hessian. We therefore partition the hessian conformably with $\underline{\theta}$ as

$$J_n(\underline{\theta}) = \begin{pmatrix} J_{n,11}(\underline{\theta}) & J_{n,12}(\underline{\theta}) & J_{n,13}(\underline{\theta}) \\ J_{n,21}(\underline{\theta}) & J_{n,22}(\underline{\theta}) & J_{n,23}(\underline{\theta}) \\ J_{n,31}(\underline{\theta}) & J_{n,32}(\underline{\theta}) & J_{n,33}(\underline{\theta}) \end{pmatrix}.$$

Since the matrix is symmetric, we consider the upper-right triangular block:

$$\begin{aligned} J_{n,11}(\underline{\theta}) &= -\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} A_{kt}(\underline{\theta})A_{lt}(\underline{\theta})z_{kt}(\underline{\theta})z_{lt}(\underline{\theta})x_{1t}^2 + \sum_{t=1}^{n}\sum_{k=1}^{J} C_{\beta\beta,kt}(\underline{\theta})z_{kt}(\underline{\theta})x_{1t}^2, \quad (70) \\ J_{n,12}(\underline{\theta}) &= -\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} A_{kt}(\underline{\theta})A_{lt}(\underline{\theta})z_{kt}(\underline{\theta})z_{lt}(\underline{\theta})x_{1t}x_{2t}' + \sum_{t=1}^{n}\sum_{k=1}^{J} C_{\beta\beta,kt}(\underline{\theta})z_{kt}(\underline{\theta})x_{1t}x_{2t}', \end{aligned}$$

$$\begin{aligned} J_{n,22}(\underline{\theta}) &= -\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} A_{kt}(\underline{\theta})A_{lt}(\underline{\theta})z_{kt}(\underline{\theta})z_{lt}(\underline{\theta})x_{2t}x_{2t}' + \sum_{t=1}^{n}\sum_{k=1}^{J} C_{\beta\beta,kt}(\underline{\theta})z_{kt}(\underline{\theta})x_{2t}x_{2t}', \\ J_{n,13}(\underline{\theta})(j) &= -\sqrt{n}\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} A_{kt}(\underline{\theta})B_{lt}(\underline{\theta},j)z_{kt}z_{lt}x_{1t} + \sqrt{n}\sum_{t=1}^{n}\sum_{k=1}^{J} C_{\beta\mu^j,kt}(\underline{\theta})z_{kt}(\underline{\theta})x_{1t}, \\ J_{n,23}(\underline{\theta})(j) &= -\sqrt{n}\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} A_{kt}(\underline{\theta})B_{lt}(\underline{\theta},j)z_{kt}z_{lt}x_{2t}' + \sqrt{n}\sum_{t=1}^{n}\sum_{k=1}^{J} C_{\beta\mu^j,kt}(\underline{\theta})z_{kt}(\underline{\theta})x_{2t}', \\ J_{n,33}(\underline{\theta}) &= J_{n,22}(\underline{\theta})), \end{aligned}$$

where we define $f_t(\underline{\theta}) = f_t(x_{1t}\alpha^1 + x_{2t}\alpha^2; \sqrt{T}\mu)$ for any function $f : R \to R$ and further define $f_t(\underline{\theta}_0)$to be the value of the function f at $\underline{\theta}_0$.

For (69) to hold, it is sufficient that

$$\begin{aligned}
n^{-3/2+2\delta} &\quad \| \quad J_{n,11}(\underline{\theta}) - J_{n,11}(\underline{\theta}_0) \| \to_p 0, \\
n^{-3/2+2\delta} &\quad \| \quad J_{n,12}(\underline{\theta}) - J_{n,12}(\underline{\theta}_0) \| \to_p 0, \\
n^{-3/2+2\delta} &\quad \| \quad J_{n,22}(\underline{\theta}) - J_{n,22}(\underline{\theta}_0) \| \to_p 0, \\
n^{-3/2+2\delta} &\quad \| \quad J_{n,13}(\underline{\theta}) - J_{n,13}(\underline{\theta}_0) \| \to_p 0, \\
n^{-3/2+2\delta} &\quad \| \quad J_{n,23}(\underline{\theta}) - J_{n,23}(\underline{\theta}_0) \| \to_p 0, \\
n^{-3/2+2\delta} &\quad \| \quad J_{n,33}(\underline{\theta}) - J_{n,33}(\underline{\theta}_0) \| \to_p 0,
\end{aligned} \tag{71}$$

uniformly for all α_1, α_2, and μ satisfying

$$\begin{aligned}
\left|\alpha^1 - \alpha_0^1\right| &\leq n^{-3/4+\delta}, \quad \| \alpha^2 - \alpha_0^2 \| \leq n^{-3/4+\delta}, \\
\| &\quad \mu - \mu_0 \| \leq n^{-3/4+\delta},
\end{aligned} \tag{72}$$

for some $\delta > 0$. We show (71) one by one. From (70) we have

$$J_{n,11}(\underline{\theta}) - J_{n,11}(\underline{\theta}_0) = \Gamma_{n,11}(\underline{\theta}^*) + \Phi_{n,11}(\underline{\theta}^*), \tag{73}$$

where

$$\begin{aligned}
\Gamma_{n,11}(\underline{\theta}) &= -\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J}(A_{kt}A_{lt}\dot{z}_{kt}z_{lt}(\underline{\theta}))x_{1t}^2\left(x_{1t}(\alpha^1-\alpha_0^1) + x_{2t}'(\alpha^2-\alpha_0^2) - \sqrt{n}(\mu-\mu_0)\right), \\
\Phi_{n,11}(\underline{\theta}) &= \sum_{t=1}^{n}\sum_{k=1}^{J}(C_{\beta\beta,k}\dot{z}_{kt}(\underline{\theta}))x_{1t}^2\left(x_{1t}(\alpha^1-\alpha_0^1) + x_{2t}'(\alpha^2-\alpha_0^2) - \sqrt{n}(\mu-\mu_0)\right),
\end{aligned}$$

and $\underline{\theta}^*$ is on the line segment connecting $\underline{\theta}$ and $\underline{\theta}_0$. For any $f : R \to R$, define $f_{\sup}$ as

$$f_{\sup}(x;y) \equiv f_{\sup}^{\varepsilon}(x;y) = \sup_{|a-a_0|\leq\varepsilon}\ \sup_{|b-b_0|\leq\varepsilon} |f(x+a;y+b)|,$$

for $\varepsilon > 0$ given. As shown in Lemma D, $f_{\sup}(x_{1t}\alpha_0^1;\sqrt{n}\mu_0) \in F_I$ if $f \in F_I$. Since $\sup_{1\leq t\leq n} x_{1t}/\sqrt{n} = O_p(1)$, $\left|\alpha^1 - \alpha_0^1\right| \leq n^{-3/4+\delta}$, $\sup_{1\leq t\leq n} \| x_{2t} \| /\sqrt{n} = O_p(1)$, $\| \alpha^2 \| \leq n^{-3/4+\delta}$, $\| \mu - \mu_0 \| \leq n^{-3/4+\delta}$, and the fact that we have for any $\varepsilon > 0$

$$|f(\underline{\theta})| = \left|f(x_{1t}\alpha^1 + x_{2t}\alpha^2;\sqrt{n}\mu)\right| \leq f_{\sup}(x_{1t}\alpha_0^1;\sqrt{n}\mu_0) + o_p(1),$$

for large n, uniformly in $1 \leq t \leq n$.

By (72), we have

$$
\begin{aligned}
\| \Gamma_{n,11}(\underline{\theta}) \| \leq & \; n^{-3/4+\delta} \sum_{t=1}^{n} \left(\sum_{k=1}^{J} \sum_{l=1}^{J} (\eta_{kl} \dot{A}_k A_l) \right)_{\sup} (x_{1t}\alpha_0^1; \sqrt{n}\mu_0) \, |x_{1t}|^3 \\
& + n^{-3/4+\delta} \sum_{t=1}^{n} \left(\sum_{k=1}^{J} \sum_{l=1}^{J} (\eta_{kl} \dot{A}_k A_l) \right)_{\sup} (x_{1t}\alpha_0^1; \sqrt{n}\mu_0) \, |x_{1t}|^2 \, \| x_{2t} \| \\
& - n^{-1/4+\delta} \sum_{t=1}^{n} \left(\sum_{k=1}^{J} \sum_{l=1}^{J} (\eta_{kl} \dot{A}_k A_l) \right)_{\sup} (x_{1t}\alpha_0^1; \sqrt{n}\mu_0) \, |x_{1t}|^2 .
\end{aligned}
$$

It therefore follows from Lemma A (a) that

$$
\| \Gamma_{n,11}(\underline{\theta}) \| = O_p(n^{5/4+\delta}), \tag{74}
$$

uniformly in $\underline{\theta}$ satisfying (72). Using exactly the same argument we can deduce that

$$
\begin{aligned}
&\| \; \Gamma_{n,12}(\underline{\theta}) \| = O_p(n^{5/4+\delta}), \quad \| \Gamma_{n,22}(\underline{\theta}) \| = O_p(n^{5/4+\delta}), \\
&\| \; \Gamma_{n,13}(\underline{\theta}) \| = O_p(n^{5/4+\delta}) \quad \| \Gamma_{n,23}(\underline{\theta}) \| = O_p(n^{5/4+\delta}), \\
&\| \; \Gamma_{n,33}(\underline{\theta}) \| = O_p(n^{5/4+\delta}), \quad \| \Phi_{n,11}(\underline{\theta}) \| = O_p(n^{5/4+\delta}), \\
&\| \; \Phi_{n,12}(\underline{\theta}) \| = O_p(n^{5/4+\delta}), \quad \| \Phi_{n,22}(\underline{\theta}) \| = O_p(n^{5/4+\delta}), \\
&\| \; \Phi_{n,13}(\underline{\theta}) \| = O_p(n^{5/4+\delta}), \quad \| \Phi_{n,23}(\underline{\theta}) \| = O_p(n^{5/4+\delta}), \\
&\| \; \Phi_{n,33}(\underline{\theta}) \| = O_p(n^{5/4+\delta}),
\end{aligned} \tag{75}
$$

uniformly in $\underline{\theta}$ satisfying (72). If we let $0 < \delta < 1/12$, we may easily deduce (71) from (74), (75) and (73). Hence, (69) holds and therefore (28).

It now follows as in the proof of Theorem 10.1 of Wooldridge (1994) that there exists a solution to the likelihood equation (66) with probability approaching one such that

$$
n^{3/4}(\underline{\hat{\theta}}_n - \underline{\theta}_0) = O_p(1).
$$

From Theorem 1, we have the joint weak convergence

$$
\left(n^{-3/4} S_n(\underline{\theta}_0), n^{-3/2} J_n(\underline{\theta}_0) \right) \Rightarrow \left(Q^{1/2} W(1), -Q \right), \tag{76}
$$

where Q is positive definite with probability one. Thus, condition (iv) of Wooldridge's theorem holds. The given limit distribution of $n^{3/4}(\underline{\hat{\theta}}_n - \underline{\theta}_0)$ now follows from (68), (69) and (76).

Proof of Corollary R6 First consider $j = 0$ denote $P_{0,x} = P_0(x; \theta_0)$, and use the following mean value expansions for $\widehat{P}_{0,x} = \widehat{P}_0(x; \widehat{\theta}_n)$, given $x_t = x$, and $z_n = z$

$$\widehat{P}_{0,x} = P_0(x'\beta_0 - z\mu_0^1) + \left(\begin{array}{cc} \frac{\partial P_0(x'\beta_n - z\mu_n^1)}{\partial \beta'} & \frac{\partial P_0(x'\beta_n - z\mu_n^1)}{\partial \mu^1} \end{array} \right) \left(\begin{array}{c} \widehat{\beta}_n - \beta_0 \\ \widehat{\mu}_n^1 - \mu_0^1 \end{array} \right),$$

where β_n and μ_n^1 are on line segements joining $\widehat{\beta}_n$ and β_0 and $\widehat{\mu}_n^1$ and μ_0^1, respectively, and where the derivatives have the form

$$\begin{aligned}
\frac{\partial P_0(x'\beta_n - z\mu_n^1)}{\partial \beta} &= -f((x'\beta_n - z\mu_n^1))x \sim -f(x'\beta_0 - z\mu_0^1)x, \\
\frac{\partial P_0(x'\beta_n - z\mu_n^1)}{\partial \mu^1} &= f(x'\beta_n - z\mu_n^1)z \sim f(x'\beta_0 - z\mu_0^1)z,
\end{aligned}$$

and

$$\left(\begin{array}{c} \widehat{\beta}_n - \beta_0 \\ \widehat{\mu}_n^1 - \mu_0^1 \end{array} \right) = R(0) \left(\begin{array}{c} \widehat{\beta}_n - \beta_0 \\ \widehat{\mu}_n - \mu_0 \end{array} \right). \tag{77}$$

Then

$$\begin{aligned}
& n^{3/4} \left(\widehat{P}_{0,x} - P_0((x'\beta_0 - z\mu_0^1)) \right) \\
= \; & f(x'\beta_0 - z\mu_0^1) \left(\begin{array}{c} -x \\ z \end{array} \right)' R(0) \left(\begin{array}{c} n^{3/4}(\widehat{\beta}_n - \beta_0)) \\ n^{3/4}(\widehat{\mu}_n - \mu_0) \end{array} \right) \\
\Rightarrow \; & f(x'\beta_0 - z\mu_0^1) \left(\begin{array}{c} -x \\ z \end{array} \right)' R(0) MN(0, GQ^{-1}G') \\
= \; & MN(0, \Upsilon(0) GQ^{-1}G' \Upsilon(0)').
\end{aligned}$$

Thus,

$$n^{3/4} \left(\widehat{P}_{0,x} - P_{0,x} \right) \Rightarrow MN \left(0, \Upsilon(0)(GQG')^{-1}\Upsilon(0)' \right),$$

where

$$\Upsilon(0) = f(x'\beta_0 - z\mu_0^1) \left(\begin{array}{c} -x \\ z \end{array} \right)' R(0).$$

Similarly, for $j = J$, denote $P_{J,x} = P_J(x; \theta_0)$, we have

$$\begin{aligned}
& n^{3/4} \left(\widehat{P}_{J,x} - P_J(x'\beta_0 - z\mu_0^J) \right) \\
= \; & f(x'\beta_0 - z\mu_0^J) \left(\begin{array}{c} x \\ -z \end{array} \right)' R(J) \left(\begin{array}{c} n^{3/4}(\widehat{\beta}_n - \beta_0)) \\ n^{3/4}(\widehat{\mu}_n - \mu_0) \end{array} \right) \\
\Rightarrow \; & f(x'\beta_0 - z\mu_0^J) \left(\begin{array}{c} x \\ -z \end{array} \right)' R(J) MN(0, GQ^{-1}G') \\
= \; & MN(0, \Upsilon(J) GQ^{-1}G' \Upsilon(J)'),
\end{aligned}$$

where

$$\Upsilon(J) = f(x'\beta_0 - z\mu_0^J) \left(\begin{array}{c} x \\ -z \end{array} \right)' R(J).$$

Now for $1 \leq j \leq J-1$, denote $P_{j,x} = P_j(x;\theta_0) = F(x'\beta_0 - z\mu_0^j) - F(x'\beta_0 - z\mu_0^{j+1})$, notice that

$$\widehat{P}_{j,x} = P_{j,x} + \begin{pmatrix} \frac{\partial P_j(x;\theta_n)}{\partial \beta_n'} & \frac{\partial P_j(x;\theta_n)}{\partial \mu_n^j} & \frac{\partial P_j(x;\theta_n)}{\partial \mu_n^{j+1}} \end{pmatrix} \begin{pmatrix} \widehat{\beta}_n - \beta_0 \\ \widehat{\mu}_n^j - \mu_0^j \\ \widehat{\mu}_n^{j+1} - \mu_0^{j+1} \end{pmatrix},$$

and

$$\begin{aligned}
\frac{\partial P_j(x;\theta_n)}{\partial \beta_n} &= p_j(x;\theta_n)x = \left[f(x'\beta_n - z\mu_n^j) - f(x'\beta_n - z\mu_n^{j+1})\right] x \\
&\sim \left[f(x'\beta_0 - z\mu_0^j) - f(x'\beta_0 - z\mu_0^{j+1})\right] x, \\
\frac{\partial P_j(x;\theta_n)}{\partial \mu_n^j} &= -f(x'\beta_0 - z\mu_0^j)z, \\
\frac{\partial P_j(x;\theta_n)}{\partial \mu_n^{j+1}} &= f(x'\beta_0 - z\mu_0^{j+1})z,
\end{aligned}$$

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

Thus, we have

$$\begin{aligned}
& n^{3/4}\left(\widehat{P}_{j,x} - P_{j,x}\right) \\
\sim & \begin{pmatrix} \left[f(x'\beta_0 - z\mu_0^j) - f(x'\beta_0 - z\mu_0^{j+1})\right] x \\ -f(x'\beta_0 - z\mu_0^j)z \\ f(x'\beta_0 - z\mu_0^{j+1})z \end{pmatrix}' R(j) \begin{pmatrix} n^{3/4}(\widehat{\beta}_n - \beta_0)) \\ n^{3/4}(\widehat{\mu}_n - \mu_0) \end{pmatrix} \\
\Rightarrow & \begin{pmatrix} \left[f(x'\beta_0 - z\mu_0^j) - f(x'\beta_0 - z\mu_0^{j+1})\right] x \\ -f(x'\beta_0 - z\mu_0^j)z \\ f(x'\beta_0 - z\mu_0^{j+1})z \end{pmatrix}' R(j) MN(0, GQ^{-1}G') \\
= & MN(0, \Upsilon(j)GQ^{-1}G'\Upsilon'(j)),
\end{aligned}$$

where

$$\Upsilon(j) = \begin{pmatrix} \left[f(x'\beta_0 - z\mu_0^j) - f(x'\beta_0 - z\mu_0^{j+1})\right] x \\ -f(x'\beta_0 - z\mu_0^j)z \\ f(x'\beta_0 - z\mu_0^{j+1})z \end{pmatrix}' R(j).$$

When stationary variates are present, the proof is similar. First consider $j = 0$, denote $P_{0,x} = P_0(x, x_3; \rho_0)$, and use the following mean value expansions for $\widehat{P}_{0,x} = \widehat{P}_0(x, x_3; \widehat{\rho}_n)$, given $x_{3t} = x_3$, $x_t = x$, and $z_n = z$,

$$\widehat{P}_{0,x} = P_0(x_3'\gamma_0 + x'\beta_0 - z\mu_0^1)$$
$$+\left(\begin{array}{ccc} \frac{\partial P_0(x_3'\gamma_n + x'\beta_n - z\mu_n^1)}{\partial\gamma'} & \frac{\partial P_0(x_3'\gamma_n + x'\beta_n - z\mu_n^1)}{\partial\beta'} & \frac{\partial P_0(x_3'\gamma_n + x'\beta_n - z\mu_n^1)}{\partial\mu^1} \end{array} \right) \left(\begin{array}{c} \widehat{\gamma}_n - \gamma_0 \\ \widehat{\beta}_n - \beta_0 \\ \widehat{\mu}_n^1 - \mu_0^1 \end{array} \right),$$

where γ_n, β_n and μ_n^1 are on line segements joining $\widehat{\gamma}_n$ and γ_0, $\widehat{\beta}_n$ and β_0 and $\widehat{\mu}_n^1$ and μ_0^1, respectively, and where the derivatives have the form

$$\begin{aligned}
\frac{\partial P_0(x_3'\gamma_n + x'\beta_n - z\mu_n^1)}{\partial\gamma} &= -f(x_3'\gamma_n + x'\beta_n - z\mu_n^1)x_3 \sim f(x_3'\gamma_0 + x'\beta_0 - z\mu_0^1)x_3, \\
\frac{\partial P_0(x_3'\gamma_n + x'\beta_n - z\mu_n^1)}{\partial\beta} &= -f((x_3'\gamma_0 + x'\beta_n - z\mu_n^1))x \sim -f(x_3'\gamma_0 + x'\beta_0 - z\mu_0^1)x, \\
\frac{\partial P_0(x_3'\gamma_n + x'\beta_n - z\mu_n^1)}{\partial\mu^1} &= f(x_3'\gamma_0 + x'\beta_n - z\mu_n^1)z \sim f(x_3'\gamma_0 + x'\beta_0 - z\mu_0^1)z,
\end{aligned}$$

and we define a matrix $R_3(0) = Diag(I_{m+m_3}, \iota_1')$ where ι_j is a vector of length J with the j'th element 1 and other elements zero. Accordingly, we may write

$$\left(\begin{array}{c} \widehat{\gamma}_n - \alpha_0 \\ \widehat{\beta}_n - \beta_0 \\ \widehat{\mu}_n^1 - \mu_0^1 \end{array} \right) = R_3(0) \left(\left(\begin{array}{c} \widehat{\gamma}_n - \alpha_0 \\ \widehat{\beta}_n - \beta_0 \\ \widehat{\mu}_n - \mu_0 \end{array} \right) \right). \tag{78}$$

Then

$$\begin{aligned}
& n^{1/4}\left(\widehat{P}_{0,x} - P_0((x_3'\gamma_0 + x'\beta_0 - z\mu_0^1))\right) \\
= \quad & f(x_3'\gamma_0 + x'\beta_0 - z\mu_0^1) \left(\begin{array}{c} -x_3 \\ -x \\ z \end{array} \right)' R_3(0) \left(\begin{array}{c} n^{1/4}(\widehat{\gamma}_n - \gamma_0) \\ n^{1/4}(\widehat{\beta}_n - \beta_0) \\ n^{1/4}(\widehat{\mu}_n - \mu_0) \end{array} \right) \\
\sim \quad & f(x_3'\gamma_0 + x'\beta_0 - z\mu_0^1) \left(\begin{array}{c} -x_3 \\ -x \\ z \end{array} \right)' R_3(0) \left(\begin{array}{c} n^{1/4}(\widehat{\gamma}_n - \gamma_0) \\ o_p(1) \\ o_p(1) \end{array} \right) \\
\sim \quad & -f(x_3'\gamma_0 + x'\beta_0 - z\mu_0^1)x_3' \left\{ n^{1/4}(\widehat{\gamma}_n - \gamma_0) \right\} + o_p(1) \\
\Rightarrow \quad & MN\left(0, f(x_3'\gamma_0 + x'\beta_0 - z\mu_0^1)^2 x_3'\Xi_{11}^{-1}x_3\right) ..
\end{aligned}$$

The proof when $j = 1, 2, ..., J-1$, and $j = J$ is similar and is omitted.

Proof of Corollary R7 For $j = 0$, denote $\upsilon_{0,x} = \upsilon_0(x;\theta_0) = p_0(x;\theta_0)\beta_0$, and we use the following mean value expansions for $\widehat{\upsilon}_{0,x} = \widehat{p}_0(x_t;\widehat{\theta}_n)\widehat{\beta}_n$, given $x_t = x$, and $z_n = z$

$$\widehat{\upsilon}_{0,x} = \upsilon_{0,x} + \left(\begin{array}{cc} \frac{\partial\upsilon_0((x'\beta_n - z\mu_n^1))}{\partial\beta'} & \frac{\partial\upsilon_0((x'\beta_n - z\mu_n^1))}{\partial\mu^1} \end{array} \right) \left(\begin{array}{c} \widehat{\beta}_n - \beta_0 \\ \widehat{\mu}_n^1 - \mu_0^1 \end{array} \right),$$

where β_n and μ_n^1 are on line segements joining $\widehat{\beta}_n$ and β_0 and $\widehat{\mu}_n^1$ and μ_0^1, respectively, and where the derivatives have the form

$$\begin{aligned}
\frac{\partial v_0(x;\theta_n)}{\partial\beta'} &= -\dot{f}((x'\beta_n - z\mu_n^1))\beta_n x' - f((x'\beta_n - z\mu_n^1))I_m \\
&\sim -\dot{f}((x'\beta_0 - z\mu_0^1))\beta_0 x' - f((x'\beta_0 - z\mu_0^1))I_m, \\
\frac{\partial v_0(x;\theta_n)}{\partial\mu^1} &\sim \dot{f}((x'\beta_0 - z\mu_0^1))z\beta_0.
\end{aligned}$$

Then

$$n^{3/4}\left(\widehat{v}_{0,x} - v_{0,x}\right) \sim_d MN\left(0, \Pi(0)GQ^{-1}G'\Pi(0)'\right),$$

where

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

and

$$\dot{f}((x'\beta_0 - z\mu_0^j)) = -f((x'\beta_0 - z\mu_0^1))(x'\beta_0 - z\mu_0^1).$$

Similarly, for $j = J$,

$$n^{3/4}\left(\widehat{v}_{J,x} - v_{J,x}\right) \sim_d MN\left(0, \Pi(J)GQ^{-1}G'\Pi(J)'\right),$$

where

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

Now for $1 \le j \le J-1$, define

$$v_j = v_j(x;\theta_0) = p_j(x;\theta_0)\beta_0 = \left[f((x'\beta_0 - z\mu_0^j)) - f((x'\beta_0 - z\mu_0^{j+1}))\right]\beta_0,$$

so that

$$\widehat{v}_j = v_j + \begin{pmatrix} \frac{\partial v_j(x;\theta_n)}{\partial\beta_n'} & \frac{\partial v_j(x;\theta_n)}{\partial\mu_n^j} & \frac{\partial v_j(x;\theta_n)}{\partial\mu_n^{j+1}} \end{pmatrix} \begin{pmatrix} \widehat{\beta}_n - \beta_0 \\ \widehat{\mu}_n^j - \mu_0^j \\ \widehat{\mu}_n^{j+1} - \mu_0^{j+1} \end{pmatrix},$$

and we have

$$\begin{aligned}
\frac{\partial v_j(x;\theta_n)}{\partial\beta'} &= \dot{p}_j(x;\theta_n)\beta_n x' + p_j(x;\theta_n)I_m \\
&= \left[\dot{f}((x'\beta_n - z\mu_n^j)) - \dot{f}((x'\beta_n - z\mu_n^{j+1}))\right]\beta_n x' + p_j(x;\theta_n)I_m \\
&\sim \left[\dot{f}((x'\beta_0 - z\mu_0^j)) - \dot{f}((x'\beta_0 - z\mu_0^{j+1}))\right]\beta_0 x' + p_j(x;\theta_0)I_m, \\
\frac{\partial v_j(x;\theta_n)}{\partial\mu^j} &= -\dot{f}((x'\beta_n - z\mu_n^j))z\beta_n \\
&\sim -\dot{f}((x'\beta_0 - z\mu_0^j))z\beta_0, \\
\frac{\partial v_j(x;\theta_n)}{\partial\mu^{j+1}} &= \dot{f}((x'\beta_n - z\mu_n^{j+1}))z\beta_n, \\
&\sim \dot{f}((x'\beta_0 - z\mu_0^{j+1}))z\beta_0,
\end{aligned}$$

$$\left(\begin{array}{c}\widehat{\beta}_n-\beta_0\\ \widehat{\mu}_n^j-\mu_0^j\\ \widehat{\mu}_n^{j+1}-\mu_0^{j+1}\end{array}\right)=R(j)\left(\begin{array}{c}\widehat{\beta}_n-\beta_0\\ \widehat{\mu}_n-\mu_0\end{array}\right).$$

Thus,

$$n^{3/4}\left(\widehat{v}_{j,x}-v_{j,x}\right)\sim_d MN\left(0,\Pi(j)GQ^{-1}G'\Pi(j)'\right),$$

where

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

The proof when we have stationary variates is similar and is omitted.

8 References

Akonom, J. (1993): "Comportement asymptotique du temps d'occupation du processus des sommes partielles," *Annales de l'Institut Henri Poincaré* **29**, 57–81.

Hu, L. and P. C. B. Phillips (2004): "Nonstationary Discrete Choice," *Journal of Econometrics,* vol. 120, 103-138.

Hu. L. and P. C. B. Phillips (2004): "Dynamics of the Federal Funds Target Rate: A Nonstationary Discrete Choice Approach," J*ournal of Applied Econometric*s, Vol. 19, 2004, pp. 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. and P. C. B. Phillips (1999): "Asymptotics for Nonlinear Transformations of Integrated Time Series," *Econometric Theory,* 15, 269-298.

Park, J.Y. and P. C. B. Phillips (2000): "Nonstationary Binary Choice," *Econometrica,* 68, 1249-1280.

Park, J.Y. and P. C. B. Phillips (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*, Vol. 64, No. 1, January 2005.

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. and J. Y. Park (1998): "Nonstationary Density Estimation and Kernel Autoregression," Yale University, Cowles Foundation Discussion Paper No 1181.

Revuz, D. and M. Yor (1994), *Continuous Martingale and Brownian Motion*, 2nd ed. Springer-Verlag: New York.

Wooldridge, J. M. (1994): "Estimation and Inference for Dependent Processes," In R. F. Engle and D. L. McFadden (Eds.) *Handbook of Econometrics*. Vol. IV. North Holland: Amsterdam.