NONSTATIONARY DISCRETE CHOICE

By

Ling Hu and Peter C.B. Phillips

May 2002

COWLES FOUNDATION DISCUSSION PAPER NO. 1364

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

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

Nonstationary Discrete Choice*

Ling Hu and Peter C. B. Phillips
Department of Economics
Yale University

September 9, 2001

Abstract

This paper develops an asymptotic theory for time series discrete choice models with explanatory variables generated as integrated processes and with multiple choices and threshold parameters determining the choices. The theory extends recent work by Park and Phillips (2000) on binary choice models. As in this earlier work, the maximum likelihood (ML) estimator is consistent and has a limit theory with multiple rates of convergence ($n^{3/4}$ and $n^{1/4}$) and mixture normal distributions where the mixing variates depend on Brownian local time as well as Brownian motion. An extended arc sine limit law is given for the sample proportions of the various choices. The new limit law exhibits a wider range of potential behavior that depends on the values taken by the threshold parameters.

Keywords: Brownian motion, Brownian local time, Discrete choice model, Dual convergence rates, Extended arc sine laws, Integrated time series, Maximum likelihood estimation, Threshold parameters.

JEL Classification Numbers: C22, C25.

*Ling Hu thanks the Cowles Foundation for the support of a Cowles Fellowship. Peter Phillips thanks the NSF for support under Grant Nos. SBR 97-30295 and SES 0092509. The paper was typed by the authors in SW2.5 and computational work was performed in GAUSS.

1 Introduction

While it is often convenient to assume continuous dependent variables in time series applications, a discrete dependent variable approach is also useful. For example, recent monetary policy models allow for the determination of an optimal policy rule by a central bank, given certain objectives relating to inflation and economic growth. In such models, the 'optimal' interest rate is determined as a continuous function of other economic variables, much as the Fisher relationship links the real rate, expected inflation and the nominal rate of interest in a continuous way. However, in practice, central banks like the Federal Reserve implement policy by intervening in the money market to achieve a target level for a short term interest rate, like the Federal Funds rate in the case of the US. By convention, this target level is adjusted in a discrete way by the monetary authority. In the US, the policy-making Federal Open Market Committee (FOMC) has regularly scheduled meetings eight times a year to direct the conduct of open market operations. Decisions at these (and other unscheduled) meetings raise the target rate, cut the target rate, or leave it unchanged. Such policy decisions are well suited to discrete choice model formulations. In addition to such macroeconomic applications, time series discrete choice models are a natural tool for modeling individual agent participation behavior over time in financial markets, markets for durable goods, and labour markets. Discrete dependent variable models are also applicable in modeling ordered data, such as ratings of bonds and stocks.

Inference in binary and multiple choice models is a standard topic covered in many econometric texts. But in the time series applications just mentioned, the covariates typically involve nonstationary data. For instance, the macroeconomic fundamentals underlying decisions by the FOMC, the history of stock prices underlying financial investment decisions, and the time profile of household income that affects labour market participation decisions may all be expected to have nonstationary characteristics. In such situations, the asymptotic theory of inference in discrete choice models may be expected to have some differences from that of the traditional cross-section textbook theory. The present paper is concerned to develop such an asymptotic theory at a level of generality that will make it useful in practical work, extending recent work of Park and Phillips (2000).

Park and Phillips developed a new limit theory for maximum likelihood (ML) estimation of a binary choice model where the covariates are integrated processes whose coefficients (β) are being estimated. One major finding in their work is that there are two convergence rates for the coefficient estimator. There is a fast rate of convergence of $n^{3/4}$ in a direction that is orthogonal to that of the true coefficient vector and a slower rate of convergence of $n^{1/4}$ in other directions. This result, which differs substantially from the stationary case, is the direct outcome of the nonlinear functions of integrated variables that arise in discrete choice modeling. Park and Phillips (2000) found further that the sample proportion of binary choices follows an arc sine law asymptotically. This result is also very different from the stationary case, where a law of large numbers holds and the limit proportion is a constant. When applied to market intervention data (such as central bank monetary policy intervention) the Park-Phillips arc sine limit law indicates that policy is likely to occur in streams of intervention or no intervention, rather than more irregular policy shifting.

The present work extends this research to a framework that is better suited to empirical applications. In particular, we allow for multiple discrete choices and parameterize the choice settings. These extensions mean that our theory accommodates more interesting empirical examples like the FOMC policy decisions on intervention, where there are three outcomes (rate cut, rate hike, or no-change) and it involves estimable parameters (μ) that set the thresholds determining the various choices. The main conclusions of our work are consistent with the binary case. We provide a limit theory for ML estimation in the discrete choice model, giving asymptotics for both the regression coefficient estimator $\hat{\beta}_n$ and the threshold estimator $\hat{\mu}_n$. We find a convergence rate of $n^{3/4}$ for $\hat{\mu}_n$, in contrast to the $n^{1/2}$ rate that applies in the stationary case and we find that, although $\hat{\beta}_n$ and $\hat{\mu}_n$ have different convergent rates in multiple choice models with integrated regressors, they are in general asymptotically dependent. We also provide an asymptotic theory for the sample proportions of the various choices and find an 'extended arc sine' limit law that these sample proportions follow. This limit law permits much more flexibility than the binary case and it is better suited for empirical implementation. For instance, in the case of market intervention, it seems particularly useful to be able to estimate the thresholds that determine decisions.

The paper is organized as follows. Section 2 outlines the model, assumptions and gives some preliminary results. Section 3 gives the main results on the limit theory of the ML estimator. Section 4 considers the case where the covariates may have a deterministic trend. Section 5 illustrates the effects of nonstationarity on estimation and Section 6 concludes. Some useful lemmas are given in Appendix A, Appendix B gives proofs of the main theorems, Appendix C summarizes notation and Appendix D lists various special functions that are used in the paper.

2 The Model, Assumptions, and Preliminary Results

Our set up is analogous to that of Park and Phillips (2000), but we allow for polychotomous choice. In particular, we consider the regression model given by

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

where x_t is a $(m \times 1)$ vector of explanatory variables and ϵ_t is an error. The dependent variable y_t^* in (1) 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. Some modifications to this specification may be needed when the covariates also have deterministic trends and this is discussed later in the paper.

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 probability distribution of y_t, $P(y_t = j) = P_j(x_t; \theta_0)$ 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:

$$\begin{aligned} m(x_t;\theta_0) = E(y_t|\mathcal{F}_{t-1}) &= \sum_{j=0}^{J} j \cdot P_j(x_t;\theta_0) \\ &= \sum_{j=1}^{J} F(x_t'\beta_0 - \sqrt{n}\mu_0^j). \end{aligned}$$

Throughout this work, let $f_t = f(x_t; \theta_0)$ for any function $f(x_t; \theta)$ evaluated at the true value θ_0. 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 $z_k(x_t;\theta_0) = u_t^k - \sigma_{kt}$. Then, $(z_{kt}, \mathcal{F}_t)$ is also a martingale difference with conditional second moments $\eta_{kl}(x_t;\theta_0) = E(z_{kt} \cdot z_{kl}|\mathcal{F}_{t-1})$. Obviously, $\sigma_{1t} = 0$ and $z_{1t} = u_t$. Further, define $\tau_{kl,t} = E(z_{kt}z_{lt} - \eta_{kl,t})^2$, giving fourth conditional moments for z_{kt}.

For our asymptotic development we need more precise assumptions on the process generating x_t and the following condition is helpful. In particular, the linear process structure and the moment conditions on the innovations assist in the use of embedding arguments that allow for a stochastic process representation of key partial sum processes, as in Lemma 1 below, which was given in Park and Phillips (2000).

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* $\kappa > 0$*.*

Lemma 1 *Let Assumption 1 hold. Then there exists a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$ *supporting sequences of random variables* U_{nt} *and* V_{nt} *satisfying the following:*

(a) *Jointly for all* $1 \le t \le n$*,*

$$(U_{1,nt}, \ldots, U_{k,nt}, \ldots, U_{K,nt}, V_{nt}) =_d \left(\frac{1}{\sqrt{n}} \sum_{i=1}^t z_{1i}, \ldots, \frac{1}{\sqrt{n}} \sum_{i=1}^t z_{ki}, \ldots, \frac{1}{\sqrt{n}} \sum_{i=1}^t z_{Ki}, \frac{1}{\sqrt{n}} \sum_{i=1}^t v_i \right).$$

(b) *For* $k = 1, \ldots, K$*, there exists a representation*

$$U_{k,nt} = U\left(\frac{T_{k,nt}}{n}\right),$$

with standard Brownian motion U_k *and time changes* $T_{k,nt}$ *in* $(\Omega, \mathcal{F}, \mathbf{P})$*. Let* $T_{k,nt} = \sum_{i=1}^t \zeta_{k,ni}$ *and define* $\mathcal{F}_{nt} = \sigma((U_k(r))_{r=1}^{T_{k,nt}/n}, (V_{ns})_{s=1}^{t+1})$*. Then* $E(\zeta_{k,nt}|\mathcal{F}_{n,t-1}) = E(z_{k,t}^2|\mathcal{F}_{t-1})$ *and* $E(\zeta_{k,nt}^r|\mathcal{F}_{n,t-1}) \le c_r E(|z_t|^{2r}|\mathcal{F}_{t-1})$ *for all* $r \ge 1$*, where* c_r *is some constant depending only upon* r*.*

(c) *Defining*

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

then $V_n \to_{a.s.} V$ *in* $D[0,1]^m$*, the m-fold Cartesian product of the space* $D[0,1]$ *endowed with the uniform topology, where* V *is Brownian motion in* $(\Omega, \mathcal{F}, \mathbf{P})$ *with variance matrix* Σ*.*

As in Park and Phillips (2000), 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 $(\alpha_0^1, \alpha_0^{2'})' = \alpha_0 = H'\beta_0$. Then we can write (1) as:

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

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) for discussion 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

$$\Lambda(t, j) = \frac{\prod_{i=0,\ldots,J \ \& \ i \neq j} (y_t - i)}{\prod_{i=0,\ldots,J \ \& \ i \neq j} (j - i)}. \tag{4}$$

It is easy to verify that $\Lambda(t, j) = 1\{y_t = j\}$, the indicator function ($\Lambda(t, j) = 1$ if $y_t = j$ and $\Lambda(t, j) = 0$ otherwise). 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

$$\frac{\partial \log L_n}{\partial \beta} = \sum_{t=1}^{n} \sum_{j=0}^{J} \frac{\Lambda(t, j)}{P_j(x_t; \theta)} p_j(x_t; \theta) x_t \tag{5}$$

$$\frac{\partial \log L_n}{\partial \mu^j} = \sqrt{n} \sum_{t=1}^{n} \left(\frac{\Lambda(t, j-1)}{P_{j-1}(x_t; \theta)} - \frac{\Lambda(t, j)}{P_j(x_t; \theta)} \right) f(x_t' \beta - \sqrt{n} \mu^j) \tag{6}$$

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 (5) and (6). 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:

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

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

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

Therefore, in a binary choice case, $g_1(x_t;0,\theta_0) = -1/(1-F)$ and similarly, $g_1(x_t;1,\theta_0) = 1/F$. Using the above results, rewrite the score functions (5) and (6) as

$$\frac{\partial \log L_n}{\partial \beta} = \sum_{t=1}^{n}\sum_{k=1}^{J} A_k(x_t;\theta) z_k(x_t;\theta) x_t, \tag{7}$$

$$\frac{\partial \log L_n}{\partial \mu^j} = \sqrt{n}\sum_{t=1}^{n}\sum_{k=1}^{J} B_k(x_t;j,\theta) z_k(x_t;\theta), \tag{8}$$

where

$$A_k(x_t;\theta) = \sum_{j=0}^{J} g_k(x_t;j,\theta)p_j(x_t;\theta),$$

and

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

Again, in the binary choice example, it is easy to see that $A(x_t;\theta) = f/(1 - F)$ 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

$$J_n(\theta) = \begin{pmatrix} J_{n,11}(\theta) & J_{n,12}(\theta) \\ J_{n,21}(\theta) & J_{n,22}(\theta) \end{pmatrix}, \tag{9}$$

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

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

We show in the next section that the Hessian matrix has elements with different stochastic orders and the matrix converges to a random limit matrix after proper normalization.

The ML estimator involves nonlinear functions of the integrated process x_t and it is helpful to be specific about the functions we need to consider. In the analysis below, we use the approach of Park and Phillips (1999) in studying nonlinear transformations of integrated processes. A function $f : \mathbf{R} \to \mathbf{R}$ is called *regular* if it is bounded, integrable, and differentiable with bounded derivative. We denote by $\mathbf{F}_R$ the class of regular functions. We also consider the class $\mathbf{F}_I$ of bounded and integrable functions and the class $\mathbf{F}_0$ of functions that are bounded and vanish at infinity. Clearly, $\mathbf{F}_R \subset \mathbf{F}_I \subset \mathbf{F}_0$. We make the following assumption about the distribution F of ϵ_t.

Assumption 2

F is three times differentiable. Further, for $k, l = 1, \ldots, J$*:*

(a) $\eta_{kl}A_kB_l, \eta_{kl}A_kA_l, \eta_{kl}B_kB_l \in \mathbf{F}_R$

(b) $\eta_{kk}A_k, \eta_{kk}B_k \in \mathbf{F}_I$

(c) $\tau_{kk}A_k^2, \tau_{kk}B_k^2, \eta_{kk}C_k \in \mathbf{F}_0$

To check whether probit and logit function satisfy these assumptions, note that each of the products in condition (a) are sums of terms in the form of $F(x)^k \frac{f(x)^2}{P(x)}$. (For example, in the binary choice case, $\eta A^2 = F \cdot f^2/(1-F), \eta B^2 = f^2/(F(1-F))$ and $\eta AB = -f^2/(1-F)$). In a logit model, $F(x) = e^x/(1+e^x)$. Let $P(x) = 1 - F(x)$ (as in the case of $P_0(x)$), then

$$F(x)^k \frac{f(x)^2}{1-F(x)} = \frac{e^{(k+2)x}}{(1+e^x)^{k+3}},$$

which is clearly bounded, integrable and differentiable and actually its derivative goes to zero as $x \to \pm\infty$. Therefore, condition (a) is satisfied. Conditions (b) and (c) can be checked in the same way.

In the probit model, let $\varphi(x)$ denote the standard normal density function and $\Phi(x)$ denote the corresponding cumulative distribution function. Note that $\Phi(x)^k\varphi(x)^2/(1-\Phi(x)) \leq \varphi(x)^2/(\Phi(x)(1-\Phi(x))$, and using Mills ratio we have

$$\frac{\varphi(x)^2}{\Phi(x)(1-\Phi(x))} = \begin{cases} \frac{x\varphi(x)}{\Phi(x)}[1+O(x^{-2})] & \text{if } x \to \infty \\ \frac{-x\varphi(x)}{1-\Phi(x)}[1+O(x^{-2})] & \text{if } x \to -\infty \end{cases},$$

so $\varphi(x)^2/(\Phi(x)(1-\Phi(x))$ is regular and satisfies condition (a). Therefore, $\Phi(x)^k\varphi(x)^2/(1-\Phi(x))$ also satisfies condition (a) and its derivative actually goes to 0 as $x \to \pm\infty$. Conditions (b) and (c) follow similarly upon some further calculations.

3 Main results

Let $\hat{\theta}_n = (\hat{\beta}_n', \hat{\mu}_n')'$ be the maximum likelihood estimator of $\theta_0 = (\beta_0', \mu_0')'$ in (1) and (2). As usual in ML limit theory, the asymptotic distribution of $\hat{\theta}_n$ will be obtained from the expansion

$$0 = S_n(\hat{\theta}_n) = S_n(\theta_0) + J_n(\tilde{\theta}_n)(\hat{\theta}_n - \theta_0), \tag{10}$$

or in partititioned form

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

where $\tilde{\theta}$ is on the line segment between $\hat{\theta}_n$ and θ_0. Corresponding to the rotation in the regressors and parameters, 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 parameter are obtained from $S_n(\underline{\theta}) = G'S_n(\theta)$ and $J_n(\underline{\theta}) = G'J_n(\theta)G$. Pre-multiplying (10) by G' we have:

$$0 = S_n(\underline{\hat{\theta}}_n) = S_n(\underline{\theta}_0) + J_n(\underline{\tilde{\theta}}_n)(\underline{\hat{\theta}}_n - \underline{\theta}_0). \tag{11}$$

The next two lemmas provide a limit theory for sample moments and covariance functions which assist in analyzing the asymptotic behavior of the score function (7), (8) and Hessian (9). These are analogous to similar results in Park and Philllips (2000).

Lemma 2 *Let Assumption 1 hold, and $f : \mathbf{R} \to \mathbf{R}$ be regular. Then we have*

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{t=1}^{n} f(x_{1t}) &\to_d L_1(1,0)\int_{-\infty}^{\infty} f(s)ds, \\
\frac{1}{n}\sum_{t=1}^{n} f(x_{1t})x_{2t} &\to_d \int_0^1 V_2(r)dL_1(r,0)\int_{-\infty}^{\infty} f(s)ds, \\
\frac{1}{n^{3/2}}\sum_{t=1}^{n} f(x_{1t})x_{2t}x_{2t}' &\to_d \int_0^1 V_2(r)V_2(r)'dL_1(r,0)\int_{-\infty}^{\infty} f(s)ds,
\end{aligned}
$$

jointly as $n \to \infty$.

Lemma 3 *Let Assumption 1 hold, and assume for $k, l, = 1, \ldots, J$, $A_k A_l \eta_{kl}$, $B_k B_l \eta_{kl}$, $A_k B_l \eta_{kl} \in \mathbf{F}_R$, and $A_k \eta_{kk}$, $B_k \eta_{kk} \in \mathbf{F}_I$, for $A_k, B_k : \mathbf{R} \to \mathbf{R}$. Then we have*

$$
\begin{pmatrix} n^{-3/4}\sum_1^n \sum_{k=1}^J A_k(x_{1t})z_{kt}x_{2t} \\ n^{-1/4}\sum_1^n \sum_{k=1}^J B_k(x_{1t}, j)z_{kt} \end{pmatrix} \to_d M^{1/2}W(1),
$$

where

$$
M = \begin{pmatrix} \int_0^1 V_2(r)V_2(r)'dL_1(r,0)\int_{-\infty}^{\infty} f_{11}(s)ds & \int_0^1 V_2(r)dL_1(r,0)\int_{-\infty}^{\infty} f_{12}(s,j)ds \\ \int_0^1 dL_1(r,0)V_2(r)'\int_{-\infty}^{\infty} f_{12}(s,j)ds & L_1(1,0)\int_{-\infty}^{\infty} f_{22}(s,j)ds \end{pmatrix},
$$

with

$$
\begin{aligned}
f_{11}(s) &= \sum_{k=1}^{J}\sum_{l=1}^{J} A_k(s)A_l(s)\eta_{kl}(s), \\
f_{12}(s,j) &= \sum_{k=1}^{J}\sum_{l=1}^{J} A_k(s)B_l(s,j)\eta_{kl}(s), \\
f_{22}(s,j) &= \sum_{k=1}^{J}\sum_{l=1}^{J} B_k(s,j)B_l(s,j)\eta_{kl}(s),
\end{aligned}
$$

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

As remarked in Park and Phillips (2000), if we let $V_{2.1} = V_2 - \sigma_{21}\sigma_{11}^{-1}V_1$, where σ_{11} and σ_{12} are respectively the variance of V_1 and V_2, then we have

$$
\int_0^1 V_2(r)dL_1(r,0) = \int_0^1 V_{2.1}(r)dL_1(r,0) \quad \text{a.s.},
$$

$$\int_0^1 V_2(r)V_2(r)'dL_1(r,0) = \int_0^1 V_{2.1}(r)V_{2.1}(r)'dL_1(r,0) \quad \text{a.s.},$$

since $\int_0^1 V_1(r)dL_1(r,0) = 0$ a.s. as $\{r : V_1(r) = 0\}$ is the support of the measure $dL_1(r,0)$. The limiting distribution in Lemma 3 is mixed Gaussian and the mixing variates are dependent upon the local time L_1 of V_1 as well as V_2. We write the limit distribution in the form $\mathrm{MN}(0, M)$.

It is also pointed out in Park and Phillips (2000) 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 $\sqrt{n}$ instead of n. Thus, suppose x_{3t} is stationary, satisfies the same conditions as v_t in Assumption 1 and is independent of u_t. Then we have:

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} f_k(x_{1t})x_{3t}x_{3t}' \to_d L_1(1,0) \int_{-\infty}^{\infty} f_k(s)ds\Sigma_{33},$$

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

$$n^{-1/4} \sum_{t=1}^{n} f_k(x_{1t})x_{3t}z_{kt} \to_d MN\left(0, L_1(1,0) \int_{-\infty}^{\infty} f_k^2(s)\eta_{kk}ds\Sigma_{33}\right).$$

Define

$$D_n = \mathrm{Diag}(n^{1/4}, n^{3/4}I_{m+J-1})$$

then $D_n^{-1}S_n(\underline{\theta}_0)$ would be

$$\begin{pmatrix} n^{-1/4}\sum_1^n \sum_{k=1}^J A_k(x_{1t};\underline{\theta})z_{kt}x_{1t} \\ n^{-3/4}\sum_1^n \sum_{k=1}^J A_k(x_{1t};\underline{\theta})z_{kt}x_{2t} \\ n^{-1/4}\sum_1^n \sum_{k=1}^J B_k(x_{1t};1,\underline{\theta})z_{kt} \\ \vdots \\ n^{-1/4}\sum_1^n \sum_{k=1}^J B_k(x_{1t};j,\underline{\theta})z_{kt} \\ \vdots \\ n^{-1/4}\sum_1^n \sum_{k=1}^J B_k(x_{1t};J,\underline{\theta})z_{kt} \end{pmatrix}$$

Using Lemma 3 and the above notion, we are now able to characterize the limit forms of the score function (7), (8) and the Hessian (9), which are given in Theorem 4.

Theorem 4 *Let Assumptions 1 and 2 hold. Then*

$$D_n^{-1}S_n(\underline{\theta}_0) \to_d Q^{1/2}W(1) \quad \text{and} \quad D_n^{-1}J_n(\underline{\theta}_0)D_n^{-1} \to_d -Q$$

jointly, where $D_n = \mathrm{Diag}(n^{1/4}, n^{3/4} I_{m+J-1})$ *and* Q *is the symmetric matrix*

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

with

$$\begin{aligned}
q_{11} &= L_1(1,0) \int_{-\infty}^{\infty} s^2 f_{11}(s) ds, \\
q_{12} &= \int_0^1 dL_1(r,0) V_2(r)' \int_{-\infty}^{\infty} s f_{11}(s) ds, \\
q_{13}(j) &= L_1(1,0) \int_{-\infty}^{\infty} s f_{12}^j(s) ds, \\
q_{22} &= \int_0^1 V_2(r) V_2(r)' dL_1(r,0) \int_{-\infty}^{\infty} f_{11}(s) ds, \\
q_{23}(j) &= \int_0^1 dL_1(r,0) V_2(r)' \int_{-\infty}^{\infty} f_{12}^j(s) ds, \\
q_{33}(i,j) &= L_1(1,0) \int_{-\infty}^{\infty} f_{22}^{ij}(s) ds,
\end{aligned}$$

and where

$$\begin{aligned}
f_{11}(s) &= \sum_{k=1}^{J} \sum_{l=1}^{J} A_k(s) A_l(s) \eta_{kl}(s), \\
f_{12}^j(s) &= \sum_{k=1}^{J} \sum_{l=1}^{J} A_k(s) B_l(s,j) \eta_{kl}(s), \\
f_{22}^{ij}(s) &= \sum_{k=1}^{J} \sum_{l=1}^{J} B_k(s,i) B_l(s,j) \eta_{kl}(s),
\end{aligned}$$

and W *is defined as in Lemma 3.*

If ϵ_t has a symmetric distribution, as in the probit and logit models, f_{11} and f_{12} are even functions. We therefore have

$$\int_{-\infty}^{\infty} s f_{11}(s) ds = 0, \text{ and } \int_{-\infty}^{\infty} s f_{12}^i(s) ds = 0,$$

so that $q_{12}, q_{13}, q_{21}, q_{31} = 0$ and Q reduces to a block diagonal matrix.

The asymptotic results for $S_n(\underline{\theta}_0)$ and $J_n(\underline{\theta}_0)$ in Theorem 4 help deliver the limit distribution of $\hat{\underline{\theta}}_n$. From the expansion (11), we expect that the normed and centered estimator satisfies

$$D_n(\hat{\underline{\theta}}_n - \underline{\theta}_0) = -(D_n^{-1}J_n(\underline{\theta}_0)D_n^{-1})^{-1}D_n^{-1}S_n(\underline{\theta}_0) + o_p(1), \tag{13}$$

a result that is established in the proof of Theorem 5 below.

Theorem 5 *Let Assumptions 1 and 2 hold. Then there exists a sequence of ML estimators for which* $\hat{\underline{\theta}}_n \to_p \underline{\theta}_0$, *and*

$$D_n(\hat{\underline{\theta}}_n - \underline{\theta}_0) \to_p Q^{-1/2}W(1),$$

in the notation introduced in Theorem 4.

REMARKS: 1. Partition the matrix Q according to the different convergence rates, i.e

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix},$$

where

$$Q_{11} = q_{11} \quad Q_{11} = \begin{pmatrix} q_{22} & q_{23} \\ q_{32} & q_{33} \end{pmatrix},$$

$$Q_{12} = \begin{pmatrix} q_{13} \\ q_{23} \end{pmatrix} \quad Q_{21} = \begin{pmatrix} q_{13} & q_{23} \end{pmatrix}.$$

Let $\hat{\alpha}_n = (\hat{\alpha}_n^1, \hat{\alpha}_n^{2\prime})'$. When $Q_{12} = Q_{21} = 0$, as in the case where ϵ_t has a symmetric distribution, we have the limits

$$n^{1/4}(\hat{\alpha}_n^1 - \alpha_0^1) \to_d Q_{11}^{-1/2}W_1(1), \tag{14}$$

$$n^{3/4}\begin{pmatrix} \hat{\alpha}_n^2 \\ \hat{\mu}_n - \mu_0 \end{pmatrix} \to_d Q_{22}^{-1/2}W_2(1) \tag{15}$$

where $W = (W_1, W_2')'$ for W defined in Theorem 5. Therefore, in this case, $\hat{\alpha}_n^1$ becomes asymptotically independent of $\hat{\alpha}_n^2$ and $\hat{\mu}_n$ conditional on x_t.

2. From Theorem 5 we get

$$D_nG'(\hat{\theta}_n - \theta_0) \to_d Q^{-1/2}W(1) = \mathrm{MN}(0, Q^{-1}). \tag{16}$$

Setting $E_n = \mathrm{Diag}(n^{1/4}I_m, n^{3/4}I_J)$ and $K = \mathrm{Diag}((h_1, 0), I_J)$, we have

$$(D_nG'E_n^{-1})^{-1} \to \mathrm{Diag}((h_1, 0), I_J) = K.$$

Therefore

$$
\begin{aligned}
E_n(\hat{\theta}_n - \theta_0) \to_d KQ^{-1/2}W(1) &= \mathrm{MN}(0, (KQ^{-1/2})(KQ^{-1/2})') \\
&= \mathrm{MN}(0, KQ^{-1}K')
\end{aligned}
$$

which we formalize as follows.

Corollary 6 *Under Assumptions 1 and 2, as $n \to \infty$,*

$$
\begin{pmatrix} n^{1/4}(\hat{\beta}_n - \beta_0) \\ n^{3/4}(\hat{\mu}_n - \mu_0) \end{pmatrix} \to_d \mathrm{MN}(0, KQ^{-1}K').
$$

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.

Theorem 7 *Under Assumptions 1 and 2,*

$$
-[E_n^{-1} J_n(\hat{\theta}_n) E_n^{-1}]^{-1},\ -[E_n^{-1} \underline{J}_n(\hat{\theta}_n) E_n^{-1}]^{-1} \to_d KQ^{-1}K',
$$

as $n \to \infty$.

Furthermore, in the case of the probit or logit model and where ϵ has a symmetric distribution and Q is block diagonal, we have

$$
\begin{aligned}
n^{1/4}(\hat{\beta}_n - \beta_0) \to_d &\quad \mathrm{MN}(0, (h_1, 0)Q_{11}^{-1}(h_1, 0)'), \\
n^{3/4}(\hat{\mu}_n - \mu_0) \to_d &\quad \mathrm{MN}(0, Q_{22}^{-1}),
\end{aligned}
$$

and, in this case, $\hat{\beta}_n$ and $\hat{\mu}_n$ are asymptotically independent.

We are also interested in $\hat{P}_j(x_t;\hat{\theta}_n)$, the predicted probability of the choice $y_t = j$, and the estimated marginal effect of x_t on $\hat{P}_j(x_t;\hat{\theta}_n)$ which is denoted by $\hat{\gamma}_{j,x} = \hat{p}_j(x_t;\hat{\theta}_n)\hat{\beta}_n$. To analyze these quantities, we define a matrix $R(0) = \text{Diag}(I_m, \iota_1')$ where ι_j is a vector of length J with the jth element 1 and other elements zero. Similarly, $R(J) = \text{Diag}(I_m, \iota_J')$ and for $1 \le j \le J-1$, $R(j) = \text{Diag}(I_m, (\iota_j, \iota_{j+1})')$. It is easy to see that

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

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 8 *Let Assumptions 1 and 2 hold. Given $x_t = x$, for $j = 0, \ldots, J$, the predicted probabilities of $y_t = j$ $(j = 0, ..., J)$ have the following asymptotic distributions as $n \to \infty$*

$$\hat{P}_j(x_t;\hat{\theta}_n) \sim_d \text{MN}\left(P_j(x;\theta_0), \frac{1}{\sqrt{n}}\Gamma'(j)KQ^{-1}K'\Gamma(j)\right),$$

where

$$\Gamma(j) = p_j(x;\theta_0)R'(j)\begin{pmatrix} x \\ -1 \end{pmatrix} \quad \text{for} \quad j = 0, ..., J$$

and

$$\Gamma(j) = R'(j)\begin{pmatrix} p_j(x;\theta_0)x \\ -f(x'\beta_0 - \sqrt{n}\mu_0^j) \\ f(x'\beta_0 - \sqrt{n}\mu_0^{j+1}) \end{pmatrix}$$

for $1 \le j \le J-1$.

Corollary 9 *Let Assumptions 1 and 2 hold. Given $x_t = x$, for $j = 0, \ldots, J$, the estimated marginal effects $\hat{\gamma}_{j,x} = \hat{p}_j(x;\hat{\theta}_n)\hat{\beta}_n$ have the following asymptotic distributions as $n \to \infty$*

$$\hat{\gamma}_{j,x} \sim_d \text{MN}\left(\gamma_j(x;\theta_0), \frac{1}{\sqrt{n}}\Psi'(j)KQ^{-1}K'\Psi(j)\right),$$

where $\gamma_j(x;\theta_0) = p_j(x;\theta_0)\beta_0$ and

$$\Psi(0) = R'(0)\begin{pmatrix} -\dot{p}_0(x;\theta_0)\beta_0 x' + p_0(x;\theta_0)I_m \\ \dot{f}(x'\beta_0 - \sqrt{n}\mu_0^1)\beta_0' \end{pmatrix},$$

$$\Psi(J) = R'(J)\begin{pmatrix} -\dot{p}_J(x;\theta_0)\beta_0 x' + p_J(x;\theta_0)I_m \\ \dot{f}(x'\beta_0 - \sqrt{n}\mu_0^J)\beta_0' \end{pmatrix},$$

and

$$\Psi(j) = R'(j)\begin{pmatrix} -\dot{p}_j(x;\theta_0)\beta_0 x' + p_j(x;\theta_0)I_m \\ -\dot{f}(x'\beta_0 - \sqrt{n}\mu_0^j)\beta_0' \\ \dot{f}(x'\beta_0 - \sqrt{n}\mu_0^{j+1})\beta_0' \end{pmatrix}$$

for $1 \leq j \leq J-1$ *and where* $\dot{f}(\cdot)$ *denotes the first derivative of* $f(\cdot)$*.*

Finally, it is of interest to study the asymptotic behavior of the empirical average of $r_n(j) = \frac{1}{n}\sum_{t=1}^{n} 1\{y_t = j\}$. The quantity r_n is an aggregate proportion and measures the proportion of $y_t = j$ outcomes in the sample data. It can also be used in a predictive manner to forecast the proportion of $y_t = j$ choices given a sequence of data on the covariates, say, $X = \{X_t : t = 1, \ldots, n\}$. In this case, we can define

$$\begin{aligned} y_{0,t}(X) &= 1\{X_t'\beta_0 \leq \sqrt{n}\mu_0^1 + \epsilon_t\} \\ y_{j,t}(X) &= 1\{\sqrt{n}\mu_0^j + \epsilon_t < X_t'\beta_0 \leq \sqrt{n}\mu_0^{j+1} + \epsilon_t\} \quad \text{for} \quad j = 1, \ldots, J-1 \\ y_{J,t}(X) &= 1\{X_t'\beta_0 > \sqrt{n}\mu_0^J + \epsilon_t\}. \end{aligned}$$

Since $y_{j,t}$ is unobserved, we could use the estimated quantities $\hat{r}_n(j, X) = n^{-1}\sum_{t=1}^{n} \hat{P}_j(X_t; \hat{\theta}_n)$ instead. The following result gives the limit theory for these empirical averages.

Theorem 10 *Let Assumptions 1 and 2 hold and define* $\omega_x^2 = \beta_0'\Sigma\beta_0$*. Suppose the time series* $X = \{X_t : t = 1, \ldots, n\}$ *is drawn independently of* x_t *from a process with properties equivalent to those of* x_t *as given in Assumption 1. Then the sample proportion* $r_n(j) = \frac{1}{n}\sum_{t=1}^{n} 1\{y_t = j\}$*, the predicted proportion* $r_n(j, X) = \frac{1}{n}\sum_{t=1}^{n} 1\{y_t(X) = j\}$*, and the estimated proportion* $\hat{r}_n(j, X) = n^{-1}\sum_{t=1}^{n} \hat{P}_j(X_t; \hat{\theta}_n)$ *all have the following limit behavior as* $n \to \infty$*:*

$$\begin{aligned} r_n(0), r_n(0, X), \hat{r}_n(0, X) &\to_d \int_0^1 1\left\{W(r) < \frac{\mu_0^1}{\omega_x}\right\} dr, \\ r_n(J), r_n(J, X), \hat{r}_n(J, X) &\to_d \int_0^1 1\left\{W(r) > \frac{\mu_0^J}{\omega_x}\right\} dr, \\ r_n(j), r_n(j, X), \hat{r}_n(j, X) &\to_d \int_0^1 1\left\{\frac{\mu_0^j}{\omega_x} < W(r) < \frac{\mu_0^{j+1}}{\omega_x}\right\} dr \\ \text{for} \quad j &= 1, \ldots, J-1. \end{aligned}$$

Borodin and Salminen (1996) give explicit forms for the probability distributions of the above limit quantities, which represent the time spent by a Brownian motion above or below certain boundaries and in certain bounded intervals. Assume that $W(0) = 0$, $\mu_0^1 < 0$, $\mu_0^J > 0$ and $0 \in (\frac{\mu_0^{\bar{j}}}{\omega_x}, \frac{\mu_0^{\bar{j}+1}}{\omega_x})$, so that $\mu_0^j > 0$ for $j > \bar{j}$ and $\mu_0^j < 0$ for $j \leq \bar{j}$. Also, for simplicity, let $a_x^j = \frac{\mu_0^j}{\omega_x}$. Then, we have the following expressions for the probability densities of these limits:

$$\text{Density of } \int_0^1 1\{W(r) > a_x^J\}\, dr : p(y) = \frac{1}{\pi\sqrt{y(1-y)}} e^{-|a_x^J|^2/2(1-y)} \tag{17}$$

$$\text{Density of } \int_0^1 1\{W(r) < a_x^1\}\, dr : p(y) = \frac{1}{\pi\sqrt{y(1-y)}} e^{-|a_x^1|^2/2(1-y)} \tag{18}$$

$$\begin{aligned} & \text{Density of } \int_0^1 1\{a_x^j < W(r) < a_x^{j+1}\}\, dr, \quad \text{for} \quad j > \bar{j} \qquad (19) \\ = \ & 2\int_0^\infty h_{1-y}(0, v + a_x^j) \text{ce}_y \left(0, 0, \frac{a_x^{j+1} - a_x^j}{2}, 0, v\right) dv + \int_0^\infty h_{1-y}(0, v + a_x^j) \\ & \times \left(\text{ec}_y \left(-1, 1, \frac{a_x^{j+1} - a_x^j}{2}, -\frac{a_x^{j+1} - a_x^j}{2}, v,\right) - \text{ec}_y \left(-1, 1, \frac{a_x^{j+1} - a_x^j}{2}, \frac{a_x^{j+1} - a_x^j}{2}, v,\right)\right) dv, \end{aligned}$$

$$\begin{aligned} & \text{Density of } \int_0^1 1\{a_x^{\bar{j}} < W(r) < a_x^{\bar{j}+1}\}\, dr, \qquad (20) \\ = \ & 2\int_0^\infty h_{1-y}(0, v) \text{ec}_y^{(1)} \left(\frac{a_x^{\bar{j}+1} + a_x^{\bar{j}}}{2}, \frac{a_x^{\bar{j}+1} - a_x^{\bar{j}}}{2}, 0, v\right) dv \\ & + \frac{1}{2}\int_0^\infty h_{1-y}(1, v) \left(\text{ec}_y \left(-1, 2, \frac{a_x^{\bar{j}+1} - a_x^{\bar{j}}}{2}, -a_x^{\bar{j}+1}, v\right) - \text{ec}_y \left(-1, 2, \frac{a_x^{\bar{j}+1} - a_x^{\bar{j}}}{2}, a_x^{\bar{j}+1}, v\right) \right. \\ & \left. + \text{ec}_y \left(-1, 2, \frac{a_x^{\bar{j}+1} - a_x^{\bar{j}}}{2}, a_x^{\bar{j}}, v\right) - \text{ec}_y \left(-1, 2, \frac{a_x^{\bar{j}+1} - a_x^{\bar{j}}}{2}, -a_x^{\bar{j}}, v\right)\right) dv, \end{aligned}$$

$$\begin{aligned} & \text{Density of } \int_0^1 1\{a_x^j < W(r) < a_x^{j+1}\}\, dr, \quad \text{for} \quad j < \bar{j} \qquad (21) \\ = \ & 2\int_0^\infty h_{1-y}(0, v - a_x^{j+1}) \text{ec}_y \left(0, 0, \frac{a_x^{j+1} - a_x^j}{2}, 0, v\right) dv + \int_0^\infty h_{1-y}(0, v - a_x^{j+1}) \\ & \times \left(\text{ec}_y \left(-1, 1, \frac{a_x^{j+1} - a_x^j}{2}, -\frac{a_x^{j+1} - a_x^j}{2}, v,\right) - \text{ec}_y \left(-1, 1, \frac{a_x^{j+1} - a_x^j}{2}, \frac{a_x^{j+1} - a_x^j}{2}, v,\right)\right) dv. \end{aligned}$$

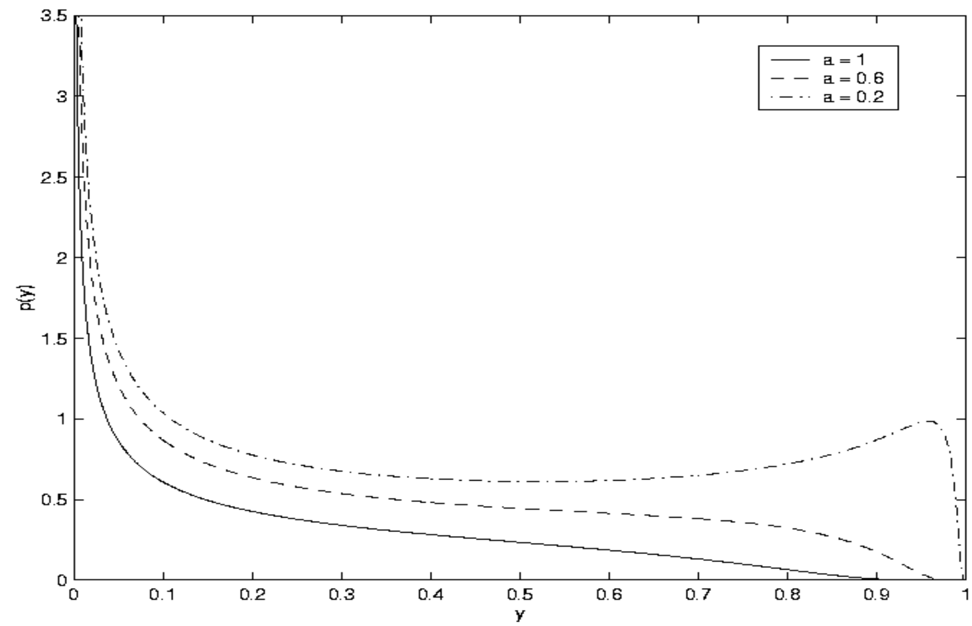

Figure 1: The density for $\int_0^1 [W(r) > a] \ dr$

Definitions of the special functions h_{1-y} and ec_y that are included in the last three formulae above are given in Appendix D. They involve recursions and are needed for the more complex case of the density of the time spent by Brownian motion in a bounded interval. These formulae are used subsequently in our computations.

Park and Phillips (2000) show that in the nonstationary binary choice case the sample proportion converges to a random variable that follows the arc sine law with probability density $1/(\pi\sqrt{y(1-y)})$ on $[0, 1]$. This case applies when $a_x^J = 0$ in (17) or when $a_x^1 = 0$ in (18)[1]. In the general multiple choice case, the limit results are much more complex and offer a range of interesting possible outcomes that extend the arc sine limit law outcome. Correspondingly, we refer to them as 'extended arc sine' laws.

As the formulae for the limit densities are quite complicated, we draw the following figures to illustrate the densities when $W(0) = 0$ and for several different parameter configurations. These reveal how the shape of the density changes as the boundary limits change and give some idea of the

[1]For example, these outcomes apply in the present case when the threshold parameters in (2) are fixed rather than of $O(\sqrt{n})$.

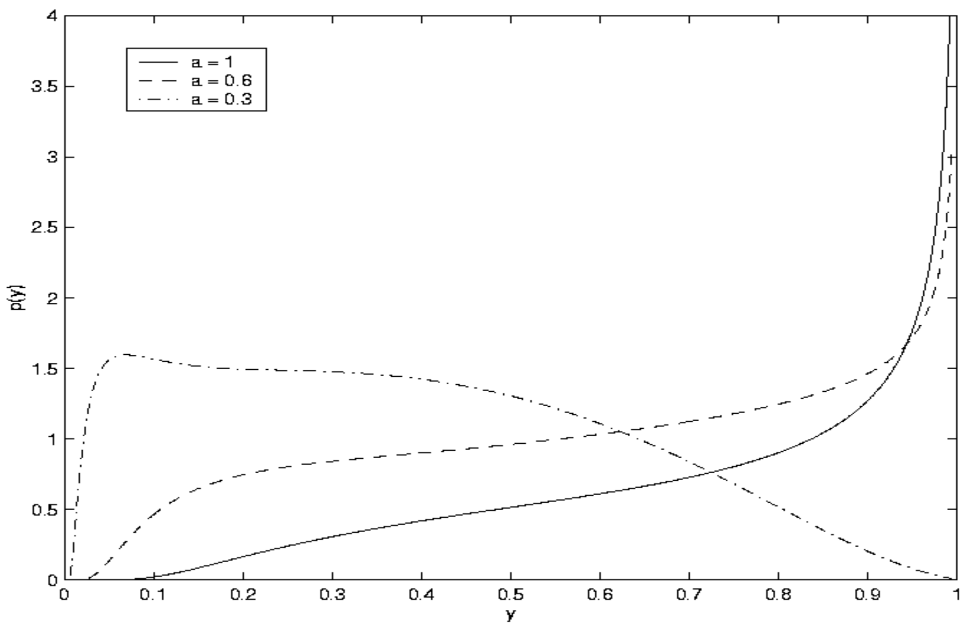

Figure 2: The density for $\int_0^1 [-a < W(r) < a]\ dr$

range of possibilities beyond the special case of the U-shaped arc sine density. Figure 1 shows the density (17). Observe that as the parameter a approaches zero, the density begins to take on the form of the U-shaped arc sine density although the density for $a = 0.2$ rapidly approaches zero at unity unlike the arc sine law. Figure 2 shows (20), giving the density of the time spent in the (symmetric) interval $[-a, a]$ about the origin. When a is small, the distribution of time spent is fairly evenly spread for $y \leq 0.5$, but the density tails off for $y \in (0.5, 1]$. When a takes larger values (here $a = 0.6, 1.0$) the density is increasing with y. Figure 3 depicts the density (19) for intervals $[a_1, a_2]$ away from the origin, where, as we might expect, the density decreases from the origin. Note that the pair (17) and (18), and the pair (19) and (21) have the same form, so we only depict one in each pair. Also note that the difference between (19), (20) and (21) depends on the relative position of the initial position of $W(0)$ to the spatial interval we are interested in. If we set $W(0) = 0$, then (19) applies when the interval is above 0, (20) applies when the interval covers 0, and finally, (21) applies when the interval is below 0.

Apparently, a wide range of possible behavior can be captured with this class of densities, depending on the precise values of the parameters determining the boundary values. In the binary choice case of Park and Phillips (2000), the arc sine law gave the limit density of the average sample proportion of $0, 1$ choices, corresponding to the limit Brownian motion process (arising from the limit of the normalized index $\beta_0' x_t / \sqrt{n}$) being on one side of the origin or the other. This is a very special case. When there are multiple choices with thresholds determining those choices, then the limit density of sample proportions of the choices depends on the thresholds and the variance of the Brownian motion. The probability distribution of the time spent by the limit process in any particular interval (and, correspondingly, the limit distribution of the sample proportions of a certain choice) can then take on a wide range of shapes. This means that in an empirical application (such as market intervention) of polychotomous choice with nonstationary covariates, we need not necessarily expect behavior such as persistent runs of the same choice.

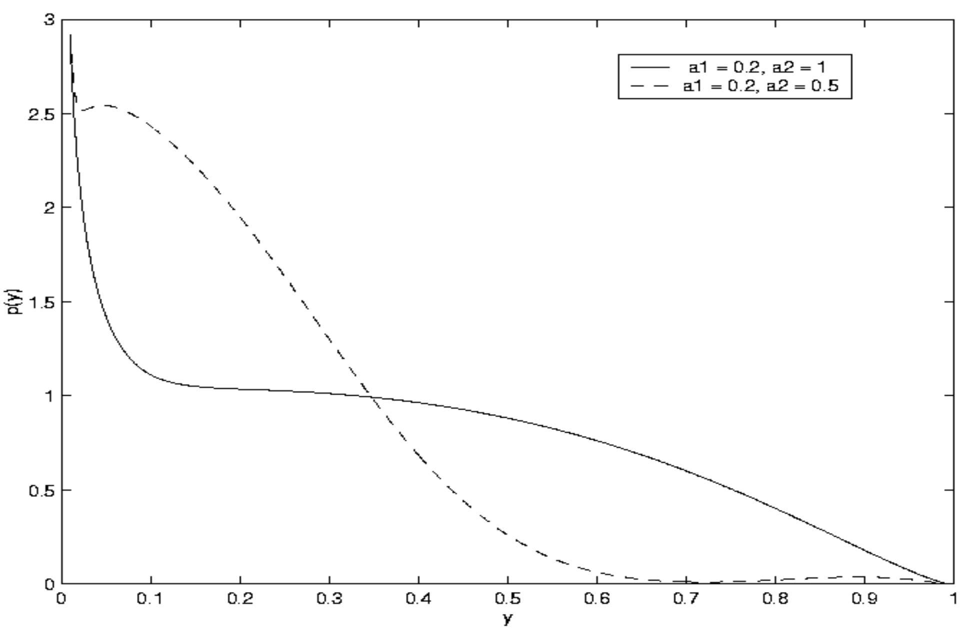

Figure 3: The density for $\int_0^1 [a_1 < W(r) < a_2] \; dr$

4 Covariates With a Deterministic Time Trend

In Assumption 1, we assume that x_t is an integrated process without drift or deterministic trend. However, many economic and financial time series, such as GDP and stock prices, show evidence of drift or trend over time. This section extends the earlier discussion by including such behavior. We re-specify the model for x_t as

$$x_t = \delta_0 t + \underline{x}_t \tag{22}$$

where $\underline{x}_t = \sum_{i=1}^{t} v_i$, or equivalently,

$$x_t = x_{t-1} + \delta_0 + v_t$$

where v_t satisfies the conditions of Assumption 1. The behavior of x_t is dominated by $\delta_0 t$. Correspondingly, in the model (1) and (2), the observed dependent variable y_t takes a constant value (viz., 0 or J) with probability approaching unity as $t \to \infty$. In particular, the conditional probability

$$\begin{aligned} P\left(y_t = J|\mathcal{F}_t\right) &= P\left(\varepsilon_t < \sqrt{n}\mu_0^J - t\beta_0'\delta_0 - \underline{x}_t|\mathcal{F}_t\right) \\ &= F\left(\sqrt{n}\mu_0^J - t\beta_0'\delta_0 - \underline{x}_t\right) \to \begin{cases} 0 & \beta_0'\delta_0 > 0 \\ 1 & \beta_0'\delta_0 < 0 \end{cases} \end{aligned}$$

with a similar expression for $P\left(y_t = 0|\mathcal{F}_t\right)$. Consistent estimation of the parameters is impossible in such cases because there is insufficient data on the various outcomes. In practice, of course, a model of the type (1) and (2) will be of interest when the choice outcomes are observed with some regularity in the sample. For this to be so in the present case, the data need to be detrended. Equivalently, the thresholds need to be adjusted to incorporare the trend. For practical applications, this is analogous to the decision maker detrending the data. For example, if (1) and (2) are used to model money market intervention by a central bank, where decisions are affected by the time path of variables with trends, the authority (either explicitly or implicitly) must be considering, not the levels of the covariates x_t, but the fluctuations of x_t about its trend path (or estimated trend paths) in making its decisions.

Suppose, therefore, we first estimate δ_0 in (22) and detrend x_t using this estimate. The (difference) least squares estimator of δ_0, which is asymptotically efficient (i.e., equivalent to generalized least squares) is

$$\hat{\delta}_n = n^{-1}\sum_{t=1}^{n}(x_t - x_{t-1}) = \delta_0 + n^{-1}\sum_{t=1}^{n} v_t,$$

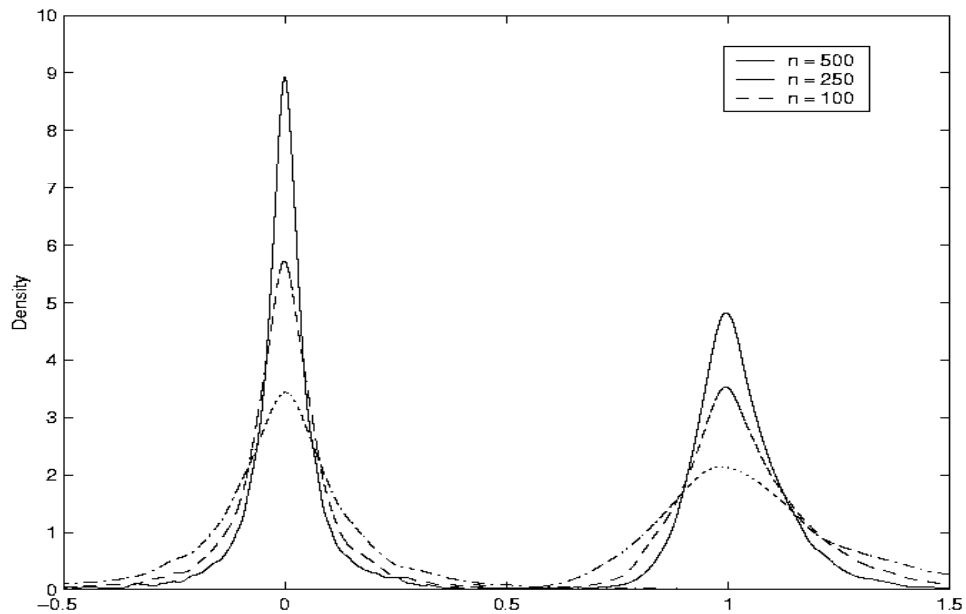

Figure 4: Probit Model: densities of estimators of $\beta_1^0 = 1, \beta_2^0 = 0$; $I(1)$ case

and as $n \to \infty$, we have

$$\sqrt{n}(\hat{\delta}_n - \delta_0) = n^{-1/2} \sum_{t=1}^{n} v_t \to V(1),$$

where $V(\cdot)$ is the Brownian motion defined in Lemma 1.

The detrended series, $\tilde{x}_t$, is

$$\tilde{x}_t = x_t - \hat{\delta} t = \underline{x}_t + (\delta_0 - \hat{\delta}_n) t,$$

hence

$$\begin{aligned} n^{-1/2} \tilde{x}_t &= n^{-1/2} \underline{x}_t - \sqrt{n}(\hat{\delta}_n - \delta_0) \frac{t}{n} \\ &\to V(r) - rV(1) = \tilde{V}(r) \end{aligned}$$

as $n \to \infty$. Therefore, in the case of unit root process with deterministic trend, we can work with the detrended series $\tilde{x}_t$ and approximate $n^{-1/2} \tilde{x}_t$ with the Brownian bridge process $\tilde{V}(r)$.

5 Simulation Evidence of the Effects of Nonstationarity

This section provides some simulation evidence on the finite sample performance of ML estimation of a polychotomous choice model under nonstationarity. We consider a model such as (1) and (2) with $m = 2$ explanatory variables and $J = 2$, giving a triple-choice dependent variable y_t. The generating mechanism for the exogenous data is the system

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix},$$

with $v_t = (v_{1t}, v_{2t})' = iid\ N(0, I_2)$. Both unit root ($a_{ii} = 1,\ i = 1, 2$) and stationary ($a_{ii} = 0.5,\ i = 1, 2$) cases were considered. The coefficient parameter vector was set at $\beta_0 =(1, 0)'$ and $\mu_0 = (\mu_0^1, \mu_0^2) = (-0.5, 0.5)'$. Thus $x_t'\beta_0 = \beta_1^0 x_{1t} = x_{1t}$ and the direction orthogonal to β_0 is $(0, 1)$, giving the coefficient $\beta_2^0 = 0$ of x_{2t}, so that this set up is analogous to that of the simulation study in Park and Phillips (2000). The number of replications was $5,000$.

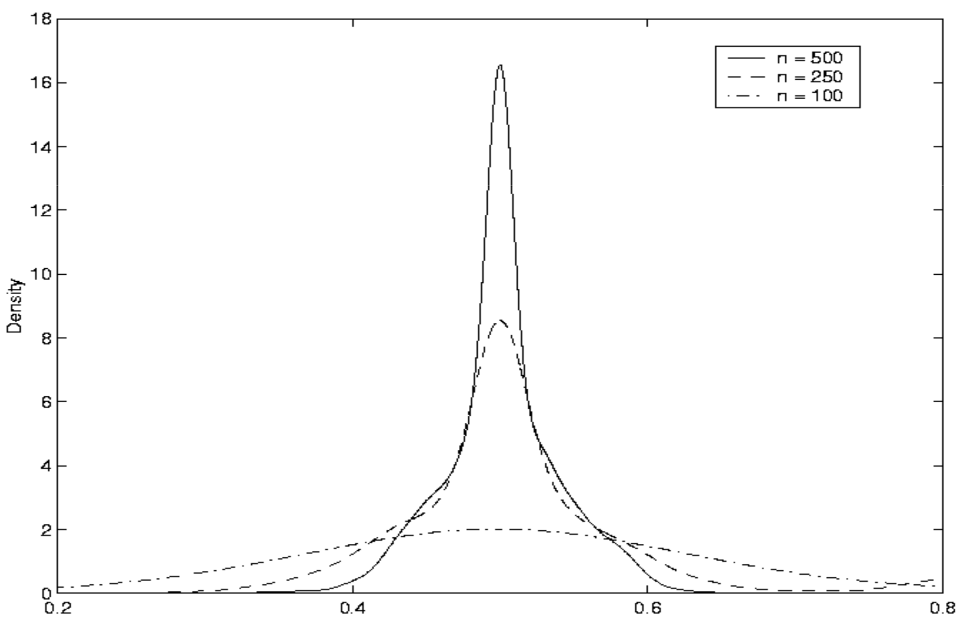

Figure 5: Probit Model: densities of estimators of $\mu_0^2 = 0.5$; $I(1)$ case

Figure 4 and Figure 5 depict kernel estimates of the sampling distribution of the probit estimates of the coefficients β_0 and μ_0^2 in the $I(1)$ case. The different convergence rate for $\widehat{\beta}_1$ and $\widehat{\beta}_2$ is apparent in Figure 4. Figure 6 gives kernel estimates of the probit estimates of the coefficient μ_0^2 in the $I(0)$ case for comparison purposes. From Figure 5 and Figure 6, it is evident that the estimator of μ_0 is more concentrated and converges faster in the nonstationary case than in the stationary case, corroborating the asymptotic theory.

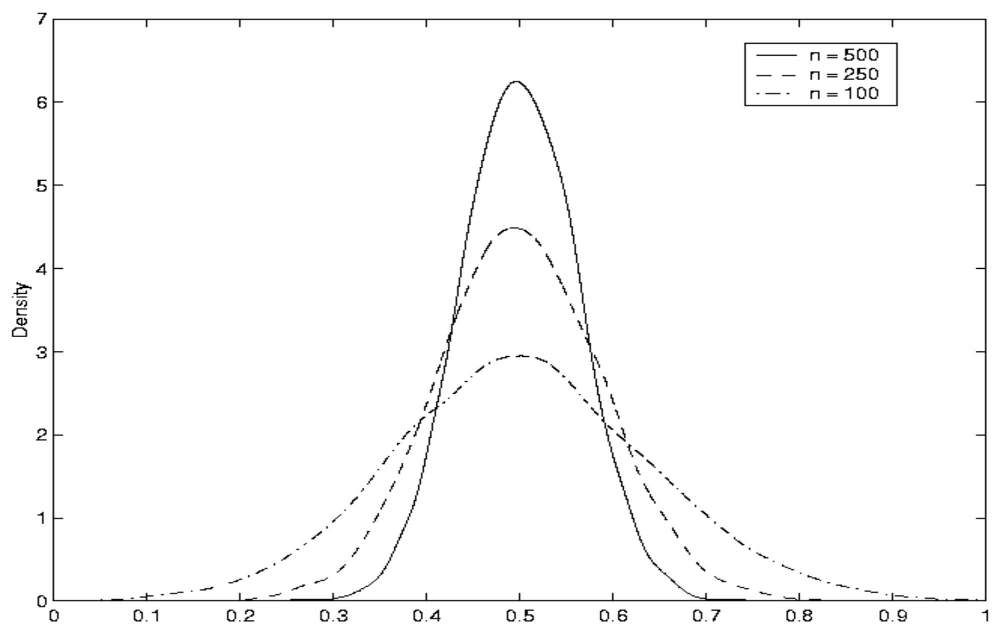

Figure 6: Probit Model: densities of estimators of $\mu_0^2 = 0.5$; $I(0)$ case

6 Conclusion

Discrete dependent variable modeling has proved to be a powerful tool in microeconometric analysis. Even though there is little empirical work to date, there appear to be plenty of potential applications of the approach to economic time series, including some in time series macroeconomics with nonstationary data. The present paper develops an asymptotic theory for maximum likelihood estimation of these models that allows for integrated explanatory variables, extending the work of Park and Phillips (2000) on binary choice to the case of polychotomous choice where there are threshold parameters to be estimated. We find different convergence rates ($n^{1/4}$ and $n^{3/4}$) for the coefficient estimates, just as in the Park-Phillips study, and a convergence rate of $n^{3/4}$ for the threshold parameters. In general, the two sets of estimates are asymptotically dependent and follow a mixed normal limit distribution which means that conventional methods of inference are possible.

A new finding in the present paper is that the sample proportion of choices of each type has a limit distribution that belongs to a family of extended arc sine laws. These laws have a wide range of possible distributional forms and thereby allow for some flexibility in applications. One application that the authors are studying in related research (Hu and Phillips, 2001) involves the practice of monetary policy and is concerned with modeling the discrete decision making structure of federal fund targeting by the Federal Reserve.

APPENDIX A: USEFUL LEMMAS AND PROOFS

Lemma 11 *Let Assumption 1 hold, and $f : \mathbf{R} \to \mathbf{R}$. Denote by x_{2t}^{κ} the κ-times tensor product of x_{2t} with itself. Define:*

$$ {}_1M_n^{\kappa} = \sum_{t=1}^{n} f(x_{1t}) x_{2t}^{\kappa}, \quad {}_2M_n^{\kappa} = \sum_{t=1}^{n} f(x_{1t}) x_{2t}^{\kappa} z_{kt}, $$

$$ {}_3M_n^{\kappa} = \sum_{t=1}^{n} f(x_{1t}) x_{2t}^{\kappa} (z_{kt}^2 - \eta_{kk,t}). $$

(a) *For $f \in \mathbf{F}_0$, ${}_1M_n^{\kappa} = o_p(n^{1+\kappa/2})$. Moreover, if $f \in \mathbf{F}_I$, then ${}_1M_n^{\kappa} = O_p(n^{(1+\kappa)/2})$.*

(b) *If $\eta_{kk} f^2 \in \mathbf{F}_0$, then ${}_2M_n^{\kappa} = o_p(n^{(1+\kappa)/2})$.*

(c) *If $\tau_{kl} f^2 \in \mathbf{F}_0$, then ${}_2M_n^{\kappa} = o_p(n^{(1+\kappa)/2})$*

Proof of Lemma 11:

Let $V = (V_1, V_2')'$. Note that

$$ \sup_{1 \le t \le n} \left\| \frac{x_{2t}}{\sqrt{n}} \right\|^{\kappa} =_d \sup_{1 \le t \le n} \|V_{2n}(r)\|^{\kappa} \le \sup_{0 \le r \le 1} \|V_2(r)\|^{\kappa} + 1 < \infty \quad \text{a.s.} $$

for all large n. For $f \in \mathbf{F}_0$, we have $n^{-1} \sum_{t=1}^{n} |f(x_{1t})| \to_d 0$, as shown in Park and Phillips (1999). If $f \in \mathbf{F}_I$, it follows from Lemma2 that

$n^{-1/2}\sum_{t=1}^{n}|f(x_{1t})| \to_d O_p(1)$ since f is bounded by a regular function. The stated results in part (a) follow. For part (b), note that

$$\begin{aligned}
n^{-(1+\kappa)}E\|{}_2M_n^\kappa\|^2 &= E\left(\frac{1}{n^{1+\kappa}}\sum_{t=1}^{n} f(x_{1t})^2\eta_{kk}(x_{1t})\|x_{2t}\|^{2\kappa}\right) \\
&= E\left(\frac{1}{n}\sum_{t=1}^{n}(\eta_{kk}f^2)(x_{1t})\left\|\frac{x_{2t}}{\sqrt{n}}\right\|^{2\kappa}\right) \\
&\leq E\left(\left(\sup_{0\le r\le 1}\|V_2(r)\|^{2\kappa}+1\right)\int_0^1(\eta_{kk}f^2)(\sqrt{n}V_{1n}(r))dr\right), \\
\to_p &\quad 0
\end{aligned}$$

by part (a) and dominated convergence. Similarly, for part (c),

$$\begin{aligned}
n^{-(1+\kappa)}E\|{}_3M_n^\kappa\|^2 &= E\left(\frac{1}{n^{1+\kappa}}\sum_{t=1}^{n} f^2(x_{1t})\tau_{kk}(x_{1t})\|x_{2t}\|^{2\kappa}\right) \\
&= E\left(\frac{1}{n}\sum_{t=1}^{n}(\tau_{kk}f^2)(x_{1t})\left\|\frac{x_{2t}}{\sqrt{n}}\right\|^{2\kappa}\right) \\
&\leq E\left(\left(\sup_{0\le r\le 1}\|V_2(r)\|^{2\kappa}+1\right)\int_0^1(\tau_{kk}f^2)(x_{1t})(\sqrt{n}V_{1n}(r))dr\right) \\
\to_p &\quad 0.
\end{aligned}$$

Q.E.D.

Lemma 12 *Let Assumption 1 hold. Assume* $\eta_{kk}f_k, \eta_{kk}g_k \in \mathbf{F}_I$ *and* $\tau_{kk}f_k^2, \tau_{kk}g_k^2 \in \mathbf{F}_0$ *for* $f_k, g_k : \mathbf{R} \to \mathbf{R}$. *Define*

$${}_1N_{nt}^2 = n^{-3/4}f_k(x_{1t})z_{kt}^2 \quad {}_2N_{nt}^2 = n^{-5/4}g_k(x_{1t})x_{2t}z_{kt}^2.$$

Then, for $i = 1, 2$ *we have, as* $n \to \infty$,

$$\sup_{1\le t\le n}\left\|\sum_{s=1}^{t} {}_iN_{ns}^2\right\| \to_p 0.$$

Proof of Lemma 12:

By part (a) in Lemma 11,

$$f_k(x_{1t})\eta_{kk,t} = O_p(n^{1/2}) \quad \text{and} \quad g_k(x_{1t})x_{2t}\eta_{kk,t} = O_p(n).$$

Next, by part (c) in Lemma 11,

$$\begin{aligned}
{}_1N_{nt}^2 &= n^{-3/4}f_k(x_{1t})z_{kt}^2 = n^{-3/4}f_k(x_{1t})\eta_{kk,t} + o_p(n^{-1/4}) \to_p 0, \\
{}_2N_{nt}^2 &= n^{-5/4}g_k(x_{1t})x_{2t}z_{kt}^2 = n^{-5/4}g_k(x_{1t})x_{2t}\eta_{kk,t} + o_p(n^{-1/4}) \to_p 0.
\end{aligned}$$

Q.E.D.

APPENDIX B: PROOF OF THE MAIN THEOREMS

Proof of Lemmas 1 and 2

See Park and Phillips (2000).

Proof of Lemma 3:

We set $m = 2$ for notational simplicity (so both x_{1t} and x_{2t} are scalars). Also, since the results hold for any $j = 1, \ldots, J$ in $B_k(x_1, j)$, we omit j for simplicity. For any $c = (c_1, c_2) \in \mathbf{R}^2$. We let

$$C_{kn}(x_1, x_2) = c_1 n^{-1/4} B_k(x_1) + c_2 n^{-3/4} A_k(x_1) x_2,$$

and define

$$\begin{aligned} M_{kn}(r_k) &= \sqrt{n} \sum_{i=1}^{t-1} C_{kn}(\sqrt{n} V_{ni}) \left(U_k \left(\frac{T_{k,ni}}{n} \right) - U_k \left(\frac{T_{kn,i-1}}{n} \right) \right) \\ &\quad + \sqrt{n} C_{kn}(\sqrt{n} V_{nt}) \left(U_k(r_k) - U_k \left(\frac{T_{kn,t-1}}{n} \right) \right). \end{aligned} \tag{23}$$

Thus, for $k = 1, \ldots, J$, M_{kn} is a continuous martingale such that

$$\sum_{t=1}^{n} C_{kn}(x_{1t}, x_{2t}) z_{kt} =_d M_{kn} \left(\frac{T_{k,nn}}{n} \right) \tag{24}$$

Therefore, $M_n = \sum_{k=1}^{J} M_{kn}$ is also a continuous martingale such that

$$\sum_{k=1}^{J} \sum_{t=1}^{n} C_{kn}(x_{1t}, x_{2t}) z_{kt} =_d \sum_{k=1}^{J} M_{kn} \left(\frac{T_{k,nn}}{n} \right). \tag{25}$$

Let $D_{kl,n}(x_1, x_2) = \eta_{kl}(x_1) C_{kn}(x_1, x_2) C_{ln}(x_1, x_2)$. Then the quadratic covariation process $[M_{kn}, M_{ln}]$ of M_{kn} and M_{ln} is given by

$$\begin{aligned} & [M_{kn}, M_{ln}](r) \\ = \; & n \sum_{i=1}^{t-1} C_{kn}(\sqrt{n} V_{ni}) C_{ln}(\sqrt{n} V_{ni}) \left(U_k \left(\frac{T_{k,ni}}{n} \right) - U_k \left(\frac{T_{ln,i-1}}{n} \right) \right) \\ & \cdot \left(U_l \left(\frac{T_{l,ni}}{n} \right) - U_l \left(\frac{T_{ln,i-1}}{n} \right) \right) \\ & + n C_{kn}(\sqrt{n} V_{nt}) C_{ln}(\sqrt{n} V_{nt}) \left(r_k - \frac{T_{kn,t-1}}{n} \right) \left(r_l - \frac{T_{ln,t-1}}{n} \right) \\ = \; & \sum_{t=1}^{n} D_{kl,n}(\sqrt{n} V_{nt}) 1 \left\{ r \geq \min \left\{ \frac{T_{k,nt}}{n}, \frac{T_{l,nt}}{n} \right\} \right\} + o_p(1), \end{aligned}$$

uniformly in $r \in [0,1]$. Consequently,

$$[M_n](r) = \sum_{k=1}^{J}\sum_{l=1}^{J}[M_{kn}, M_{ln}](r).$$

Therefore, we have

$$[M_n](r) \to_p c'M(r)c, \tag{26}$$

uniformly in $r \in [0,1]$, where

$$M = \begin{pmatrix} \int_0^1 V_2(r)V_2(r)'dL_1(r,0)\int_{-\infty}^{\infty} f_{11}(s)ds & \int_0^1 V_2(r)dL_1(r,0)\int_{-\infty}^{\infty} f_{12}(s,j)ds \\ \int_0^1 dL_1(r,0)V_2(r)'\int_{-\infty}^{\infty} f_{12}(s,j)ds & L_1(1,0)\int_{-\infty}^{\infty} f_{22}(s,j)ds \end{pmatrix},$$

due to the results in Lemma 2, where

$$\begin{aligned} f_{11}(s) &= \sum_{k=1}^{J}\sum_{l=1}^{J} A_k(s)B_l(s)\eta_{kl}(s), \\ f_{12}(s) &= \sum_{k=1}^{J}\sum_{l=1}^{J} A_k(s)B_l(s)\eta_{kl}(s), \\ f_{22}(s) &= \sum_{k=1}^{J}\sum_{l=1}^{J} B_k(s)B_l(s)\eta_{kl}(s). \end{aligned}$$

Moreover, if we let $\sigma_{uv}(k)$ be the covariance of U_k and V and

$$E_{kn}(x_1, x_2) = \eta_{kk}(x_1)C_{kn}(x_1, x_2),$$

then the quadratic covariation process $[M_{kn}, V]$ of M_{kn} and V is:

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

uniformly in $r \in [0,1]$, by Lemma 12. It follows, in particular, that for $k = 1, \ldots, J-1$,

$$[M_{kn}, V](\rho_{kn}(r)) \to_p 0, \tag{27}$$

where $\rho_{kn}(r) = \inf\{s \in [0,1] : [M_{kn}](s) > r\}$ is a sequence of time changes.

The asymptotic distribution of the continuous martingale M_n in (23) is completely determined by (26) and (27), as shown in Revuz and Yor (1994, Theorem 2.3). Now define

$$W_{kn}(r) = M_{kn}(\rho_{kn}(r)).$$

The process W_{kn} is the DDS (or Dambis, Dubins-Schwarz) Brownian motion (see Revuz and Yor, 1994) of the continuous martingale M_{kn}. It follows that (V, M_{kn}) converges jointly in distribution to two independent standard linear Brownian motions (V, W_k), say. Therefore,

$$M_n = \sum_{k=1}^{J} M_{kn}\left(\frac{T_{kn,n}}{n}\right) \to_d W(c'Mc),$$

which, in view of (25), completes the proof.

Q.E.D.

Proof of Theorem 4

The results for the score function directly follow lemma 3. For the Hessian matrix $J_n(\underline{\theta}_n) = G'J_n(\theta_n)G$, we partition the matrix as

$$\begin{pmatrix} J_{n,11}(\underline{\theta}_0) & J_{n,12}(\underline{\theta}_0) & J_{n,13}(\underline{\theta}_0) \\ J_{n,21}(\underline{\theta}_0) & J_{n,22}(\underline{\theta}_0) & J_{n,23}(\underline{\theta}_0) \\ J_{n,31}(\underline{\theta}_0) & J_{n,32}(\underline{\theta}_0) & J_{n,33}(\underline{\theta}_0) \end{pmatrix} \tag{28}$$

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

$$\begin{aligned}
J_{n,11}(\underline{\theta}_0) &= -\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} A_k(x_{1t})A_l(x_{1t})z_{kt}z_{lt}x_{1t}^2 + \sum_{t=1}^{n}\sum_{k=1}^{J} C_{\beta\beta,k}(x_{1t})z_{kt}x_{1t}^2, \\
J_{n,12}(\underline{\theta}_0) &= -\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} A_k(x_{1t})A_l(x_{1t})z_{kt}z_{lt}x_{1t}x_{2t}' + \sum_{t=1}^{n}\sum_{k=1}^{J} C_{\beta\beta,k}(x_{1t})z_{kt}x_{1t}x_{2t}', \\
J_{n,22}(\underline{\theta}_0) &= -\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} A_k(x_{1t})A_l(x_{1t})z_{kt}z_{lt}x_{2t}x_{2t}' + \sum_{t=1}^{n}\sum_{k=1}^{J} C_{\beta\beta,k}(x_{1t})z_{kt}x_{2t}x_{2t}', \\
J_{n,13}(\underline{\theta}_0)(i) &= -\sqrt{n}\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} A_k(x_{1t})B_l(x_{1t},i)z_{kt}z_{lt}x_{1t} + \sqrt{n}\sum_{t=1}^{n}\sum_{k=1}^{J} C_{\beta\mu^i,k}(x_{1t})z_{kt}x_{1t}, \\
J_{n,23}(\underline{\theta}_0)(i) &= -\sqrt{n}\sum_{t=1}^{n}\sum_{k=1}^{J}\sum_{l=1}^{J} A_k(x_{1t})B_l(x_{1t},i)z_{kt}z_{lt}x_{2t}' + \sqrt{n}\sum_{t=1}^{n}\sum_{k=1}^{J} C_{\beta\mu^i,k}(x_{1t})z_{kt}x_{2t}', \\
J_{n,33}(\underline{\theta}_0) &= J_{n,22}(\theta_0).
\end{aligned}$$

Then, $D_n^{-1} J_n(\underline{\theta}_0) D_n^{-1}$ has the form

$$\begin{pmatrix} n^{-1/2} J_{n,11}(\underline{\theta}_0) & n^{-1} J_{n,12}(\underline{\theta}_0) & n^{-1} J_{n,13}(\underline{\theta}_0) \\ n^{-1} J_{n,21}(\underline{\theta}_0) & n^{-3/2} J_{n,22}(\underline{\theta}_0) & n^{-3/2} J_{n,23}(\underline{\theta}_0) \\ n^{-1} J_{n,31}(\underline{\theta}_0) & n^{-3/2} J_{n,32}(\underline{\theta}_0) & n^{-3/2} J_{n,33}(\underline{\theta}_0) \end{pmatrix}.$$

First, note that all the terms with z_k are $o_p(1)$ by Lemma 11, i.e.

$$\begin{array}{cc} n^{-1/2} \sum_{t=1}^n \sum_{k=1}^J C_{\beta\beta,k}(x_{1t}) z_{kt} x_{1t}^2 & n^{-1} \sum_{t=1}^n \sum_{k=1}^J C_{\beta\beta,k}(x_{1t}) z_{kt} x_{1t} x_{2t}', \\ n^{-3/2} \sum_{t=1}^n \sum_{k=1}^J C_{\beta\beta,k}(x_{1t}) z_{kt} x_{2t} x_{2t}' & n^{-1/2} \sum_{t=1}^n \sum_{k=1}^J C_{\beta\mu^i,k}(x_{1t}) z_{kt} x_{1t}, \\ n^{-1} \sum_{t=1}^n \sum_{k=1}^J C_{\beta\mu^i,k}(x_{1t}) z_{kt} x_{2t}' & n^{-1/2} \sum_{t=1}^n \sum_{k=1}^J C_{\mu^i\mu^i,k}(x_{1t}) z_{kt} \end{array}$$

are all $o_p(1)$ by Lemma 11. The asymptotic results of the remaining terms then follow Lemma 2 .

Q.E.D.

Proof of Theorem 5:

As in Park and Phillips (2000), we can apply Theorem 10.1 of Wooldridge (1994) to show that (13) holds and thus there is a consistent local solution to the likelihood equation. The proof follows precisely as in theorem 2 of Park and Phillips (2000) and is not repeated here.

Proof of Theorem 7:

Write

$$\begin{aligned} -[E_n^{-1} J_n(\hat{\theta}_n) E_n^{-1}]^{-1} &= -E_n D_n^{-1} G [D_n^{-1} G' J_n(\hat{\theta}_n) G D_n^{-1}]^{-1} G' D_n^{-1} E_n \\ &= -E_n D_n^{-1} G [D_n^{-1} J_n(\hat{\underline{\theta}}_n) D_n^{-1}]^{-1} G' D_n^{-1} E_n. \end{aligned}$$

By Theorem 4 and 5 we have

$$-D_n^{-1} J_n(\hat{\underline{\theta}}_n) D_n^{-1} = -D_n^{-1} J_n(\underline{\theta}_0) D_n^{-1} + o_p(1) \to_d Q,$$

and we have that $E_n D_n^{-1} G \to K$. Therefore,

$$-[E_n^{-1} J_n(\hat{\theta}_n) E_n^{-1}]^{-1} \to_d K Q^{-1} K',$$

as expected. In the case $-[E_n^{-1} \underline{J}_n(\hat{\theta}_n) E_n^{-1}]^{-1}$, as shown in the proof of Theorem 4, we have

$$D_n^{-1} J_n(\hat{\theta}_n) D_n^{-1} = [D_n^{-1} \underline{J}_n(\hat{\theta}_n) D_n^{-1}]^{-1} + o_p(1),$$

so that the same results hold.

Q.E.D.

Proof of Corollary 8

First consider $j = 0$:

$$\hat{P}_0 = P_0(x;\theta_0) + \begin{pmatrix} \frac{\partial P_0(x;\theta_n)}{\partial \beta_n} & \frac{\partial P_0(x;\theta_n)}{\partial \mu_n^1} \end{pmatrix} \begin{pmatrix} \hat{\beta}_n - \beta_0 \\ \hat{\mu}_n^1 - \mu_0^1 \end{pmatrix},$$

and we have

$$\begin{aligned} \frac{\partial P_0(x;\theta_n)}{\partial \beta_n} &= p_0(x;\theta_n)x, \\ \frac{\partial P_0(x;\theta_n)}{\partial \mu_n^1} &= -\sqrt{n} p_0(x;\theta_n), \end{aligned}$$

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

Then

$$n^{1/4}(\hat{P}_0 - P_0(x;\theta_0)) = \Gamma(0) \begin{pmatrix} n^{1/4}(\hat{\beta}_n - \beta_0) \\ n^{3/4}(\hat{\mu}_n - \mu_0) \end{pmatrix}.$$

The approach is similar for $1 \le j \le J$, except that in analyzing $\hat{P}_j(x;\hat{\theta}_n)$ we also need to take derivatives with respect to μ_n^{j+1} for $1 \le j \le J-1$.

Q.E.D.

Proof of Corollary 9

For $j = 0$,

$$\hat{\gamma}_{0,x} = \gamma_0(x;\theta_0) + \begin{pmatrix} \frac{\partial \gamma_0(x;\theta_0)}{\partial \beta_0} & \frac{\partial \gamma_0(x;\theta_0)}{\partial \mu_0^1} \end{pmatrix} \begin{pmatrix} \hat{\beta}_n - \beta_0 \\ \hat{\mu}_n^1 - \mu_0^1 \end{pmatrix},$$

and

$$\begin{aligned} \frac{\partial \gamma_0(x;\theta_0)}{\partial \beta_0} &= -p_0'(x;\theta_0)\beta_0 x' + p_0(x;\theta_0) I_m, \\ \frac{\partial \gamma_0(x;\theta_0)}{\partial \mu_0^1} &= f'(x'\beta_0 - \sqrt{n}\mu_0^1)\beta_0. \end{aligned}$$

Then,

$$n^{1/4}(\hat{\gamma}_{0,x} - \gamma_0(x;\theta_0)) = \Psi(0) \begin{pmatrix} n^{1/4}(\hat{\beta}_n - \beta_0) \\ n^{3/4}(\hat{\mu}_n - \mu_0) \end{pmatrix}.$$

Again, it would be the same for $1 \le j \le J$, except that for $\hat{\gamma}_j(x;\hat{\theta}_n)$ we also need to take derivatives with respect to μ_n^{j+1} for $1 \le j \le J-1$.

Q.E.D.

Proof of Theorem 10

Since the z_{kt} are martingale differences, we have

$$\begin{aligned} r_n(j) &= \frac{1}{n}\sum_{t=1}^{n}\Lambda(t,j) \\ &= \frac{1}{n}\sum_{t=1}^{n}P_j(x_t;\theta_0) + \frac{1}{n}\sum_{t=1}^{n}\left(P_j(x_t;\theta_0)\sum_{k=1}^{J}g_k(x_t;j,\theta_0)z_{kt}\right) \\ &= \frac{1}{n}\sum_{t=1}^{n}P_j(x_t;\theta_0) + o_p(1). \end{aligned}$$

Therefore,

$$\begin{aligned} r_n(0) &= 1 - \frac{1}{n}\sum_{t=1}^{n}F(x_t'\beta_0 - \sqrt{n}\mu_0^1)) + o_p(1) \\ &= \frac{1}{n}\sum_{t=1}^{n}1\{x_t'\beta_0 < \sqrt{n}\mu_0^1\} + o_p(1) \\ &= \frac{1}{n}\sum_{t=1}^{n}1\left\{\frac{x_t'\beta_0}{\sqrt{n}} < \mu_0^1\right\} + o_p(1). \end{aligned}$$

By virtue of Assumption 1, we have $\frac{x_t}{\sqrt{n}} \to_d V(r) =_d BM(\Sigma)$, and then $\frac{x_t'\beta_0}{\sqrt{n}} \to_d \beta_0'V(r) =_d BM(\beta_0'\Sigma\beta)$. Define ω_x such that $\omega_x W(r) =_d BM(\beta_0'\Sigma\beta_0)$. Therefore,

$$r_n(0) = \frac{1}{n}\sum_{t=1}^{n}1\left\{\frac{x_t'\beta_0}{\sqrt{n}} < \mu_0^1\right\} + o_p(1) \to_d \int_0^1 1\left\{W(r) < \frac{\mu_0^1}{\omega_x}\right\}\, dr.$$

Similarly, for $1 < j < J$,

$$\begin{aligned} r_n(j) &= \frac{1}{n}\sum_{t=1}^{n}F(x_t'\beta_0 - \sqrt{n}\mu_0^j) - \frac{1}{n}\sum_{t=1}^{n}F(x_t'\beta_0 - \sqrt{n}\mu_0^{j+1}) + o_p(1) \\ \to_d &\quad 1 - \int_0^1 1\left\{W(r) < \frac{\mu_0^j}{\omega_x}\right\}\, dr + \int_0^1 1\left\{W(r) < \frac{\mu_0^{j+1}}{\omega_x}\right\}\, dr \\ &= \int_0^1 1\left\{\frac{\mu_0^j}{\omega_x} < W(r) < \frac{\mu_0^{j+1}}{\omega_x}\right\}\, dr, \end{aligned}$$

and

$$
\begin{aligned}
r_n(J) &= \frac{1}{n}\sum_{t=1}^{n} F(x_t'\beta_0 - \sqrt{n}\mu_0^J) + o_p(1) \\
\to_d &\quad 1 - \int_0^1 1\left\{W(r) < \frac{\mu_0^J}{\omega_x}\right\} dr \\
&= \int_0^1 1\left\{W(r) > \frac{\mu_0^J}{\omega_x}\right\} dr,
\end{aligned}
$$

as expected.

The proof for the predicted proportion $r_n(j, X) = \frac{1}{n}\sum_{t=1}^{n} 1\{y_t(X) = j\}$ follows in the same manner. In the estimated case, $\hat{r}_n(j, X) = n^{-1}\sum_{t=1}^{n} \hat{P}_j(X_t; \hat{\theta}_n)$. By the mean value expansion as in the proof of Corollary 9,

$$\hat{r}_n(j, X) = r_n(j, X) + O_p(n^{-1/4}),$$

and thus $\hat{r}_n(j, X)$ has the same limit as $r_n(j, X)$.

Q.E.D.

APPENDIX C: NOTATION

$\to_{a.s}$	almost sure convergence.
$\to_p$	convergence in probabitlity.
$\to_d$	weak convergence.
$o_p(1)$	tends to zero in probabitlity.
$=_d$	distributional equivalence.
$\sim_d$	asymptotically distributed as.
W, V_1, V_2	standard Brownian motions.
$L_V(t, s)$	Local time of V at time t and spatial point s
$MN(0, V)$	mixed normal distribution with variance V.
$\Vert \cdot \Vert$	Euclidean norm in $\mathbf{R}^k$.
$\mathbf{F}_R$	class of reguar functions.
$\mathbf{F}_I$	class of bounded integrable functions.
$\mathbf{F}_0$	class of bounded functions vanishing at infinity.
$\mathcal{L}_\gamma^{-1}$	inverse Laplace transform with respect to γ.

APPENDIX D: SPECIAL FUNCTIONS

$$\mathrm{He}_n(x) := (-1)^n e^{x^2/2} \frac{d^n}{dx^n}(e^{-x^2/2})$$

$$h_v(n,x) := \mathcal{L}_\gamma^{-1}((2\gamma)^{n/2-1/2} e^{-v\sqrt{2\gamma}}) = \frac{e^{-v^2/2x}}{\sqrt{2\pi} x^{(n+1)/2}} \mathrm{He}_n\left(\frac{v}{\sqrt{x}}\right),\ 0 < v$$

$${}_1F_1\left(a,b;z\right) = \sum_{j=0}^{\infty} \frac{(a)_j}{(b)_j\, j!} z^j,\ \ (a)_j = a\left(a+1\right) ... \left(a+j-1\right)$$

$$D_v(x) = 2^{\frac{v}{2}} e^{-x^2/4} \left[\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{v+1}{2})}\ {}_1F_1\left(-\frac{v}{2}, \frac{1}{2}; \frac{x^2}{2}\right) + \frac{x}{2^{\frac{1}{2}}} \frac{\Gamma(-\frac{1}{2})}{\Gamma(-\frac{v}{2})}\ {}_1F_1\left(\frac{1}{2} - \frac{v}{2}, \frac{3}{2}; \frac{x^2}{2}\right)\right]$$

$$\mathrm{c}_y(\mu,\nu,t,z) = 2^\nu \sum_{k=0}^{\infty} \frac{(-1)^k \Gamma(\nu+k) e^{-(\nu t+z+2kt)^2/4y}}{\sqrt{2\pi} y^{1+\mu/2} \Gamma(\nu) k!} D_{\mu+1}\left(\frac{\nu t + z + 2kt}{\sqrt{y}}\right)$$

for $\nu \geq 0, \nu t + z > 0$

$$\mathrm{ec}_y(\mu,\nu,t,x,z) = \sum_{k=0}^{\infty} \frac{(-z)^k}{k!} \mathrm{c}_y(\mu+k, \nu+k, t, x+z+kt),\ \ \nu \geq 0,\ \ \nu t + x + z > 0$$

$$\mathrm{ce}_y^{(1)}(\nu,t,x,z) = \frac{1}{2}\mathrm{ec}_y(0,1,t,x-\nu,z) + \frac{1}{2}\mathrm{ec}_y(0,1,t,x+\nu,z)$$

for $t + x + z - \nu > 0, t > 0$

References

Borodin A. N. and P. Salminen (1996) *Handbook of Brownian Motion - Facts and Formulae* Birkhäuser Verlag: Boston

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and its Application.* Academic Press: New York.

Hu, L. and P. C. B. Phillips (2001). "Dynamics of the Federal Funds Target Rate: a Nonstationary Discrete Choice Approach". Yale University, mimeographed.

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). "Econometric analysis of the Fisher equation". Cowles Foundation Discussion Paper, No. 1180.

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.

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer-Verlag: New York.

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.