Econometrica, Vol. 68, No. 5 (September, 2000), 1249–1280

NONSTATIONARY BINARY CHOICE

BY JOON Y. PARK AND PETER C. B. PHILLIPS[1]

This paper develops an asymptotic theory for time series binary choice models with nonstationary explanatory variables generated as integrated processes. Both logit and probit models are covered. The maximum likelihood (ML) estimator is consistent but a new phenomenon arises in its limit distribution theory. The estimator consists of a mixture of two components, one of which is parallel to and the other orthogonal to the direction of the true parameter vector, with the latter being the principal component. The ML estimator is shown to converge at a rate of $n^{3/4}$ along its principal component but has the slower rate of $n^{1/4}$ convergence in all other directions. This is the first instance known to the authors of multiple convergence rates in models where the regressors have the same (full rank) stochastic order and where the parameters appear in linear forms of these regressors. It is a consequence of the fact that the estimating equations involve nonlinear integrable transformations of linear forms of integrated processes as well as polynomials in these processes, and the asymptotic behavior of these elements is quite different. The limit distribution of the ML estimator is derived and is shown to be a mixture of two mixed normal distributions with mixing variates that are dependent upon Brownian local time as well as Brownian motion.

It is further shown that the sample proportion of binary choices follows an arc sine law and therefore spends most of its time in the neighborhood of zero or unity. The result has implications for policy decision making that involves binary choices and where the decisions depend on economic fundamentals that involve stochastic trends. Our limit theory shows that, in such conditions, policy is likely to manifest streams of little intervention or intensive intervention.

KEYWORDS: Binary choice model, Brownian motion, Brownian local time, dual convergence rates, integrated time series, maximum likelihood estimation.

1. INTRODUCTION

BINARY CHOICE MODELS ARE NOW A STANDARD TOOL of microeconometrics and have been widely used in empirical research. While most of the empirical applications have been to cross-section data, there are many situations in time series modeling where binary choice dependent variables arise. Ongoing economic decisions by individual agents over time constitute prime examples: consumers decide to buy certain durable goods, firms choose to retain or fire employees, unions decide to strike or capitulate, women choose to join the workforce, and so on. At the national level, monetary authorities decide to intervene in the market for securities to change interest rates, or intervene in the foreign exchange market to influence exchange rates. One way of formally

[1] Park thanks the Department of Economics at Rice University, where he is an Adjunct Professor, for its continuing hospitality, and the Cowles Foundation for support during a visit in June, 1998. Phillips thanks the NSF for support under Grant Nos. SBR 94-22922 and SBR 97-30295. The paper was typed by the authors in SW2.5 and computational work was performed in GAUSS.

characterizing these situations is through a binary choice model with covariates that are relevant to the choice being made. In such time series applications, the covariates will often involve nonstationary data reflecting relevant economic conditions. For instance, monetary authority decisions can be assumed to be based on macroeconomic fundamentals, which will include such variables as real output and its components, inflation, unemployment rates, and other indicators of economic conditions and performance. Many of these variables display nonstationary characteristics. In such situations, we may well expect the econometric theory of binary choice models to be different from that of traditional cross-section theory which relies heavily on the simplifying assumption of independence across observations.

This paper seeks to develop a new theory for binary choice regressions that accommodates nonstationary data. In particular, we study binary choice models with covariates that are integrated time series and develop a new limit theory for the maximum likelihood (ML) estimation of such systems. Since ML estimation in discrete choice models involves nonlinear optimization, the asymptotic theory in this paper involves asymptotics for nonlinear functions of integrated time series. Some recent work by the authors (Park and Phillips (1998, 1999)) has provided techniques for analyzing nonlinear regressions with nonstationary time series, and this paper shows how to utilize some of those techniques in the context of discrete choice models. The main theoretical contribution of the paper is to derive an asymptotic theory for the ML estimation of nonlinear discrete choice models with covariates that are integrated time series. Some of the results obtained here are directly applicable in the wider context of M estimation. They also lay the groundwork for the asymptotic analysis of index models consisting of integrated time series.

One of the main findings of the paper is that there are dual rates of convergence in binary choice models with integrated regressors. There is a fast rate of convergence of $n^{3/4}$ in a direction that is orthogonal to that of the true coefficient vector β_0. A slower rate of convergence of $n^{1/4}$ applies in all other directions. Such dual convergence rates are unexpected in econometric models where the covariates have the *same* full rank stochastic order, as they do here, and the parameters appear in linear forms of these regressors. In fact, to the best of our present knowledge, this is the first instance of the phenomenon in asymptotic statistical theory. The explanation for the phenomenon in the present case is that the nonlinearity arising from the discrete choice probability framework confabulates the signal from the regressors. Since the signal works through the probability function, it involves sample moments of the covariates x_t in conjunction with scalar functionals of a distribution function evaluated at the linear form $x_t'\beta_0$. These nonlinear sample moment functionals have stochastic orders that depend on the direction in which they are evaluated. In effect, there is more sample information about β in directions orthogonal to β_0 than there are in other directions and, in particular, along β_0. Heuristically, the signal from x_t is attenuated along β_0 because large deviations of $x_t'\beta_0$ contribute less to the sample second moment since they are attenuated by the scaling of a density function that is evaluated in the same direction and that, by its very nature, downweights large deviations. Thus, by virtue of the formulation of the model, and somewhat ironically, there is less information in the data in the direction of the true parameter than there is in the orthogonal direction.

Another finding of the paper is that in nonstationary binary choice the sample proportion of binary choices follows an arc sine law and therefore spends most of its time in the neighborhood of zero or unity, just as a random walk spends most of its time on one side of the origin or the other. This result has some testable empirical implications for policy decision making. In particular, if policy involves market interventions and these are influenced by any factor that involves a stochastic trend, then the theory suggests that market interventions will most likely occur in streams of little intervention or large numbers of interventions.

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 gives a brief numerical illustration of the effects on nonstationarity on logit and probit estimates. Section 5 concludes. Some useful lemmas are given in Appendix A, Appendix B gives proofs of the main theorems, and notation is summarized in Appendix C.

2. THE MODEL, ASSUMPTIONS, AND PRELIMINARY RESULTS

We consider the regression model given by

$$y_t^* = x_t' \beta_0 - \varepsilon_t, \tag{1}$$

where x_t is a vector of explanatory variables and ε_t is an error. The dependent variable y_t^* is assumed to be latent and the observed variable is simply the indicator

$$y_t = 1\{ y_t^* \geq 0\}. \tag{2}$$

The model given by (1) and (2) is a standard binary choice model. As usual, we assume that x_t is predetermined, i.e., x_{t+1} is adapted to some filtration (F_t), with respect to which ε_t is measurable.

The theory of the binary choice model in (1) and (2) when x_t is a stationary and ergodic process or short memory time series is obtained by standard methods. The simple case where x_t is iid is included in many econometrics texts (e.g., Amemiya (1985)) and review articles (Dhrymes (1986)), while dynamic cases with weakly dependent data are covered in the theory in Wooldridge (1994) and White (1994). The present paper looks, instead, at the case where x_t is nonstationary. In particular, we assume that x_t is an integrated time series, possibly of the ARIMA type. All the remaining features of the model are identical to those of the standard parametric binary choice model. Thus, we let ε_t be iid conditionally on F_{t-1} with distribution function F, which is assumed to be known and standardized, the primary cases of interest being the standard normal (leading to the probit model) and the standard logistic (leading to the logit model). Also, we let β_0 be an interior point of a subset of $\mathbf{R}^m$ which we assume to be compact and convex.

In this framework we have $\mathbf{E}(y_t \mid F_{t-1}) = \mathbf{P}(y_t = 1 \mid F_{t-1}) = F(x_t'\beta_0)$. Then, defining u_t as the residual in

$$(3) \qquad y_t = F(x_t'\beta_0) + u_t,$$

it is apparent that (u_t, F_t) is a martingale difference sequence with conditional variance

$$(4) \qquad \mathbf{E}(u_t^2 \mid F_{t-1}) = \sigma^2(x_t'\beta_0),$$

where $\sigma^2(z) = F(z)(1 - F(z))$. If the conditioning set or the dynamics are misspecified and $\mathbf{E}(y_t \mid F_{t-1}) \neq F(x_t'\beta_0)$, then u_t in (3) is no longer a martingale difference and the theory developed here does not cover that case. Thus, the present paper contributes to correctly specified dynamic binary choice models of given (distributional) functional form.

The following assumption on the process generating x_t underpins the asymptotic development. In particular, the linear process structure and the moment conditions on the innovations assist in the use of embedding arguments that allow for stochastic process representations of key partial sum processes, as shown in Lemma 1 below.

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

$$v_t = \Pi(L)e_t = \sum_{i=0}^{\infty} \Pi_i e_{t-i},$$

with $\Pi(1)$ *nonsingular and* $\sum_{i=0}^{\infty} i\|\Pi_i\| < \infty$. *The innovations* e_t *are iid with mean zero and* $\mathbf{E}\|e_t\|^r < \infty$ *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* $(\Omega, F, \mathbf{P})$ *supporting sequences of random variables* U_{nt} *and* V_{nt} *satisfying the following*:

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

$$(U_{nt}, V_{nt}) =_d \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{t} u_i, \frac{1}{\sqrt{n}} \sum_{i=1}^{t} v_i \right).$$

(b) *There exists a representation*

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

with standard Brownian motion U *and time changes* T_{nt} *in* $(\Omega, F, \mathbf{P})$. *Let* $T_{nt} = \sum_{i=1}^{t} \tau_{ni}$ *and define* $F_{nt} = \sigma((U(r))_{r=1}^{T_{nt}/n}, (V_{ns})_{s=1}^{t+1})$. *Then* $\mathbf{E}(\tau_{nt} \mid F_{n,t-1}) = \mathbf{E}(u_t^2 \mid$

$F_{t-1})$ *and* $\mathbf{E}(\tau_{nt}^r \mid F_{n,t-1}) \leq c_r \mathbf{E}(|u_t|^{2r} \mid F_{t-1})$ *for all* $r \geq 1$, *where* c_r *is some constant depending only upon* r.

(c) *Defining*

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

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* $(\Omega, F, \mathbf{P})$ *with variance matrix* Σ.

For the development of our theory, it is convenient to assume that $\beta_0 \neq 0$ and to rotate the regressor space using an orthogonal matrix $H = (h_1, H_2)$ with $h_1 = \beta_0 / (\beta_0' \beta_0)^{1/2}$. We then define

(5) $\qquad x_{1t} = h_1' x_t \qquad$ and $\qquad x_{2t} = H_2' x_t.$

Conformable with this rotation, define

(6) $\qquad V_1 = h_1' V \qquad$ and $\qquad V_2 = H_2' V,$

which are Brownian motions of dimensions 1 and $(m-1)$, respectively.

Our subsequent theory involves the local time of the 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), Phillips and Park (1998), and Park and Phillips (1998, 1999) for applications of this process in time series. In the representation of our limit theory it is especially convenient to use the scaled local time of V_1 given by

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

where σ_{11} is the variance of V_1. The process $L_1(t,s)$ is called chronological local time in Phillips and Park (1998) because it measures the sojourn time in chronological units that the process spends in the vicinity of the spatial point s. It is defined by the limit

$$L_1(t,s) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^t 1\{|V_1(r) - s| < \varepsilon\} \, dr.$$

The notation in (6) and (7) will be used frequently in the paper without any further reference.

Since we will be dealing with nonlinear functions of the integrated process x_t, it aids our theory to be more specific about the class of functions we will consider. In the development that follows we draw upon the general approach of Park and Phillips (1999) in studying nonlinear transformations of integrated processes. Here we will concentrate on functions that typically arise in the context of binary choice models such as (3) above, viz. distribution functions, probability densities and their derivatives. These functions play an important role in the subsequent development of our theory.

A function $f:\mathbf{R}\to\mathbf{R}$ will be 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, $\mathbf{F}_R\subset\mathbf{F}_I\subset\mathbf{F}_0$. Further, we define f_κ by $f_\kappa(x)=x^\kappa f(x)$ for any function f. By convention, $f_\kappa\in\mathbf{F}_R$, $\mathbf{F}_I$ or $\mathbf{F}_0$ when $f_i\in\mathbf{F}_R$, $\mathbf{F}_I$ or $\mathbf{F}_0$ for all $i=0,\ldots,\kappa$. The first and second derivatives of f are denoted, respectively, by $\dot f$ and $\ddot f$, when they exist.

We now make some assumptions on the distribution function F of ε_t. We will make extensive use of the following notation:

$$G=\dot F/F(1-F);\; K=G\dot F=G^2F(1-F). \tag{8}$$

Assumption 2: *F is three times differentiable so that $\dot F$, $\ddot F$, $\dot G$ and $\ddot G$ all exist. Further*: (*a*) $K_2\in\mathbf{F}_R$, (*b*) $\dot F_1$, $(\dot G\dot F)_2$, $(G\ddot F)_2$, $(\ddot G F^{1/2}(1-F)^{1/2})_2\in\mathbf{F}_I$, *and* (*c*) $(\dot G F^{1/2}(1-F)^{1/2})_2$, $(G^3\dot F)_4\in\mathbf{F}_0$.

For the logit model, $F(x)=e^x/(1+e^x)$ and we have $G=1$. Consequently, $K=\dot F$ and so $K(x)=e^x/(1+e^x)^2$, the density of the logistic distribution. Condition (a) of Assumption 2 is therefore clearly satisfied. Moreover, since $\dot G=0$, it is simple to check the conditions in (b) and (c) of Assumption 2, all of which hold trivially.

For the probit model, it is also not difficult to verify that the conditions in Assumption 2 hold. Indeed, using Mills ratio we have

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

where φ is the standard normal density function and Φ is the corresponding cumulative distribution function. It is apparent from these formulae that $G(x)=O(|x|)$ for large $|x|$, and that G is differentiable with a bounded derivative. It further follows that

$$K(x)=G(x)\dot F(x)=\frac{\varphi(x)^2}{\Phi(x)(1-\Phi(x))}=\begin{cases}\dfrac{x\varphi(x)}{\Phi(x)}[1+O(x^{-2})] & \text{as } x\to\infty,\\[2ex] \dfrac{-x\varphi(x)}{1-\Phi(x)}[1+O(x^{-2})] & \text{as } x\to-\infty,\end{cases}$$

and so K is regular and satisfies condition (a) of Assumption 2. Conditions (b) and (c) of Assumption 2 follow upon some further routine calculations.

For each t, the conditional log likelihood of y_t given F_{t-1} is $y_t \log F(x_t'\beta) + (1-y_t)\log[1-F(x_t'\beta)]$. The implied (conditional) log likelihood for the model (1) and (2) has the familiar form

$$\log L_n(\beta) = \sum_{t=1}^{n} y_t \log F(x_t'\beta) + \sum_{t=1}^{n} (1-y_t)\log[1-F(x_t'\beta)], \tag{9}$$

and this function is well known to be globally concave in the logit and probit cases. For example, in the logit case we have (c.f. Amemiya (1985, p. 273))

$$\frac{\partial^2 \log L_n}{\partial\beta\partial\beta'} = -\sum_{t=1}^{n} \Lambda_t(1-\Lambda_t)x_t x_t', \qquad \Lambda_t = \Lambda(x_t'\beta) = \frac{e^{x_t'\beta}}{1+e^{x_t'\beta}}, \tag{10}$$

which is negative definite for all β, just as it is when the covariates comprise cross-section observations.

The hessian (10) is a weighted sample second moment of the integrated process x_t. The weights are given by the logistic density at $x_t'\beta$, viz.

$$\Lambda_t(1-\Lambda_t) = \frac{e^{x_t'\beta}}{(1+e^{x_t'\beta})^2},$$

and are nonlinear integrable functions of x_t. In contrast to the case of cross-section data that is studied in the literature, or the case where x_t is a stationary and ergodic time series, the normed hessian

$$\frac{1}{n}\frac{\partial^2 \log L_n}{\partial\beta\partial\beta'} = -\frac{1}{n}\sum_{t=1}^{n} \Lambda_t(1-\Lambda_t)x_t x_t'$$

does not converge in probability to a negative definite matrix when x_t is an integrated process. In fact, as we will show, the hessian matrix $\partial^2 \log L_n/\partial\beta\partial\beta'$ has elements with different stochastic orders in different directions and, when appropriately normed to accommodate this divergence, the matrix converges weakly to a random limit matrix, not a constant matrix. The random limit matrix is negative definite almost surely and so the limit function is also globally concave.

3. MAIN RESULTS

Let $\hat{\beta}_n$ be the maximum likelihood estimator of β_0 in (1). From (9) and using the notation (8) we can write the score $S_n(\beta)$ and hessian $J_n(\beta)$ as

$$S_n(\beta) = \sum_{t=1}^{n} G(x_t'\beta)x_t(y_t - F(x_t'\beta)), \tag{11}$$

$$J_n(\beta) = -\sum_{t=1}^{n} K(x_t'\beta)x_t x_t' + \sum_{t=1}^{n} \dot{G}(x_t'\beta)x_t x_t'(y_t - F(x_t'\beta)). \tag{12}$$

As usual in ML limit theory, the asymptotic distribution of $\hat{\beta}_n$ will be obtained from the expansion

$$0 = S_n(\hat{\beta}_n) = S_n(\beta_0) + J_n(\beta_n)(\hat{\beta}_n - \beta_0), \tag{13}$$

where $S_n(\hat{\beta}_n)$ and $S_n(\beta_0)$ are the scores at $\hat{\beta}_n$ and β_0 respectively, and $J_n(\beta_n)$ is the hessian matrix with rows evaluated at mean values between $\hat{\beta}_n$ and β_0. As it stands, this is a conventional problem in extremum estimation. However, in the present case, the limits of the score $S_n(\beta_0)$ and the hessian $J_n(\beta_0)$ are nonstandard and our first step is to characterize these limits.

To analyze the asymptotic behavior of $S_n(\beta_0)$ and $J_n(\beta_0)$, we need to rotate the coordinate system and reparameterize the model. Corresponding to the transformation of x_t in (5), let

$$\alpha_1 = h_1' \beta \qquad \text{and} \qquad \alpha_2 = H_2' \beta,$$

and define $\alpha = (\alpha_1, \alpha_2')' = H'\beta$ and $\alpha_0 = (\alpha_1^0, \alpha_2^{0\prime})' = ((\beta_0' \beta_0)^{1/2}, 0)' = H'\beta_0$. Since maximum likelihood estimation is invariant with respect to reparameterization, if $\hat{\alpha}_n$ is the maximum likelihood estimator of α_0, then $\hat{\alpha}_n = H'\hat{\beta}_n$ or $\hat{\beta}_n = H\hat{\alpha}_n$. The score function $S_n(\alpha)$ and hessian $J_n(\alpha)$ for the parameter α can be obtained from those of α simply by using the relationships $S_n(\alpha) = H'S_n(\beta)$ and $J_n(\alpha) = H'J_n(\beta)H$. Furthermore, by premultiplying the expansion (13) by H' it is apparent that the expansion remains valid for $S_n(\hat{\alpha}_n)$, i.e.,

$$0 = S_n(\hat{\alpha}_n) = S_n(\alpha_0) + J_n(\alpha_n)(\hat{\alpha}_n - \alpha_0), \tag{14}$$

where $J_n(\beta_n)$ is replaced by $J_n(\alpha_n) = H'J_n(\beta_n)H$.

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 (11) and hessian (12).

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

(a) $$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} f(x_{1t}) \to_d L_1(1,0) \int_{-\infty}^{\infty} f(s)\, ds,$$

(b) $$\frac{1}{n} \sum_{t=1}^{n} f(x_{1t}) x_{2t} \to_d \int_0^1 V_2(r)\, dL_1(r,0) \int_{-\infty}^{\infty} f(s)\, ds,$$

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

jointly as $n \to \infty$.

LEMMA 3: *Let Assumption* 1 *hold, and assume* $\sigma^2 f^2, \sigma^2 g^2 \in \mathbf{F}_R$, $\sigma^2 f, \sigma^2 g \in \mathbf{F}_I$ *and* $\sigma^2 f^4, \sigma^2 g^4 \in \mathbf{F}_0$ *for* $f, g : \mathbf{R} \to \mathbf{R}$. *Then we have*

$$\begin{pmatrix} n^{-1/4} \sum_{t=1}^{n} f(x_{1t}) u_t \\ n^{-3/4} \sum_{t=1}^{n} g(x_{1t}) x_{2t} u_t \end{pmatrix} \to_d M^{1/2} W(1), \tag{15}$$

where

$$M = \begin{pmatrix} L_1(1,0)\int_{-\infty}^{\infty} f_\sigma^2(s)\,ds & \int_0^1 dL_1(r,0)V_2(r)'\int_{-\infty}^{\infty} f_\sigma(s)g_\sigma(s)\,ds \\ \int_0^1 V_2(r)\,dL_1(r,0)\int_{-\infty}^{\infty} g_\sigma(s)f_\sigma(s)\,ds & \int_0^1 V_2(r)V_2(r)'\,dL_1(r,0)\int_{-\infty}^{\infty} g_\sigma^2(s)\,ds \end{pmatrix},$$

with $f_\sigma = \sigma f$, $g_\sigma = \sigma g$ *and* W *is m-dimensional Brownian motion with covariance matrix* I, *which is independent of* V.

REMARKS: 1. 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 the covariance of V_1 and V_2. 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., a result that is easy to deduce because $\{r : V_1(r) = 0\}$ is the support of the measure $dL_1(r,0)$ (e.g. Revuz and Yor (1994, Ch. VI)). We may therefore regard V_2 as being independent of V_1 (and hence of L_1) in so far as the process V_2 arises in the representations of the limit distributions in Lemmas 2 and 3. The limiting distribution in Lemma 3 is mixed Gaussian. The mixing variates are dependent upon the local time L_1 of V_1 as well as V_2. We denote the limit distribution by MN(0, M).

2. Note that the components in (15) are asymptotically dependent in general, in spite of their differing rates of convergence. A special case in applications when the conditional covariance matrix M is block diagonal is discussed below.

3. In part (a) of Lemma 2 the sample mean of $f(x_{1t})$ is standardized by $\sqrt{n}$ rather than the usual n. This reduction in norming arises because the function f is integrable and therefore attenuates contributions from large values of x_{1t}. The increase in the norming factor that appears in parts (b) and (c) arises from the presence of the integrated regressor x_{2t} in the sample moment.

4. 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 return to $\sqrt{n}$. Thus, suppose x_{3t} is stationary, satisfies the same conditions as v_t in Assumption 1 and is independent of u_t. Then, a simple derivation along the lines of the proof of Lemma 2 reveals that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} f(x_{1t})x_{3t}x_{3t}' \to_d L_1(1,0)\int_{-\infty}^{\infty} f(s)\,ds\Sigma_{33}, \tag{16}$$

where $\Sigma_{33} = \mathbf{E}(x_{3t}x_{3t}')$. Likewise, we find that

$$n^{-1/4}\sum_{t=1}^{n} g(x_{1t})x_{3t}u_t \to_d \text{MN}\left(0,\ L_1(1,0)\int_{-\infty}^{\infty} g_\sigma^2(s)\,ds\Sigma_{33}\right). \tag{17}$$

Thus, the presence of stationary elements in the regressors does not affect the $\sqrt{n}$ convergence rate of the sample mean in Lemma 2(a) or the $\sqrt[4]{n}$ convergence rate of the sample covariance in Lemma 3(a). Results (16) and (17) ensure that the coefficients of stationary regressors in binary choice models where there are also some integrated regressors behave in a similar way to the coefficients of x_{1t}.

Using these results we are now able to characterize the limit forms of the score function (11) and hessian (12).

THEOREM 1: *Let Assumptions* 1 *and* 2 *hold. Then*

$$D_n^{-1} S_n(\alpha_0) \to_d Q^{1/2} W(1) \qquad \text{and} \qquad D_n^{-1} J_n(\alpha_0) D_n^{-1} \to_d -Q,$$

jointly, where $D_n = \operatorname{diag}(n^{1/4}, n^{3/4} I_{m-1})$,

$$Q = \begin{pmatrix} L_1(1,0)\int_{-\infty}^{\infty} s^2 K(\alpha_1^0 s)\, ds & \int_0^1 dL_1(r,0) V_2(r)' \int_{-\infty}^{\infty} sK(\alpha_1^0 s)\, ds \\ \int_0^1 V_2(r)\, dL_1(r,0) \int_{-\infty}^{\infty} sK(\alpha_1^0 s)\, ds & \int_0^1 V_2(r) V_2(r)'\, dL_1(r,0) \int_{-\infty}^{\infty} K(\alpha_1^0 s)\, ds \end{pmatrix}, \tag{18}$$

$\alpha_1^0 = (\beta_0' \beta_0)^{1/2}$, *and* W *is defined as in Lemma* 3.

If ε_t has a symmetric distribution, as in the probit and logit models, K is an even function. We therefore have

$$\int_{-\infty}^{\infty} sK(s)\, ds = 0.$$

In this case, the matrix Q given in Theorem 2 reduces to a block diagonal matrix.

The asymptotic results for $S_n(\alpha_0)$ and $J_n(\alpha_0)$ presented in Theorem 1 aid in deriving the limiting distribution of $\hat{\alpha}_n$. Indeed, from the expansion (14) we may expect that the normed and centered estimator satisfies

$$D_n(\hat{\alpha}_n - \alpha_0) = -\left(D_n^{-1} J_n(\alpha_0) D_n^{-1}\right)^{-1} D_n^{-1} S_n(\alpha_0) + o_p(1). \tag{19}$$

Indeed, (19) is established in the proof of Theorem 2 below. This expansion is enough to deliver the form of the limit distribution of $D_n(\hat{\alpha}_n - \alpha_0)$.

THEOREM 2: *Let Assumptions* 1 *and* 2 *hold. Then, there exists a sequence of ML estimators for which* $\hat{\alpha}_n \to_p \alpha_0$, *and*

$$D_n(\hat{\alpha}_n - \alpha_0) \to_d Q^{-1/2} W(1),$$

in the notation introduced in Theorem 1.

REMARKS: 1. As usual for local extremum estimation problems, Theorem 2 establishes the existence of a consistent root of the likelihood equation. The log likelihood function is well known to be concave in the logit and probit cases (and, indeed, in all cases where $\log F(x)$ and $\log(1 - F(x))$ are concave functions; see Pratt (1981)), and in such cases the consistent root is unique and is the global maximum.

2. Let $\hat{\alpha}_n = (\hat{\alpha}_{1n}, \hat{\alpha}_{2n}')'$. When $\int_{-\infty}^{\infty} sK(s)\,ds = 0$, we have the limits

$$n^{1/4}(\hat{\alpha}_{1n} - 1) \to_d \left(L_1(1,0) \int_{-\infty}^{\infty} s^2 K(s)\,ds \right)^{-1/2} W_1(1),$$

$$n^{3/4}\hat{\alpha}_{2n} \to_d \left(\int_0^1 V_2(r) V_2(r)'\,dL_1(r,0) \int_{-\infty}^{\infty} s^2 K(s)\,ds \right)^{-1/2} W_2(1),$$

where $W = (W_1, W_2')'$ for W defined in Theorem 2. The limiting distributions of $\hat{\alpha}_{1n}$ and $\hat{\alpha}_{2n}$ are therefore dependent only through their mixing variates given by V. Consequently, in this case $\hat{\alpha}_{1n}$ and $\hat{\alpha}_{2n}$ become asymptotically independent conditional on x_t.

3. It follows from Theorem 2 that

$$D_n H'\left(\hat{\beta}_n - \beta_0\right) \to_d Q^{-1/2} W(1) = \mathrm{MN}(0, Q^{-1}). \tag{20}$$

Setting $E_n = n^{-1/4} D_n = \mathrm{diag}(1, \sqrt{n}\, I_{m-1})$ we have

$$(D_n H' n^{-1/4})^{-1} = H E_n^{-1} \to (h_1, 0)$$

so that

$$\sqrt[4]{n}\left(\hat{\beta}_n - \beta_0\right) \to_d (h_1, 0) Q^{-1/2} W(1) = \mathrm{MN}(0, (h_1,0) Q^{-1} (h_1, 0)')$$
$$= \mathrm{MN}(0, h_1 h_1' q_{11.2}^{-1}),$$

where

$$q_{11.2} = q_{11} - q_{12} Q_{22}^{-1} q_{21}, \tag{21}$$

which is expressed in terms of the elements of the partitioned matrix

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

with Q defined in (18). We formalize this result as follows.

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

$$\sqrt[4]{n}\left(\hat{\beta}_n - \beta_0\right) \to_d P_{\beta_0}\, \mathrm{MN}(0, q_{11.2}^{-1}),$$

where $q_{11.2}$ *is given by* (21) *and* (22) *and* $P_{\beta_0} = \beta_0(\beta_0'\beta_0)^{-1}\beta_0'$.

According to this Corollary we have the asymptotic approximation

$$\hat{\beta}_n \sim_d \mathrm{MN}\left(\beta_0, \frac{1}{\sqrt{n}} P_{\beta_0} q_{11.2}^{-1} \right). \tag{23}$$

A natural estimate of the (conditional) covariance matrix of $\hat{\beta}_n$ is provided by the hessian inverse $-J_n(\hat{\beta}_n)^{-1}$, or the more commonly used alternative $-\underline{J}_n(\hat{\beta}_n)^{-1}$, where

$$\underline{J}_n(\beta) = -\sum_{t=1}^{n} K(x_t'\beta)x_t x_t'$$

is the negative definite first component of $J_n(\beta)$ in (12). The next result shows that these matrices accurately represent the asymptotic covariance matrix of $\hat{\beta}_n$ in (23).

THEOREM 3: *Under Assumptions* 1 *and* 2,

$$-\left[n^{-1/2}J_n\left(\hat{\beta}_n\right)\right]^{-1}, -\left[n^{-1/2}\underline{J}_n\left(\hat{\beta}_n\right)\right]^{-1} \to_d P_{\beta_0} q_{11.2}^{-1}$$

as $n \to \infty$.

It follows that the usual construction for asymptotic standard errors of $\hat{\beta}_n$ applies, even though there are different rates of convergence in the components of the estimator. The situation is analogous to that considered by Park and Phillips (1989) in regressions with cointegrated regressors. Moreover, by (20) the limit distribution of $\hat{\beta}_n$ is mixed normal in all directions, albeit at different rates, and this fact ensures that Wald tests of restrictions on β_0 have asymptotic chi-squared distributions, so that statistical inference can proceed in the usual manner.

COROLLARY 2: *Let Assumptions* 1 *and* 2 *hold. If* $\int_{-\infty}^{\infty} sK(s)\,ds = 0$, *then*

$$\sqrt[4]{n}\left(\hat{\beta}_n - \beta_0\right) \to_d \frac{\beta_0}{\|\beta_0\|}\left(L_1(1,0)\int_{-\infty}^{\infty} s^2 K(s\|\beta_0\|)\,ds\right)^{-1/2} W(1)$$

where $\|\beta_0\| = (\beta_0'\beta_0)^{1/2}$, *and* W *is univariate standard Brownian motion independent of* V.

It is clear from Corollaries 1 and 2 that $\hat{\beta}_n$ is asymptotically dominated by the component that converges slower. The overall convergence rate is thus given by $n^{1/4}$, but the limit distribution is singular. It is degenerate along the direction that is orthogonal to the true parameter vector β_0, and, in that direction, $\hat{\beta}_n$ converges at the faster rate of $n^{3/4}$.

Also of interest are the predicted probabilities $\hat{F} = F(x_t'\hat{\beta}_n)$ and the estimated marginal effects $f(x_t'\hat{\beta}_n)\hat{\beta}_n$, where f is the density function corresponding to F. The limit theory for these functionals is given in the following result.

COROLLARY 3: *Let Assumptions* 1 *and* 2 *hold. Given* $x_t = x$, *the predicted probability* $\hat{F} = F(x'\hat{\beta}_n)$ *and estimated marginal effect* $\hat{\gamma}_x = f(x'\hat{\beta}_n)\hat{\beta}_n$ *have the* *following asymptotic distributions as* $n \to \infty$

$$F\left(x'\hat{\beta}_n\right) \sim_d \mathrm{MN}\left(F(x'\beta_0), \frac{1}{\sqrt{n}} f(x'\beta_0)^2 x' P_{\beta_0} x q_{11.2}^{-1}\right), \tag{24}$$

and

$$\hat{\gamma}_x \sim_d \mathrm{MN}\left(\gamma(x'\beta_0), \frac{1}{\sqrt{n}} [f(x'\beta_0) + f'(x'\beta_0) x'\beta_0]^2 P_{\beta_0} q_{11.2}^{-1}\right), \tag{25}$$

where $\gamma(x'\beta_0) = f(x'\beta_0)\beta_0$.

In both cases, the rate of convergence is $n^{1/4}$. However, the limit distribution of the estimated marginal effects is singular and the faster rate $n^{3/4}$ applies in directions orthogonal to β_0. In the Probit case, the asymptotic variance has the very simple form

$$\frac{1}{\sqrt{n}} \varphi(x'\beta_0)^2 \left[1 - (x'\beta_0)^2\right]^2 P_{\beta_0} q_{11.2}^{-1}.$$

The asymptotic variance formulae given in (24) and (25) correspond to those given in the literature for the iid or stationary case (e.g. Greene (1997)), although the singularity in (25) does not arise in that case. Notwithstanding the singularity, one can estimate the asymptotic variance matrix of the coefficient estimator by the inverse of the hessian as indicated in Theorem 3, and it is apparent that standard errors for the predicted probabilities and the estimated marginal effects may be computed in the usual manner. Thus, the main effect of nonstationarity is to slow down the rate of convergence in these estimates.

Finally, it is of interest to study the asymptotic behavior of the empirical average $r_n = n^{-1}\sum_{t=1}^n y_t$. The quantity r_n is an aggregate proportion and measures the proportion of positive choices (i.e. $y_t = 1$ outcomes) in the sample data. It can also be used in a predictive manner to forecast the proportion of positive choices for some given sequence of data on the covariates, say $X = \{X_t : t = 1, \dots, n\}$. In this event, we can define $y_t(X) = 1\{X_t'\beta_0 \geq \varepsilon_t\}$. Of course, in practical applications $y_t(X)$ is unobserved, and we would therefore use the estimate $\hat{r}_n(X) = n^{-1}\sum_{t=1}^n \hat{F}_t(X)$, where $\hat{F}_t(X) = F(X_t'\hat{\beta}_n)$. Such estimates are useful in policy situations in assessing the likely proportionate number of positive choices arising in the event of the data $\{X_t : t = 1, \dots, n\}$. An example in monetary policy would be the likely number of market interventions needed for a given scenario of economic fundamentals.

The following result gives the limit theory for such quantities.

THEOREM 4: *Let Assumptions* 1 *and* 2 *hold. Suppose the time series* $X = \{X_t : t = 1, \dots, n\}$ *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* $r_n = n^{-1}\sum_{t=1}^n y_t$, *the predicted proportion* $r_n(X) = n^{-1}\sum_{t=1}^n y_t(X)$, *and the estimated proportion*

$\hat{r}_n(X) = n^{-1}\sum_{t=1}^{n} \hat{F}_t(X_t' \hat{\beta}_n)$ *all have the following limit behavior as* $n \to \infty$:

$$r_n, r_n(X), \hat{r}_n(X) \to_d \int_0^\infty L_1(1, s)\, ds. \tag{26}$$

An elementary calculation shows that $L_1(1, s) =_d \sigma^{-1} L(1, s/\sigma)$ where $\sigma = \sigma_{11}^{1/2}$ and $L(1, s)$ is the local time of standard Brownian motion. Then

$$\int_0^\infty L_1(1, s)\, ds =_d \sigma^{-1} \int_0^\infty L\left(1, \frac{s}{\sigma}\right) ds$$
$$= \int_0^\infty L(1, t)\, dt = \int_0^1 1\{W(r) > 0\}\, dr,$$

a quantity that is well known (e.g., Revuz and Yor (1994, p. 232–233)) to be a random variable that follows the arc sine law with probability density

$$\frac{1}{\pi\sqrt{x(1-x)}} \tag{27}$$

on $[0,1]$. We deduce that the empirical average r_n and predictive averages $r_n(X)$, and $\hat{r}_n(X)$ all have the same limit distribution given by the arc sine law (27). This result is decidedly different from the stationary case, where $r_n \to_p \mathbf{E}[F(x_t' \beta_0)]$ and the expectation is taken with respect to the stationary distribution of x_t.

The fact that r_n follows an arc sine law in the limit means that the sample proportion of positive choices spends most of its time in the neighborhood of zero or unity, just as a random walk spends most of its time on one side of the origin or the other. To take the explicit example of a monetary policy intervention, the theorem has the following implication. If the economic fundamentals that determine monetary policy intervention include a stochastic trend, then the likely number of market interventions needed for any given scenario of fundamentals is determined by the arc sine law (27). Thus, the theorem tells us that it is most likely that interventions will occur in streams giving periods where there is very little intervention or periods where there are a large number of interventions. This result holds for any given trajectory of fundamentals and for any particular policy determining mechanism, i.e., any linear form $x_t' \beta_0$ provided it has a stochastic trend.

4. ILLUSTRATION OF THE EFFECTS OF NONSTATIONARITY

This section reports a brief numerical exercise that illustrates the effects of nonstationarity on a binary choice regression. Data were generated from (1) and (2) using both logit and probit formulations for F and with exogenous covariates x_t, which were generated by a bivariate vector autoregression of the form

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

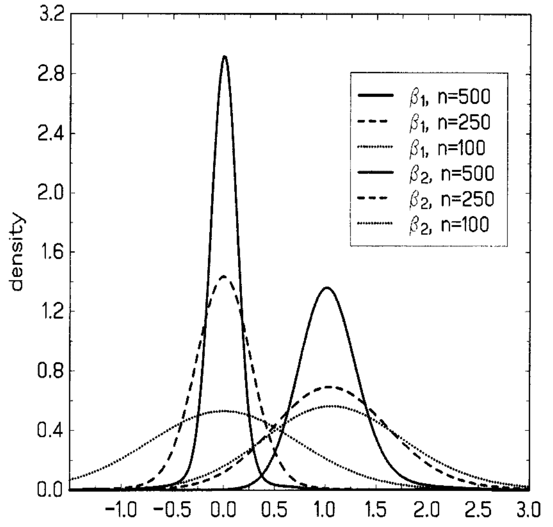

FIGURE 1.—Logit model—densities of estimators of $\beta_1^0 = 1$, $\beta_2^0 = 0$; $I(1)$ case.

with $v_t = (v_{1t}, v_{2t})' \equiv iiN(0, I_2)$. Both unit root ($a_{ii} = 1, i = 1, 2$) and stationary ($a_{ii} = 0.5, i = 1, 2$) cases were examined. The parameter value was set at $\beta_0 = (1, 0)'$. 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}. The number of replications was 5,000.

Figures 1 and 2 show kernel estimates of the sampling distributions of the (correctly specified) logit and probit estimates of the coefficients β_1^0 and β_2^0 in

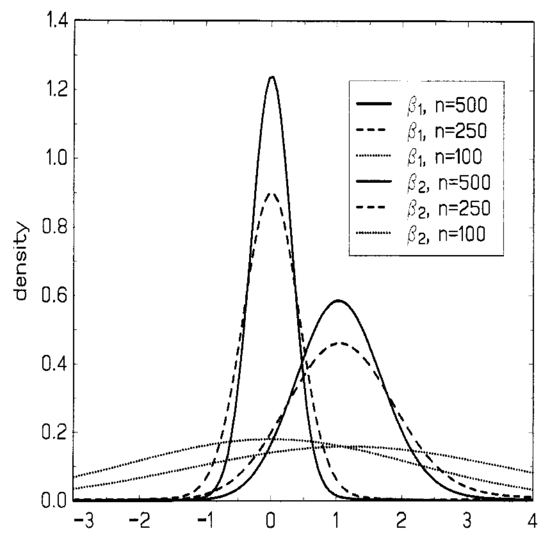

FIGURE 2.—Probit model—densities of estimators of $\beta_1^0 = 1$, $\beta_2^0 = 0$; $I(1)$ case.

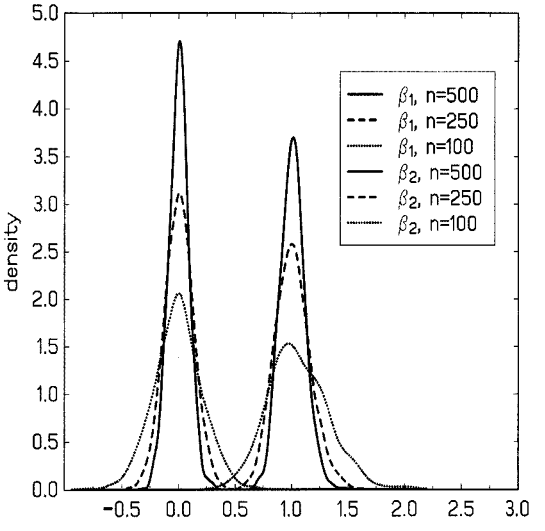

FIGURE 3.—Logit model—densities of estimators of $\beta_1^0 = 1$, $\beta_2^0 = 0$; stationary case.

the unit root case $(a_{ii} = 1, i = 1, 2)$ for sample sizes $n = 100$, 250, 500. The greater concentration of the estimates of β_2^0 and the differing rates of convergence between the two coefficients are quite apparent in the figures. Comparing the probit results (in Figure 2) with those of the logit (Figure 1), reveals the effect of the longer tails of the logit distribution—the probit estimates have substantially greater dispersion for both coefficients. The reason is simply that the probit function attenuates the signal from the nonstationary regressors more severely than the logit function because of the thinner tails of the normal density. Interestingly, this relationship between the logit and probit estimates in a nonstationary regression (viz. that probit estimates are more dispersed than logit estimates) is the opposite of the well known (approximate) scaling relationship that applies in the standard case, viz. that the logit estimates tend to be larger than the probit estimates by a factor of close to $\pi/\sqrt{3}$, a feature that arises from the variance of the logit distribution being $\pi^2/3$, compared with that of the normal being unity (c.f. Greene (1997)).

Figures 3 and 4 show the corresponding estimates for the stationary case $(a_{ii} = 0.5, i = 1, 2)$. There is much less difference between the densities for the two coefficients in this case, just as asymptotic theory for the stationary case indicates. However, estimates of β_2^0 are still somewhat more concentrated than those of estimates of β_1^0. Interestingly, there is a small but noticeable difference between the probit and logit densities, with the probit estimates now being more concentrated, as theory suggests for the stationary case, which is quite the opposite of the relative behavior in the nonstationary case.

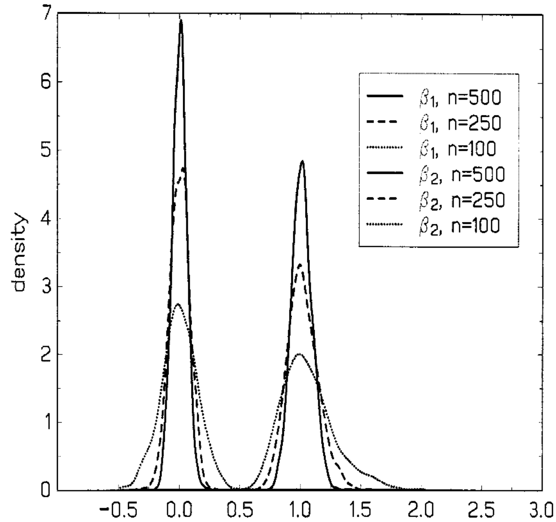

FIGURE 4.—Probit model—densities of estimators of $\beta_1^0 = 1$, $\beta_2^0 = 0$; stationary case.

5. CONCLUSION

While binary choice models have been a popular tool of microeconometrics for several decades, time series and panel data applications of these models are also important and may well increasingly be so in the future. The present paper provides an asymptotic theory for maximum likelihood estimation of these models in time series contexts. Our principal finding is that dual rates of convergence operate in these models when there are multiple integrated regressors, even though the regressors have the same (full rank) stochastic order. This outcome is unusual and is the first instance of the phenomena in asymptotic statistical theory that the authors have encountered. As we have seen, the probability function formulation of conventional binary choice models serves to attenuate the signal emanating from an integrated regressor by dint of the fact that large values of $x_t' \beta_0$ contribute little to the score and hessian because they arise in these quantities by way of a probability density that vanishes at infinity. The effect is that significant signal attenuation occurs in all directions except those that are orthogonal to β_0. It is further shown that the limit distribution theory of the ML estimator is mixed normal and that conventional methods of inference remain valid.

Another finding of some interest is that the sample proportion of positive (or negative) choices follows an arc sine law and therefore spends most of its time in the neighborhood of zero or unity, just as a random walk spends most of its time on one side of the origin or the other. This result has some empirical implications for policy decision making, and in particular market interventions. Thus, in a situation where any of the determining factors involves a stochastic trend, market interventions will most likely occur in streams of little intervention or large numbers of interventions. This is obviously a testable empirical implication of the theory that we hope to explore in later work.

There is scope for extending the results of the paper to some related models. A partial list of extensions that seem valuable for empirical work includes multivariate models, poly-chotomous choice, panel data situations and nonparametric approaches, thereby covering the common extensions of binary choice that arise in the microeconometric context. As indicated earlier, it is also possible to extend our theory to include cointegrated regressors (or combinations of integrated and stationary regressors). In fact, the outcomes in this case are anticipated in results (16) and (17) and so they have not been detailed here as they involve little that is new beyond the results already provided.

School of Economics, Seoul National University, Seoul 151-742, Korea; jpark@plaza.snu.ac.kr; http://econ.snu.ac.kr/faculty/jpark/index_e.html

and

Cowles Foundation for Research in Economics, Yale University, Box 208281, New Haven, CT 06520-8281, U.S.A.; peter.phillips@yale.edu; http://korora.econ.yale.edu

Manuscript received October, 1998; final revision received April, 1999.

APPENDIX A: USEFUL LEMMAS AND PROOFS

LEMMA A1: *Let $f:\mathbf{R}\to\mathbf{R}$ and $f\in\mathbf{F}_I$.*

(a) *Let $\varepsilon>0$ and define*

$$\bar f_\varepsilon(x)=\sup_{|y|\le\varepsilon}|f(x+y)|.$$

Then $\bar f_\varepsilon\in\mathbf{F}_I$.

(b) *Let $K\subset\mathbf{R}$ be compact, and define*

$$\bar f_K(x)=\sup_{c\in K}|f(cx)|.$$

Then $\bar f_K\in\mathbf{F}_I$.

PROOF OF LEMMA A1: It is obvious that both $\bar f_\varepsilon$ and $\bar f_K$ defined in (a) and (b) are bounded. It therefore suffices to show that they are integrable. To simplify the proofs, we assume that f is a symmetric and integrable function that is monotone increasing (decreasing) on $\mathbf{R}_-(\mathbf{R}_+)$. This causes no loss in generality, since any integrable function is bounded by such a function. To deduce part (a), we only need to note that $\bar f_\varepsilon(x)=f(0)$ for $|x|\le\varepsilon$, and $\bar f_\varepsilon(x)=f(x-\varepsilon)$ and $f(x+\varepsilon)$ respectively for $x>\varepsilon$ and $x<-\varepsilon$. Clearly, $\bar f_\varepsilon$ is integrable if f is. To prove part (b), we choose an arbitrary $c_0\in K$, $c_0>0$, and its neighborhood $N_0=[c_0-\varepsilon,c_0+\varepsilon]$ for some $\varepsilon>0$. Define $\bar f_0(x)=f((c_0-\varepsilon)x)$ and $f((c_0+\varepsilon)x)$ respectively for $x\ge0$ and $x<0$. By construction, we then have $f(cx)\le\bar f_0(x)$ for all $c\in N_0$ and $x\in\mathbf{R}$. It is obvious that $\bar f_0$ is integrable, and the stated result follows immediately from the compactness of K. *Q.E.D.*

LEMMA A2: *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*

$$M_n^{\kappa}=\sum_{t=1}^{n} f(x_{1t})x_{2t}^{\kappa} \qquad and \qquad N_n^{\kappa}=\sum_{t=1}^{n} f(x_{1t})x_{2t}^{\kappa}u_t.$$

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

PROOF OF LEMMA A2: Let $V=(V_1,V_2')'$. Note that

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

for all large n. For $f\in\mathbf{F}_0$, we have $n^{-1}\sum_{t=1}^{n}|f(x_{1t})|\to_p 0$, as shown in Park and Phillips (1999). If $f\in\mathbf{F}_I$, it follows from Lemma 2 that $n^{-1/2}\sum_{t=1}^{n}|f(x_{1t})|=O_p(1)$, since f is bounded by a regular function. The stated results in part (a) therefore may easily be deduced. To prove part (b), we notice that

$$n^{-(1+\kappa)}\mathbf{E}\|N_n^{\kappa}\|^2=\mathbf{E}\left(\frac{1}{n^{1+\kappa}}\sum_{t=1}^{n}(\sigma^2f^2)(x_{1t})\|x_{2t}\|^{2\kappa}\right)$$
$$\le \mathbf{E}\left(\left(\sup_{0\le r\le 1}\|V_2(r)\|^{2\kappa}+1\right)\int_0^1(\sigma^2f^2)(\sqrt{n}\,V_{1n}(r))\,dr\right)\to_p 0,$$

by dominated convergence. *Q.E.D.*

LEMMA A3: *Let Assumption 1 hold. Assume* σ^2f^2, $\sigma^2g^2\in\mathbf{F}_R$ *and* σ^2f^4, $\sigma^2g^4\in\mathbf{F}_0$ *for* $f,g:\mathbf{R}\to\mathbf{R}$. *Define*

$${}_1P_{nt}^2=n^{-1/2}f^2(x_{1t})u_t^2 \qquad and \qquad {}_2P_{nt}^2=n^{-3/2}g^2(x_{1t})x_{2t}x_{2t}'u_t^2,$$

and ${}_iQ_{nt}^2=\mathbf{E}({}_iP_{nt}^2\,|\,F_{t-1})$ *for* $i=1,2$. *Then we have for* $i=1,2$

$$\sup_{1\le t\le n}\left\|\sum_{s=1}^{t}{}_iP_{ns}^2-\sum_{s=1}^{t}{}_iQ_{ns}^2\right\|\to_p 0,$$

as $n\to\infty$.

PROOF OF LEMMA A3: Notice first that

$${}_1Q_{nt}^2=n^{-1/2}(\sigma^2f^2)(x_{1t}) \qquad \text{and} \qquad {}_2Q_{nt}^2=n^{-3/2}(\sigma^2g^2)(x_{1t})x_{2t}x_{2t}'.$$

It follows from Lemma 2 that

$$\sum_{t=1}^{n}{}_1Q_{nt}^2\to_d L_1(1,0)\int_{-\infty}^{\infty}(\sigma^2f^2)(s)\,ds,$$
$$\sum_{t=1}^{n}{}_2Q_{nt}^2\to_d \int_0^1 V_2(r)V_2(r)'\,dL_1(r,0)\int_{-\infty}^{\infty}(\sigma^2g^2)(s)\,ds,$$

and therefore, both are obviously tight.

Now we show that

$$\sum_{t=1}^{n}\mathbf{E}({}_iP_{nt}^4\,|\,F_{t-1})\to_p 0,$$

for $i=1,2$. We may assume w.l.o.g. that x_{2t} is scalar by considering each component separately. It follows that

$$\mathbf{E}(u_t^4 \mid F_{t-1}) = \sigma_t^2(1-3F_t+3F_t^2),$$

where we write $F_t = F(x_{1t})$ and $\sigma_t^2 = F_t(1-F_t)$ for short. We therefore have

$$\sum_{t=1}^{n} \mathbf{E}({}_1P_{nt}^4 \mid F_{t-1}) = \frac{1}{n}\sum_{t=1}^{n}(\sigma^2 f^4)(1-3F+3F^2)(x_{1t}) \to_p 0,$$

$$\sum_{t=1}^{n} \mathbf{E}({}_2P_{nt}^4 \mid F_{t-1}) = \frac{1}{n^3}\sum_{t=1}^{n}(\sigma^2 g^4)(1-3F+3F^2)(x_{1t})x_{2t}^4 \to_p 0,$$

due to Lemma A2(a). The stated result now follows from Theorem 2.23 of Hall and Heyde (1980). *Q.E.D.*

LEMMA A4: *Let Assumption* 1 *hold. Assume* $\sigma^2 f$, $\sigma^2 g \in \mathbf{F}_I$ *and* $\sigma^2 f^2$, $\sigma^2 g^2 \in \mathbf{F}_0$ *for* $f, g: \mathbf{R} \to \mathbf{R}$. *Define*

$${}_1N_{nt}^2 = n^{-3/4} f(x_{1t})u_t^2 \qquad \textit{and} \qquad {}_2N_{nt}^2 = n^{-5/4} g(x_{1t})x_{2t}u_t^2.$$

Then we have for $i=1,2$

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

as $n\to\infty$.

PROOF OF LEMMA A4: We let ${}_iM_{nt}^2 = \mathbf{E}({}_iN_{nt}^2 \mid F_{t-1})$ for $i=1,2$, so that

$${}_1M_{nt}^2 = n^{-3/4}(\sigma^2 f)(x_{1t}) \qquad \text{and} \qquad {}_2M_{nt}^2 = n^{-5/4}(\sigma^2 g)(x_{1t})x_{2t}.$$

It follows from Lemma A2 that $\sum_{t=1}^{n} {}_iM_{nt}^2 \to_p 0$ for $i=1,2$. Moreover,

$$\sum_{t=1}^{n} \mathbf{E}({}_1N_{nt}^4 \mid F_{t-1}) = n^{-3/2}\sum_{t=1}^{n}(\sigma^2 f^2)(1-3F+3F^2)(x_{1t}) \to_p 0,$$

$$\sum_{t=1}^{n} \mathbf{E}({}_2N_{nt}^4 \mid F_{t-1}) = n^{-5/2}\sum_{t=1}^{n}(\sigma^2 g^2)(1-3F+3F^2)(x_{1t})x_{2t}^2 \to_p 0,$$

where we assume that x_{2t} is scalar, as in the proof of Lemma A3. The stated result therefore follows as in Lemma A3. *Q.E.D.*

LEMMA A5: *Let Assumption* 1 *hold. Define*

$$\iota_{nk}(r) = 1\{k\delta_n \le \sqrt{n}\,V_{1n}(r) < (k+1)\delta_n\},$$

$$\iota_n(r) = 1\{0 \le \sqrt{n}\,V_{1n}(r) < \delta_n\},$$

$$\iota^n(r) = 1\{0 \le \sqrt{n}\,V_1(r) < \delta_n\}.$$

Then we have

$$\mathbf{E}\left(\int_0^1 |\iota_{nk}(r) - \iota_n(r)|\,dr\right)^2 \le \frac{c\delta_n}{n^{3/2}}(1 + k\delta_n^2 \log n)$$

for all large n, where c is some constant. Moreover,

$$\int_0^1 |\iota_n(r) - \iota^n(r)|\, dr = o_p(n^{-5/8})$$

for any $\delta_n \geq n^{-1/3}$.

PROOF OF LEMMA A5: The stated result follows from Akonom (1993), precisely as for Lemma 2.5 of Park and Phillips (1999). We may just apply his results to $\int_0^1 |\iota_{nk}(r) - \iota_n(r)|\, dr$ and $\int_0^1 |\iota_n(r) - \iota^n(r)|\, dr$, instead of $\int_0^1 (\iota_{nk}(r) - \iota_n(r))\, dr$ and $\int_0^1 (\iota_n(r) - \iota^n(r))\, dr$. Though not stated explicitly, it is obvious that all his results are applicable for $\int_0^1 |\iota_{nk}(r) - \iota_n(r)|\, dr$ and $\int_0^1 |\iota_n(r) - \iota^n(r)|\, dr$ as well as $\int_0^1 (\iota_{nk}(r) - \iota_n(r))\, dr$ and $\int_0^1 (\iota_n(r) - \iota^n(r))\, dr$. *Q.E.D.*

APPENDIX B: PROOFS OF THE MAIN THEOREMS

PROOF OF LEMMA 1: We show how to construct sequences (U_{nt}) and (V_{nt}) satisfying conditions (a)–(c). In the subsequent construction, let $(\Omega, F, \mathbf{P})$ be any probability space rich enough to support the Brownian motions U and V in addition to the other random elements that it includes. Also, define the filtrations

$$G_t = \sigma(u_t, F_{t-1}) \qquad \text{and} \qquad G_{nt} = \sigma(U_{nt}, F_{n,t-1}),$$

and denote by $\cdot \mid F_0$ the distribution conditional on a sub-σ-field F_0.

Let n be given and fixed. First, let V_{n1} be any random variable on $(\Omega, F, \mathbf{P})$, which has the same distribution as $n^{-1/2} v_1$, i.e., $V_{n1} =_d n^{-1/2} v_1$. Second, let τ_{n1} be a stopping time defined on $(\Omega, F, \mathbf{P})$, for which $U(\tau_{n1}/n) \mid F_{n0} =_d n^{-1/2} u_1$. Such a stopping time exists, as shown in Hall and Heyde (1980, Theorem A1). We then define a random variable on $(\Omega, F, \mathbf{P})$, denoted by V_{n2}, such that $V_{n2} - V_{n1} \mid G_{n1} =_d n^{-1/2} v_2 \mid G_1$, and so on. It is obvious that we may proceed in this way to find $(T_{nt})_{t=1}^n$ and $(V_{nt})_{t=1}^n$ on $(\Omega, F, \mathbf{P})$ successively so that

$$U\left(\frac{T_{nt}}{n}\right) \Bigg| F_{n,t-1} =_d \frac{1}{\sqrt{n}} \sum_{i=1}^t u_i \Bigg| F_{t-1} \qquad \text{and} \qquad V_{nt} \Bigg| G_{nt} =_d \frac{1}{\sqrt{n}} \sum_{i=1}^t v_i \Bigg| G_t,$$

in a zig-zag fashion. The equalities of the distributions in (a) and the representation of U_{nt} in (b) are fulfilled by construction. The moment conditions for the stopping times τ_{nt} follow from Hall and Heyde (1980, Theorem A1). Moreover, if we define V_n as in (c), then an invariance principle holds for V_n, as in Phillips and Solo (1992). In particular, it follows that V_n converges weakly to V in $D[0,1]^m$ endowed with uniform topology (see, e.g., Billingsley (1968, pp. 150–153) for the uniform topology in $D[0,1]$). We may therefore redefine V_n so that the distribution of (U_{nt}, V_{nt}) is unchanged and $V_n \to_{a.s.} V$ uniformly on $[0,1]$. This is possible due to the representation theorem in, e.g., Pollard (1984, pp. 71–72) of weakly convergent probability measures by the almost sure convergent sequences. *Q.E.D.*

PROOF OF LEMMA 2: Part (a) follows directly from Park and Phillips (1999, Theorem 5.1). To prove part (b), we let $V_n = (V_{1n}, V_{2n}')'$, where V_n is given in Lemma 1. Define

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

where κ_n and δ_n are sequences of numbers satisfying conditions in the proof of Theorem 5.1 in Park and Phillips (1999). In particular, $\kappa_n \to \infty$ and $\delta_n \to 0$. We may show using the notation in Lemma A5 that

$$\begin{aligned}
(28) \qquad \sqrt{n} \int_0^1 f(\sqrt{n} V_{1n}(r)) V_{2n}(r)\, dr &= \sqrt{n} \int_0^1 f_n(\sqrt{n} V_{1n}(r)) V_{2n}(r)\, dr + o_p(1) \\
&= \sqrt{n} \sum_{k=-\kappa_n}^{\kappa_n} f(k\delta_n) \int_0^1 \iota_{nk}(r) V_{2n}(r)\, dr + o_p(1) \\
&= \left(\int_{-\infty}^{\infty} f_n(s)\, ds \right) \frac{\sqrt{n}}{\delta_n} \int_0^1 \iota_n(r) V_{2n}(r)\, dr + o_p(1),
\end{aligned}$$

following the proof of Theorem 5.1, Park and Phillips (1999). Note that $V_{2n} \to_{a.s.} V_2$ uniformly and therefore, V_{2n} is bounded uniformly in n. Also, we have

$$\left| \int_0^1 \iota_{nk}(r) V_{2n}(r)\, dr - \int_0^1 \iota_n(r) V_{2n}(r)\, dr \right| \le \int_0^1 |\iota_{nk}(r) - \iota_n(r)|\, |V_{2n}(r)|\, dr$$

$$\le \left(1 + \sup_{0 \le r \le 1} |V_2(r)|\right) \int_0^1 |\iota_{nk}(r) - \iota_n(r)|\, dr.$$

The stated result therefore follows from Lemma A5 exactly as in the proof of Theorem 5.1, Park and Phillips (1999).

We have

$$\int_{-\infty}^{\infty} f_n(s)\, ds \to \int_{-\infty}^{\infty} f(s)\, ds. \tag{29}$$

Moreover, if we choose $\delta_n = n^{-\delta}$ with $0 < \delta < 1/8$ and let $\pi_n = \delta_n / \sqrt{n}$, then $n^{5/8}\pi_n \to \infty$. We therefore have from Lemma A5 that

$$\frac{1}{\pi_n} \int_0^1 \iota_n(r) V_{2n}(r)\, dr = \frac{1}{\pi_n} \int_0^1 \iota^n(r) V_2(r)\, dr + o_p(1) \tag{30}$$

since

$$\left| \int_0^1 \iota_n(r) V_{2n}(r)\, dr - \int_0^1 \iota^n(r) V_2(r)\, dr \right|$$

$$\le \int_0^1 |\iota_n(r) - \iota^n(r)|\, |V_{2n}(r)|\, dr + \int_0^1 |\iota^n(r)|\, |V_{2n}(r) - V_2(r)|\, dr$$

$$\le \left(1 + \sup_{0 \le r \le 1} |V_2(r)|\right) \int_0^1 |\iota_n(r) - \iota^n(r)|\, dr + \sup_{0 \le r \le 1} |V_{2n}(r) - V_2(r)|.$$

We may write

$$\frac{1}{\pi_n} \int_0^1 \iota^n(r) V_2(r)\, dr = \int_0^1 \int_0^1 V_2(r)\, dL_1(r, \pi_n s)\, ds \tag{31}$$

using the extended occupation times formula (see, e.g., Revuz and Yor (1994, Exercise 1.15, p. 222)).

Define

$$R_n = \int_0^1 \int_0^1 V_2(r)\, dL_1(r, \pi_n s)\, ds - \int_0^1 V_2(r)\, dL_1(r, 0). \tag{32}$$

Due to (28)–(31), it suffices to show that $R_n \to_{a.s.} 0$, which we now set out to do. Let c_n be a sequence of numbers such that

$$c_n \to \infty \qquad \text{and} \qquad c_n \pi_n^{1/2-\varepsilon} \to 0,$$

for some $0 < \varepsilon < 1/2$. Define

$$I(x) = \int_0^1 V_2(r)\, dL_1(r, x),$$

$$I_n(x) = \sum_{i=1}^{c_n} V_2\left(\frac{i}{c_n}\right)\left(L_1\left(\frac{i+1}{c_n}, x\right) - L_1\left(\frac{i}{c_n}, x\right)\right).$$

Then we have

$$|R_n| \le A_n + B_n + C_n,$$

where

$$A_n = \int_0^1 |I(\pi_n s) - I_n(\pi_n s)|\, ds,$$

$$B_n = \int_0^1 |I_n(\pi_n s) - I_n(0)|\, ds,$$

$$C_n = \int_0^1 |I_n(0) - I(0)|\, ds.$$

We will show that $A_n,\ B_n,\ C_n \to_{a.s.} 0$.

It is obvious that $C_n \to_{a.s.} 0$ since $I(0)$ exists. To show $B_n \to_{a.s.} 0$, notice that we have

$$\sup_{r\in[0,1]} |L_1(r, \pi_n s) - L_1(r,0)| \le \kappa_1 |\pi_n s|^{1/2-\varepsilon} \quad \text{a.s.,} \tag{33}$$

for some constant κ_1, due to the uniform Hölder continuity of the local time $L_1(r,\cdot)$ (see, e.g., Revuz and Yor (1994, Corollary 1.8, p. 217)). Therefore,

$$|I_n(\pi_n s) - I_n(0)| \le 2\kappa_1 \left(\frac{1}{c_n} \sum_{i=1}^{c_n} \left| V_2\left(\frac{i}{c_n}\right)\right| \right) c_n |\pi_n s|^{1/2-\varepsilon},$$

and consequently,

$$B_n \le 2\kappa_1 c_n \pi_n^{1/2-\varepsilon} \left(\frac{1}{c_n} \sum_{i=1}^{c_n} \left| V_2\left(\frac{i}{c_n}\right)\right| \right) \int_0^1 |s|^{1/2-\varepsilon}\, ds \to_{a.s.} 0,$$

as was to be shown.

Finally, we let

$$V_2^n(r) = \sum_{i=1}^{c_n} V_2\left(\frac{i}{c_n}\right) 1\left\{ \frac{i}{c_n} \le r < \frac{i+1}{c_n} \right\},$$

so that

$$I_n(x) = \int_0^1 V_2^n(r)\, dL_1(r, x).$$

As is well known

$$\sup_{0\le r\le 1} |V_2(r) - V_2^n(r)| \le \kappa_2 c_n^{-1/2+\varepsilon} \quad \text{a.s.} \tag{34}$$

It follows from (33) and (34) that

$$\begin{aligned} |I(\pi_n s) - I_n(\pi_n s)| &\le \kappa_2 c_n^{-1/2+\varepsilon} \int_0^1 |dL_1(r, \pi_n s)| \\ &\le \kappa_2 c_n^{-1/2+\varepsilon} \left(\int_0^1 |dL_1(r,0)| + 2\kappa_1 |\pi_n s|^{1/2-\varepsilon} \right), \end{aligned}$$

and therefore,

$$A_n \le \kappa_2 c_n^{-1/2+\varepsilon} \int_0^1 |dL_1(r,0)| + 2\kappa_1\kappa_2 c_n^{-1/2+\varepsilon} \pi_n^{1/2-\varepsilon} \int_0^1 |s|^{1/2-\varepsilon}\, ds \to_{a.s.} 0.$$

The proof for part (b) is thereby complete.

The proof for part (c) is analogous to that of part (b). We have

$$\begin{aligned}
&\frac{1}{n^{3/2}}\sum_{t=1}^{n} f(x_{1t})x_{2t}x_{2t}' \\
&\quad =_d \sqrt{n}\int_0^1 f(\sqrt{n}\,V_{1n}(r))V_{2n}(r)V_{2n}(r)'\,dr \\
&\quad = \sqrt{n}\int_0^1 f_n(\sqrt{n}\,V_{1n}(r))V_{2n}(r)V_{2n}(r)'\,dr + o_p(1) \\
&\quad = \sqrt{n}\sum_{k=-\kappa_n}^{\kappa_n} f(k\delta_n)\int_0^1 1\{k\delta_n \le \sqrt{n}\,V_{1n}(r) \le (k+1)\delta_n\}V_{2n}(r)V_{2n}(r)'\,dr + o_p(1) \\
&\quad = \left(\delta_n \sum_{k=-\kappa_n}^{\kappa_n} f(k\delta_n)\right)\frac{\sqrt{n}}{\delta_n}\int_0^1 1\{0 \le \sqrt{n}\,V_{1n}(r) < \delta_n\}V_{2n}(r)V_{2n}(r)'\,dr + o_p(1) \\
&\quad = \left(\int_{-\infty}^{\infty} f(s)\,ds\right)\frac{1}{\pi_n}\int_0^1 1\{0 \le V_1(r) < \pi_n\}V_2(r)V_2(r)'\,dr + o_p(1) \\
&\quad = \left(\int_{-\infty}^{\infty} f(s)\,ds\right)\int_0^1\int_0^1 V_2(r)V_2(r)'\,dL_1(r,\pi_n s)\,ds + o_p(1) \\
&\quad = \left(\int_{-\infty}^{\infty} f(s)\,ds\right)\int_0^1 V_2(r)V_2(r)'\,dL_1(r,0) + o_p(1).
\end{aligned}$$

Each step can be shown rigorously using the arguments in the proof of part (b). *Q.E.D.*

PROOF OF LEMMA 3: We set $m=2$. This is just for notational simplicity. The proof for the general case is essentially identical. For any $c=(c_1,c_2)\in \mathbf{R}^2$, we let

$$A_n(x_1,x_2) = c_1 n^{-1/4} f(x_1) + c_2 n^{-3/4} f(x_1)x_2,$$

and define

$$M_n(r) = \sqrt{n}\sum_{i=1}^{t-1} A_n(\sqrt{n}\,V_{ni})\left(U\left(\frac{T_{ni}}{n}\right) - U\left(\frac{T_{n,i-1}}{n}\right)\right) + \sqrt{n}\,A_n(\sqrt{n}\,V_{nt})\left(U(r) - U\left(\frac{T_{n,t-1}}{n}\right)\right), \tag{35}$$

for $T_{n,t-1}/n < r \le T_{nt}/n$, where T_{nt}, $t=1,\ldots,n$, are the time changes introduced in Lemma 1. One may easily see that M_n is a continuous martingale such that

$$\sum_{t=1}^{n} A_n(x_{1t},x_{2t})u_t =_d M_n\left(\frac{T_{nn}}{n}\right), \tag{36}$$

for all n.

Let $B_n(x_1,x_2) = \sigma^2(x_1)A_n^2(x_1,x_2)$. The quadratic variation $[M_n]$ of M_n is given by

$$\begin{aligned}
[M_n](r) &= n\sum_{i=1}^{t-1} A_n^2(\sqrt{n}\,V_{ni})\left(\frac{T_{ni}}{n} - \frac{T_{n,i-1}}{n}\right) + nA_n^2(\sqrt{n}\,V_{nt})\left(r - \frac{T_{n,t-1}}{n}\right) \\
&= \sum_{t=1}^{n} B_n(\sqrt{n}\,V_{nt})1\left\{r \ge \frac{T_{nt}}{n}\right\} + o_p(1),
\end{aligned}$$

uniformly in $r \in [0,1]$, by Lemma A3. Consequently, we have

(37) $$[M_n](r) \to_p c' M(r) c,$$

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

$$M(r) = \begin{pmatrix} L_1(r,0)\int_{-\infty}^{\infty} f_\sigma^2(s)\,ds & \int_0^r dL_1(s,0)V_2(s)'\int_{-\infty}^{\infty} f_\sigma(s) g_\sigma(s)\,ds \\ \int_0^r V_2(s)\,dL_1(s,0)\int_{-\infty}^{\infty} g_\sigma(s) f_\sigma(s)\,ds & \int_0^r V_2(s)V_2(s)'\,dL_1(s,0)\int_{-\infty}^{\infty} g_\sigma^2(s)\,ds \end{pmatrix},$$

due to the results in Lemma 2, and where the notation $f_\sigma(s) = \sigma(s) f(s)$ and $g_\sigma(s) = \sigma(s) g(s)$ is being used.

Moreover, if we let σ_{uv} be the covariance of U and V and

$$C_n(x_1, x_2) = \sigma^2(x_1) A_n(x_1, x_2),$$

then the quadratic covariation process $[M_n, V]$ of M_n and V is

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

uniformly in $r \in [0,1]$, by Lemma A4. It follows, in particular, that

(38) $$[M_n, V](\rho_n(r)) \to_p 0,$$

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

The asymptotic distribution of the continuous martingale M_n in (35) is completely determined by (37) and (38), as shown in Revuz and Yor (1994, Theorem 2.3, page 496). Now define

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

The process W_n is the DDS (or Dambis, Dubins-Schwarz) Brownian motion of the continuous martingale M_n (see, for example, Revuz and Yor (1994, Theorem 1.6, page 173)). It now follows that (V, W_n) converges jointly in distribution to two independent standard linear Brownian motions (V, W), say. Therefore,

$$M_n\left(\frac{T_{nn}}{n} \right) \to_d W(c' M c),$$

which, due to (36), completes the proof for the first part. *Q.E.D.*

PROOF OF THEOREM 1: The limiting distribution of $D_n^{-1} S_n(\alpha_0)$ is derived by applying Lemma 3 with $f(x) = xG(x)$ and $g(x) = G(x)$. It is easy to check that the conditions in Lemma 3 hold. Note that $G^2 F(1-F) = K$, for which $K_2 \in \mathbf{F}_R$ by Assumption 2(a). Also, we have $GF(1-F) = \dot{F}$ and $\dot{F}_1 \in \mathbf{F}_1$ by Assumption 2(b). Moreover, $G^4 F(1-F) = G^3 \dot{F}$ and we require $G^3 \dot{F}_4 \in \mathbf{F}_0$ in Assumption 2(c). Thus, with $\alpha_1^0 = (\beta_0' \beta_0)^{1/2}$ and using Lemma 3, we obtain

$$\begin{aligned} D_n^{-1} S_n(\alpha_0) &= D_n^{-1} \sum_{t=1}^{n} G(x_t' \beta_0) H' x_t (y_t - F(x_t' \beta_0)) \\ &= \begin{pmatrix} n^{-1/4} \sum_{t=1}^n G(x_{1t}\alpha_1^0) x_{1t} u_t \\ n^{-3/4} \sum_{t=1}^n G(x_{1t}\alpha_1^0) x_{2t} u_t \end{pmatrix} \\ &= \begin{pmatrix} n^{-1/4} (\alpha_1^0)^{-1} \sum_{t=1}^n f(x_{1t}\alpha_1^0) u_t \\ n^{-3/4} \sum_{t=1}^n g(x_{1t}\alpha_1^0) x_{2t} u_t \end{pmatrix} \\ &\to_d M^{1/2} W(1), \end{aligned}$$

with

$$M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & M_{22} \end{pmatrix}$$

where

$$m_{11} = (\alpha_1^0)^{-2} L_1(1,0) \int_{-\infty}^{\infty} f_\sigma^2(s\alpha_1^0)\, ds,$$

$$m_{12} = (\alpha_1^0)^{-1} \int_0^1 dL_1(r,0) V_2(r)' \int_{-\infty}^{\infty} f_\sigma(s\alpha_1^0) g_\sigma(s\alpha_1^0)\, ds,$$

$$m_{21} = (\alpha_1^0)^{-1} \int_0^1 V_2(r)\, dL_1(r,0) \int_{-\infty}^{\infty} g_\sigma(s\alpha_1^0) f_\sigma(s\alpha_1^0)\, ds,$$

$$M_{22} = \int_0^1 V_2(r) V_2(r)'\, dL_1(r,0) \int_{-\infty}^{\infty} g_\sigma^2(s\alpha_1^0)\, ds,$$

and where

$$f_\sigma(x) = f(x)^2 \sigma(x)^2 = x^2 G(x)^2 F(x)[1 - F(x)] = x^2 K(x),$$

$$g_\sigma(x)^2 = g(x)^2 \sigma(x)^2 = G(x)^2 F(x)[1 - F(x)] = K(x),$$

$$f_\sigma(x) g_\sigma(x) = f(x) g(x) \sigma(x)^2 = xG(x)^2 F(x)[1 - F(x)] = xK(x).$$

Thus,

$$M = \begin{pmatrix} L_1(1,0)\int_{-\infty}^{\infty} s^2 K(\alpha_1^0 s)\, ds & \int_0^1 dL_1(r,0) V_2(r)' \int_{-\infty}^{\infty} sK(\alpha_1^0 s)\, ds \\ \int_0^1 V_2(r)\, dL_1(r,0) \int_{-\infty}^{\infty} sK(\alpha_1^0 s)\, ds & \int_0^1 V_2(r) V_2(r)'\, dL_1(r,0) \int_{-\infty}^{\infty} K(\alpha_1^0 s)\, ds \end{pmatrix},$$

and, with $M = Q$, we have the stated result for the limit of the score process $D_n^{-1} S_n(\alpha_0)$.

To get the stated asymptotic result for $J_n(\alpha_0)$, we let $\dot{G}_t = \dot{G}(x_{1t})$ and observe that

$$n^{-1/2} \sum_{t=1}^{n} \dot{G}_t x_{1t}^2 u_t,\ n^{-1} \sum_{t=1}^{n} \dot{G}_t x_{1t} x_{2t} u_t, \quad n^{-3/2} \sum_{t=1}^{n} \dot{G}_t x_{2t} x_{2t}' u_t = o_p(1).$$

These follow from a direct application of Lemma A2(b) with $f(x) = \dot{G}(x)$, $x\dot{G}(x)_i$, $x^2 \dot{G}(x)$ and using the fact that $[\dot{G}F^{1/2}(1-F)^{1/2}]_2 \in \mathbf{F}_0$ by Assumption 2(c). We therefore have

$$\begin{aligned} & D_n^{-1} J_n(\alpha_0) D_n(\alpha_0) \\ & = -D_n^{-1} \sum_{t=1}^{n} K(x_{1t}\alpha_1^0) H' x_t x_t' H D_n^{-1} + o_p(1) \\ & = \begin{pmatrix} n^{-1/2}\sum_{t=1}^n K(x_{1t}\alpha_1^0) x_{1t}^2 & n^{-1}\sum_{t=1}^n K(x_{1t}\alpha_1^0) x_{1t} x_{2t}' \\ n^{-1}\sum_{t=1}^n K(x_{1t}\alpha_1^0) x_{2t} x_{1t} & n^{-3/2}\sum_{t=1}^n K(x_{1t}\alpha_1^0) x_{2t} x_{2t}' \end{pmatrix} + o_p(1). \end{aligned} \tag{39}$$

The stated results are now immediate from Lemma 2. *Q.E.D.*

PROOF OF THEOREM 2: We employ a standard approach to local extremum estimation. In particular, Theorem 10.1 of Wooldridge (1994) allows for varying rates of convergence in the components of the estimator and is well suited to the present problem. The idea is simply to show that equation (19) holds and that there is a consistent local solution to the likelihood equation. Conditions (i) and (ii) of Wooldridge's theorem hold trivially by our assumption about β_0 and our Assumption 2, so it remains to verify conditions (iii) and (iv) of that theorem.

The likelihood equation for the ML estimator $\hat{\alpha}_n$ is

$$S_n(\hat{\alpha}_n) = 0, \tag{40}$$

which has the expansion

$$S_n(\hat{\alpha}_n) = S_n(\alpha_0) + J_n(\alpha_n)(\hat{\alpha}_n - \alpha_0) = 0,$$

or

$$S_n(\alpha_0) + J_n(\alpha_0)(\hat{\alpha}_n - \alpha_0) + [J_n(\alpha_n) - J_n(\alpha_0)](\hat{\alpha}_n - \alpha_0) = 0, \tag{41}$$

where $S_n(\hat{\alpha}_n)$ and $S_n(\alpha_0)$ are the scores respectively at $\hat{\alpha}_n$ and α_0, and $J_n(\alpha_n)$ is the hessian matrix with rows evaluated at mean values that lie on the line segment connecting $\hat{\alpha}_n$ and α_0. Then (41) can be written as

$$\begin{aligned} 0 &= D_n^{-1} S_n(\alpha_0) + [D_n^{-1} J_n(\alpha_0) D_n^{-1}] D_n(\hat{\alpha}_n - \alpha_0) \\ &\quad + (D_n^{-1}[J_n(\alpha_n) - J_n(\alpha_0)] D_n^{-1}) D_n(\hat{\alpha}_n - \alpha_0), \end{aligned}$$

or

$$\begin{aligned} 0 &= D_n^{-1} S_n(\alpha_0) + [D_n^{-1} J_n(\alpha_0) D_n^{-1}] D_n(\hat{\alpha}_n - \alpha_0) \\ &\quad + n^{-2\delta}(C_n^{-1}[J_n(\alpha_n) - J_n(\alpha_0)] C_n^{-1}) D_n(\hat{\alpha}_n - \alpha_0), \end{aligned} \tag{42}$$

where $C_n = D_n n^{-\delta}$ for some $\delta > 0$, so that $C_n D_n^{-1} = o(1)$ as $n \to \infty$, as in condition (iii) (a) of Wooldridge's Theorem 10.1. Equation (19) now follows from (42) if the final term of (42) is $o_p(1)$. This will be so, if condition (iii) (b) of Wooldridge's theorem holds.

To show this is so, we need to establish that

$$\sup_{\{\alpha : \|C_n(\alpha - \alpha_0)\| \le 1\}} \|C_n^{-1}[J_n(\alpha) - J_n(\alpha_0)] C_n^{-1}\| = o_p(1). \tag{43}$$

Our proof involves looking at the components of the hessian. We therefore partition the hessian conformably with α as

$$J_n(\alpha) = \begin{pmatrix} J_{11}^n(\alpha) & J_{12}^n(\alpha) \\ J_{21}^n(\alpha) & J_{22}^n(\alpha) \end{pmatrix},$$

wherein, from (12),

$$J_{ij}^n(\alpha) = -\sum_{t=1}^n K_t(\alpha) x_{it} x_{jt} - \sum_{t=1}^n \dot{G}_t(\alpha) x_{it} x_{jt} (F_t(\alpha) - F_t) + \sum_{t=1}^n \dot{G}_t(\alpha) x_{it} x_{jt} u_t, \tag{44}$$

for $i, j = 1, 2$ and where we define $f_t(\alpha) = f(x_{1t}\alpha_1 + x_{2t}'\alpha_2)$ for any function $f : \mathbf{R} \to \mathbf{R}$ and further define f_t to be the value of the function f at $\alpha_0 = (\alpha_1^0, \alpha_2^{0\prime})'$.

For (43) to hold it is sufficient that

$$\begin{aligned} &n^{-1/2+2\delta} |J_{11}^n(\alpha) - J_{11}^n(\alpha_0)| \to_p 0, \\ &n^{-1+2\delta} \|J_{12}^n(\alpha) - J_{12}^n(\alpha_0)\| \to_p 0, \\ &n^{-3/2+2\delta} \|J_{22}^n(\alpha) - J_{22}^n(\alpha_0)\| \to_p 0, \end{aligned} \tag{45}$$

uniformly for all α_1 and α_2 satisfying

$$|\alpha_1 - 1| \le n^{-1/4+\delta} \quad \text{and} \quad \|\alpha_2\| \le n^{-3/4+\delta}, \tag{46}$$

for some $\delta > 0$. Now, from (44) we have

(47) $$J_{ij}^n(\alpha) - J_{ij}^n(\alpha_0) = A_{ij}^n(\alpha^*) + B_{ij}^n(\alpha^*) + C_{ij}^n(\alpha^*),$$

where

$$A_{ij}^n(\alpha) = -\sum_{t=1}^n \dot{K}_t(\alpha) x_{it} x_{jt} (x_{1t}(\alpha_1 - \alpha_1^0) + x_{2t}'(\alpha_2 - \alpha_2^0)),$$

$$B_{ij}^n(\alpha) = -\sum_{t=1}^n (\dot{G}\dot{F})_t(\alpha) x_{it} x_{jt} (x_{1t}(\alpha_1 - \alpha_1^0) + x_{2t}'(\alpha_2 - \alpha_2^0)),$$

$$C_{ij}^n(\alpha) = \sum_{t=1}^n \ddot{G}_t(\alpha) x_{it} x_{jt} u_t (x_{1t}(\alpha_1 - \alpha_1^0) + x_{2t}'(\alpha_2 - \alpha_2^0)),$$

and α^* is on the line segment connecting α and α_0. For $f: \mathbf{R} \to \mathbf{R}$, define $\bar{f}$ by

$$\bar{f}(x) = \sup_{|a-1| \le \varepsilon} \sup_{|b| \le \varepsilon} |f(ax+b)|,$$

for $\varepsilon > 0$ given. We denote $f_t = \bar{f}(x_{1t})$. As shown in Lemma A1, $\bar{f} \in \mathbf{F}_I$ if $f \in \mathbf{F}_I$. Since $\sup_{1 \le t \le n} \|x_{2t}\| / \sqrt{n} = O_p(1)$, $|\alpha_1 - \alpha_1^0| \le n^{-1/4+\delta}$, and $\|\alpha_2\| \le n^{-3/4+\delta}$, we have for any $\varepsilon > 0$

$$|f(x_{1t}\alpha_1 + x_{2t}\alpha_2)| \le \bar{f}(x_{1t}) + o_p(1),$$

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

By virtue of (46) we have

$$\|A_{ij}^n(\alpha)\| \le n^{-1/4+\delta} \sum_{t=1}^n \underline{\dot{K}}_t \|x_{it}\| \|x_{jt}\| |x_{1t}| + n^{-3/4+\delta} \sum_{t=1}^n \underline{\dot{K}}_t \|x_{it}\| \|x_{jt}\| \|x_{2t}\|.$$

It therefore follows from Lemma A2(a) that

(48) $$\|A_{11}^n(\alpha)\| = O_p(n^{1/4+\delta}), \qquad \|A_{12}^n(\alpha)\| = O_p(n^{3/4+\delta}), \qquad \|A_{22}^n(\alpha)\| = O_p(n^{5/4+\delta}),$$

uniformly in α satisfying (46). Note that $\dot{K} = \dot{G}\dot{F} + G\ddot{F}$ and hence $\dot{K}_2 \in \mathbf{F}_I$ by Assumption 2(b). We may use exactly the same argument to deduce that

(49) $$\|B_{11}^n(\alpha)\| = O_p(n^{1/4+\delta}), \qquad \|B_{12}^n(\alpha)\| = O_p(n^{3/4+\delta}), \qquad \|B_{22}^n(\alpha)\| = O_p(n^{5/4+\delta}).$$

Finally, we have

$$\|C_{ij}^n(\alpha)\| \le n^{-1/4+\delta} \sum_{t=1}^n \underline{\ddot{G}}_t (F_t(1-F_t))^{1/2} \|x_{it}\| \|x_{jt}\| |x_{1t}| + n^{-3/4+\delta} \sum_{t=1}^n \underline{\ddot{G}}_t (F_t(1-F_t))^{1/2} \|x_{it}\| \|x_{jt}\| \|x_{2t}\|,$$

since

$$(\mathbf{E}(|u_t| F_{t-1}))^2 \le \mathbf{E}(u_t^2 | F_{t-1}) = F_t(1-F_t),$$

by the conditional Jensen inequality. It follows directly from Assumption 2(b) and Lemma A2(a) that

(50) $$\|C_{11}^n(\alpha)\| = O_p(n^{1/4+\delta}), \qquad \|C_{12}^n(\alpha)\| = O_p(n^{3/4+\delta}), \qquad \|C_{22}^n(\alpha)\| = O_p(n^{5/4+\delta}),$$

uniformly in α given by (46). If we let $0 < \delta < 1/12$, we may now easily deduce (45) from (48), (49), and (50) together with (47). Hence, (43) holds and therefore (19).

It now follows as in the proof of Theorem 10.1 of Wooldridge (1994) that there exists a solution to the likelihood equation (40) with probability approaching one such that

$$D_n(\hat{\alpha}_n - \alpha_0) = O_p(1).$$

From Theorem 1, we have the joint weak convergence

$$(D_n^{-1} S_n(\alpha_0), D_n^{-1} J_n(\alpha_0) D_n^{-1}) \to_d (Q^{1/2} W(1), -Q), \tag{51}$$

where Q is positive definite with probability one. Thus, condition (iv) of Wooldridge's theorem holds. The given limit distribution of $D_n(\hat{\alpha}_n - \alpha_0)$ now follows directly from (42), (43), and (51). *Q.E.D.*

PROOF OF THEOREM 3: Write

$$\begin{aligned} -J_n(\hat{\beta}_n)^{-1} &= -H[H' J_n(\hat{\beta}_n) H]^{-1} H' \\ &= -H[J_n(\hat{\alpha}_n)]^{-1} H' \\ &= -H D_n^{-1} [D_n^{-1} J_n(\hat{\alpha}_n) D_n^{-1}]^{-1} D_n^{-1} H'. \end{aligned}$$

By Theorems 1 and 2 we have

$$-D_n^{-1} J_n(\hat{\alpha}_n) D_n^{-1} = -D_n^{-1} J_n(\alpha_0) D_n^{-1} + o_p(1) \to_d Q.$$

Partitioning the hessian matrix conformably with (22) and using (39), we have

$$\begin{aligned} & D_n^{-1} \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix} D_n^{-1} \\ &\quad = D_n^{-1} \begin{pmatrix} \sum_{t=1}^n K(x_{1t}\alpha_1^0) x_{1t}^2 & \sum_{t=1}^n K(x_{1t}\alpha_1^0) x_{1t} x_{2t}' \\ \sum_{t=1}^n K(x_{1t}\alpha_1^0) x_{2t} x_{1t} & \sum_{t=1}^n K(x_{1t}\alpha_1^0) x_{2t} x_{2t}' \end{pmatrix} D_n^{-1} + o_p(1). \end{aligned} \tag{52}$$

Then

$$\left[\frac{J_n(\hat{\alpha}_n)}{\sqrt{n}} \right]^{-1} = \left[\frac{1}{\sqrt{n}} \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix} \right]^{-1} = \sqrt{n} \begin{pmatrix} J_{11.2}^{-1} & -J_{11.2}^{-1} J_{12} J_{22}^{-1} \\ -J_{22}^{-1} J_{21} J_{11.2}^{-1} & J_{22}^{-1} \end{pmatrix}.$$

Next observe that

$$\begin{aligned} n^{-1/2} J_{11.2} &= \left(n^{-1/2} \sum_{t=1}^n K(x_{1t}\alpha_1^0) x_{1t}^2 + o_p(1) \right) - \left(n^{-1} \sum_{t=1}^n K(x_{1t}\alpha_1^0) x_{1t} x_{2t}' + o_p(1) \right) \\ &\quad \cdot \left(n^{-3/2} \sum_{t=1}^n K(x_{1t}\alpha_1^0) x_{2t} x_{2t}' + o_p(1) \right)^{-1} \left(n^{-1} \sum_{t=1}^n K(x_{1t}\alpha_1^0) x_{2t} x_{1t} + o_p(1) \right) \\ &\to_d q_{11} - q_{12} Q_{22}^{-1} q_{21} = q_{11.2}, \end{aligned}$$

$$\sqrt{n}\, J_{11.2}^{-1} J_{12} J_{22}^{-1} = \frac{1}{\sqrt{n}} (\sqrt{n}\, J_{11.2}^{-1})(n^{-1} J_{12})(n^{-3/2} J_{22})^{-1} = O_p(n^{-1/2}),$$

and

$$\sqrt{n}\, J_{22}^{-1} = \frac{\sqrt{n}}{n^{3/2}} (n^{-3/2} J_{22})^{-1} = O_p(n^{-1}).$$

It follows that

$$\left[\frac{J_n(\hat{\alpha}_n)}{\sqrt{n}}\right]^{-1} \to_d \begin{pmatrix} q_{11.2}^{-1} & 0 \\ 0 & 0 \end{pmatrix},$$

and thus

$$[D_n^{-1} J_n(\hat{\alpha}_n) D_n^{-1}]^{-1} \to_d \begin{pmatrix} q_{11.2}^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

Consequently,

$$\begin{aligned} -\sqrt{n}\, J_n(\hat{\beta}_n)^{-1} &= -HD_n^{-1}\sqrt{n}\,[D_n^{-1} J_n(\hat{\alpha}_n) D_n^{-1}]^{-1} D_n^{-1} H' \\ &\to_d h_1 q_{11.2}^{-1} h_1' = \beta_0(\beta_0'\beta_0)^{-1}\beta_0' q_{11.2}^{-1}, \end{aligned}$$

giving the stated result for $-\sqrt{n}\, J_n(\hat{\beta}_n)^{-1}$. The result for $-\sqrt{n}\, \underline{J}_n(\hat{\beta}_n)^{-1}$ follows from (52) using the same argument. *Q.E.D.*

PROOF OF COROLLARY 2: The stated result follows immediately, since

$$(\hat{\beta}_n - \beta_0) = J_1(\hat{\alpha}_{1n} - \alpha_1^0) + J_2(\hat{\alpha}_{2n} - \alpha_2^0),$$

and

$$\sqrt[4]{n}\,(\hat{\beta}_n - \beta_0) = \sqrt[4]{n}\, J_1(\hat{\alpha}_{1n} - \alpha_1^0) + O_p(n^{-1/2}),$$

for large n. *Q.E.D.*

PROOF OF COROLLARY 3: Use the following mean value expansions for $\hat{F} = F(x'\hat{\beta}_n)$ and $\hat{\gamma}_x = f(x'\hat{\beta}_n)\hat{\beta}_n$

$$\begin{aligned} \hat{F} &= F(x_t'\beta_0) + f(x_t'\beta_n)x_t'(\hat{\beta}_n - \beta_0), \\ \hat{\gamma}_x &= \gamma(x'\beta_0) + \Gamma(x'\beta_n)(\hat{\beta}_n - \beta_0), \end{aligned} \tag{53}$$

where

$$\Gamma(x'\beta_n) = f(x'\beta_n)I_m + f'(x'\beta_n)\beta_n x'.$$

Then

$$\begin{aligned} \sqrt[4]{n}\,(\hat{F} - F(x'\beta_0)) &= f(x'\beta_n)x'(\hat{\beta}_n - \beta_0)n^{1/4} \sim_d f(x'\beta_0)x'\,\mathrm{MN}(0, P_{\beta_0} q_{11.2}^{-1}) \\ &=_d \mathrm{MN}\Big(0, f(x'\beta_0)^2 x' P_{\beta_0} x q_{11.2}^{-1}\Big), \end{aligned} \tag{54}$$

and

$$\begin{aligned} \sqrt[4]{n}\,(\hat{\gamma}_x - \gamma(x'\beta_0)) &= \Gamma(x'\beta_n)(\hat{\beta}_n - \beta_0)n^{1/4} \sim_d \Gamma(x'\beta_0)\mathrm{MN}(0, P_{\beta_0} q_{11.2}^{-1}) \\ &=_d \mathrm{MN}(0, \Gamma(x'\beta_0)P_{\beta_0}\Gamma(x'\beta_0)' q_{11.2}^{-1}). \end{aligned}$$

A simple calculation reveals that

$$\Gamma(x'\beta_0)P_{\beta_0}\Gamma(x'\beta_0)' = (f(x'\beta_0) + f'(x'\beta_0)x'\beta_0)^2 P_{\beta_0},$$

and the stated results follow. *Q.E.D.*

PROOF OF THEOREM 4: Since $y_t = F(x_t'\beta_0) + u_t$ and u_t is a martingale difference, we have

$$r_n = n^{-1}\sum_{t=1}^{n} y_t = n^{-1}\sum_{t=1}^{n} F(x_t'\beta_0) + o_p(1).$$

The function $F(z)$ is a cumulative distribution function and is asymptotically homogeneous of degree zero as $z \to \infty$ in the sense that

$$F(\lambda z) = 1\{z > 0\} + R(z, \lambda),$$

and $R(z, \lambda)$ is dominated by a locally integrable function that vanishes at infinity. It therefore follows from Park and Phillips (1999, Theorem 5.3) that

$$n^{-1} \sum_{t=1}^{n} F(x_t' \beta_0) \to_d \int_{-\infty}^{\infty} 1\{s > 0\} L_1(1, s)\, ds = \int_{0}^{\infty} L_1(1, s)\, ds,$$

as required.

The proof for the predicted proportion $r_n(X) = n^{-1} \sum_{t=1}^{n} y_t(X)$ follows in the same manner. In the estimated case, $\hat{r}_n(X) = n^{-1} \sum_{t=1}^{n} \hat{F}_t(X)$, with $\hat{F}_t(X) = F(X_t' \hat{\beta}_n)$. By the mean value expansion in (54) and using Lemma 2 (a) and (b) we find that

$$\begin{aligned} \hat{r}_n(X) &= r_n(X) + n^{-1} \sum_{t=1}^{n} f(x' \beta_n) x' (\hat{\beta}_n - \beta_0) \\ &= r_n(X) + O_p(n^{-1/4}), \end{aligned}$$

and thus $\hat{r}_n(X)$ has the same limit as $r_n(X)$. *Q.E.D.*

APPENDIX C: NOTATION

$\to_{a.s.}$	almost surely converge.
$\to_p$	convergence in probability.
$\to_d$	weak convergence.
$o_p(1)$	tends to zero in probability.
$o_{a.s.}(1)$	tends to zero almost surely.
$=_d$	distributional equivalence.
$\sim_d$	asymptotically distributed as.
$\equiv$	equivalence.
W, V_1, V_2	standard Brownian motions.
$\mathrm{MN}(0, V)$	mixed normal distribution with variance V.
$\lVert \cdot \rVert$	Euclidean norm in $\mathbf{R}^k$.
$\mathbf{F}_R$	class of regular functions.
$\mathbf{F}_I$	class of bounded integrable functions.
$\mathbf{F}_0$	class of bounded functions vanishing at infinity.

REFERENCES

AKONOM, J. (1993): "Comportement Asymptotique du Temps d'Occupation du Processus des Sommes Partielles," *Annales de l'Institut Henri Poincaré*, 29, 57–81.

AMEMIYA, T. (1985): *Advanced Econometrics*. Cambridge: Harvard University Press.

BILLINGSLEY, P. (1968): *Convergence of Probability Measures*. New York: Wiley.

DHRYMES, P. (1986): "Limited Dependent Variables," Ch. 27 in *Handbook of Econometrics*, Vol. III, ed. by Z. Griliches and M. D. Intriligator. Amsterdam: North Holland.

GREENE, W. (1997): *Econometric Analysis*, Third Edition. New York: MacMillan.

HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and its Application*. New York: Academic Press.

PARK, J. Y., AND P. C. B. PHILLIPS (1989): "Statistical Inference in Regressions with Integrated Processes: Part 2," *Econometric Theory*, 5, 95–132.

——— (1998): "Nonlinear Regression with Integrated Processes," Cowles Foundation Discussion Paper No. 1182.

——— (1999): ''Asymptotics for Nonlinear Transformations of Integrated Time Series,'' *Econometric Theory*, 15, 269–298.

PHILLIPS, P. C. B. (1998): ''Econometric Analysis of the Fisher Equation,'' Cowles Foundation Discussion Paper No. 1180.

PHILLIPS, P. C. B., AND J. Y. PARK (1998): ''Nonstationary Density Estimation and Kernel Autoregression,'' Yale University, Cowles Foundation Discussion Paper No. 1181.

PHILLIPS, P. C. B., AND V. SOLO (1992): ''Asymptotics for Linear Processes,'' *Annals of Statistics*, 20, 971–1001.

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

PRATT, J. W. (1981): ''Concavity of the Log Likelihood,'' *Journal of the American Statistical Association*, 76, 103–106.

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

WHITE, H. (1994): *Estimation, Inference and Specification Analysis*. Cambridge: Cambridge University Press.

WOOLDRIDGE, J. M. (1994): ''Estimation and Inference for Dependent Processes,'' in *Handbook of Econometrics*, Vol. IV, ed. by R. F. Engle and D. L. McFadden. Amsterdam: North Holland.