Maximum Score Estimation of a Nonstationary Binary Choice Model

Hyungsik Roger Moon

USC Center for Law, Economics & Organization
Research Paper No. C03-15

USC
THE LAW
SCHOOL

CENTER FOR LAW, ECONOMICS AND ORGANIZATION RESEARCH PAPER SERIES

Sponsored by the John M. Olin Foundation

University of Southern California Law School
Los Angeles, CA 90089-0071

Maximum Score Estimation of a Nonstationary Binary Choice Model

Hyungsik Roger Moon*
Department of Economics
University of Southern California

December 2002

Abstract

This paper studies the estimation of a simple binary choice model in which explanatory variables include nonstationary variables and the distribution of the model is not known. We find a set of conditions under which the coefficients of the nonstationary variables are identified. We show that the maximum score estimator of the nonstationary coefficients is consistent.

1 Introduction

The main purpose of this paper is to investigate asymptotic properties of the maximum score estimator of a nonstationary binary choice model. The nonstationary binary choice model is particularly favored when we estimate the decisions that may be affected by fundamental macroeconomic or financial variables, many of which are known to show nonstationary characteristics. Park and Phillips (2000) recently investigated a parametric nonstationary binary choice model in which the known distribution of the model belongs to a certain regular class. The important asymptotic results were that the maximum likelihood estimator of the parametric nonstationary binary choice model converges at a rate of $n^{1/4}$ and its limiting distribution is a mixture of two mixed normal distributions.[1]

In this paper, we consider a nonstationary binary choice model where the distribution of the binary variable is unknown. The model allows for the latent variable to be heterogeneous due, for example, to structural breaks. The model also allows for endogenous regressors.

Various semiparametric estimation methods have been proposed for the conventional binary choice models for random samples with unknown error distributions. (See Horowitz, 1998, for a survey of these developments.) One of them is the maximum score estimation method originally proposed by Manski (1975), which is known to be robust to the heterogeneity of the model. Assuming that the median of the error term in the equation for the latent variable is zero, Manski (1985) proved the (strong) consistency of the maximum score estimator of the parameter in the binary choice model with randomly sampled data.

*I wish to thank the Associate Editor and an anonymous referee for valuable comments. I am grateful to Emmanuel Guerre, Geert Ridder, Quang Vuong, Robert Dekle, and Yong Kim for helpful discussions. In particular, Emmanuel Guerre pointed out an error in Section 2 of an earlier version and gave me very helpful suggestions. I also thank Katherine Goodman for proofreading.

[1] This asymptotic result holds under the assumption that the coefficient vector of the nonstationary variables is different from zero. When it is zero, the maximum likelihood estimator has a convergence rate of order n (see Guerre and Moon, 2002).

Later, Cavanagh(1987) and Kim and Pollard (1990) showed that the convergence order of the maximum score estimator is $n^{1/3}$, slower than the usual rate of $n^{1/2}$, and its limiting distribution is nonstandard. Smoothing the non-differentiable score function and imposing further restrictions on the model, Horowitz (1992) shows that the maximum score estimator can achieve a faster convergence rate (at least $n^{2/5}$ and can make arbitrarily close to $n^{1/2}$) and has a normal limit distribution.

When the model is nonstationary, none of these asymptotic results about the maximum score estimator are known. There are two major findings in this paper. First, we find a set of conditions under which the coefficients of the nonstationary regressors are identified. Second, we show that the maximum score estimator of the coefficient of the nonstationary variables is consistent. For this, we derive the uniform (weak) limit of the sample score function of the data that include nonstationary observations. We prove that the limit function is maximized uniquely at the true parameter of the nonstationary variables. When the data is randomly selected, the uniform convergence of the sample score function can be derived using conventional empirical process theories (see, for example, Kim and Pollard, 1990). However, when the data include nonstationary samples, the conventional empirical process theory cannot be applied. In this context, establishing the uniform convergence result can be considered one of the major theoretical contributions of this paper.

The paper is organized as follows. In Section 2, we introduce a nonstationary binary choice model and regulatory conditions. In Section 3, we define a maximum score estimator, derive the uniform limit of the sample score function, and investigate consistency of the maximum score estimator. The appendix contains all the technical proofs.

Some words on notation: Notation "$\overset{d}{=}$" signifies equivalence in distribution, "$\overset{p}{\to}$" convergence in probability, "$\overset{a.s.}{\to}$" almost sure convergence, and "$\Rightarrow$" convergence in distribution. We denote $\|x\|$ to be the Euclidean norm of vector x. When A is a set, $1\{A\}$ denotes the indicator function of the set A.

2 Model and Assumptions

We start by introducing a binary choice model that includes nonstationary regressors. For a real number a, we denote $sgn\,(a) = 1$, if $a \geq 0$, and $sgn\,(a) = -1$, if $a < 0$. The model assumes that an observable binary variable y_t is generated by

$$y_t = sgn\left(\beta_0' x_t + \gamma_0' z_t - u_t\right), \tag{1}$$

where x_t is a $k-vector$ valued explanatory variable, z_t is an $m-vector$ valued explanatory variable, and u_t is an unobservable error term, satisfying the following assumption:

Assumption 1 *(i) Let $x_t = x_{t-1} + v_t$ with $x_0 = 0$. Then, $v_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$, where $\sum_{j=0}^{\infty} j\,\|c_j\| < \infty$, $\sum_{j=0}^{\infty} c_j \neq 0$, and ε_t is a $k-vector$ valued iid process with zero mean and $E\,\|\varepsilon_t\|^p < \infty$ for some $p > 2$. Define $C = \sum_{j=0}^{\infty} c_j$. (ii) $\Sigma_v = CC'$ is positive definite. Let $s_t = (z_t', u_t)'$. (iii) s_t satisfies $\max_{1\leq t\leq n} E\,\|s_t\|^{2+\delta} < \infty$ for some $\delta > 0$.*

According to Assumption 1, the explanatory variable x_t is integrated implying nonstationary (Assumption 1(i)) and the elements of x_t are not cointegrated (Assumption 1(ii)). Assumption 1(iii) assumes that variables z_t and u_t are not integrated and they have uniformly finite moments that are higher than two.[2]

Note that the model does not assume that the distribution of u_t is known. The model does not impose any restriction on the quantiles of the (conditional) distribution of u_t,

[2]Here we implicitly assume that the integration order of the variables is known or pre-tested.

either. Instead, the model imposes a moment condition as in Assumption 1(iii). Also notice that under Assumption 1 the regressor z_t and the error term v_t generating x_t could be correlated with u_t, and so we allow for endogeneity in the model. Finally, the model allows that u_t could be heterogenous over time and may have structural breaks. The nonstationary binary choice model studied by Park and Phillips (2000) assumes that the error term u_t is conditionally identically and independently distributed (iid) on the information set generated by x_s and $z_s, s \leq t$, the functional form of the density of u_t is known, and the density function belongs to a certain regular parametric family.

Next, we assume that the parameters in the binary choice model are normalized. This normalization assumption is required for the identification of the parameters, which has been assumed in most semiparametric binary choice models for cross section data (for example, Manski 1975, 1985 and Horowitz, 1992).

Assumption 2 *(i)* $\beta_0 \neq 0$. *(ii) The parameter set for* $(\beta', \gamma')'$ *is denoted by* $\mathbb{B} \times \Gamma$, *where* $\mathbb{B}$ *is a unit sphere*[3] *in* $\mathbb{R}^k$ *and* Γ *is a compact subset in* $\mathbb{R}^m$.

Imposing Assumption 2, this paper considers only the nontrivial case, $k \geq 2$.

In many empirical applications, binary choice model (1) could be interpreted using a latent variable. Suppose that there is an unobservable latent variable y_t^* that is generated by

$$y_t^* = \beta_0' x_t + \gamma_0' z_t - u_t. \tag{2}$$

Then, model (1) is equivalent to

$$y_t = sgn\left(y_t^*\right), \tag{3}$$

where y_t is the observable indicator. Under Assumption 1, the nonstationary latent variable y_t^* and the explanatory variables x_t are cointegrated and the coefficient β_0 measures the long-run relationship between y_t^* and x_t. When y_t^* is observable, it is well known that the long-run relationship β_0 can be consistently estimated. However, this result is not known when we observe only the indicator y_t and its distribution is unknown. The main goal of this paper is to show that the long-run relationship parameter β_0 is identified (up to the scale normalization) and find a consistent estimation procedure for β_0. As is well known, when β_0 is identified up to a scale normalization, the consistent estimate of β_0 is useful in determining the (long-run) direction or measuring the (long-run) relative effect of the unobserved latent variable y_t^* with respect to the change of corresponding nonstationary covariates.

Finally, we would like to point out that the parameter γ_0 is not identified under Assumptions 1 and 2. To discuss this in more detail, let $\mathcal{P}$ denote the set of all distributions for a sequence of observations $\{(y_t, x_t, z_t)\}_{t=1,2,\ldots}$. Let $\mathbf{P}$ denote the distribution of the sequence $\{(x_t, z_t, u_t)\}_{t=1,2,\ldots}$. Then, in view of model (1), a typical element in $\mathcal{P}$ is indexed by a parameter $\theta = (\beta, \gamma, \mathbf{P})$, and we denote it $\mathbb{P}_\theta$. The regularity conditions imposed on the parameter θ are that $\|\beta\| = 1$, $\gamma \in \Gamma$ (Assumption 2) and the unknown distribution $\mathbf{P}$ satisfies Assumption 1. Let Θ be the set of all the admissible parameters that satisfy the regularity conditions. We say that a sub-parameter $g(\theta)$ is identified if and only if, for θ and θ' in Θ, $\mathbb{P}_\theta = \mathbb{P}_{\theta'}$ implies $g(\theta) = g(\theta')$.

Without loss of generality, we assume that 0 is admissible for γ. Now, letting $\mathbf{P}_1$ be the distribution of $\{(x_t, z_t, u_t - \gamma_0' z_t)\}$, consider $\theta_0 = (\beta_0, \gamma_0, \mathbf{P}_0)$ with $\gamma_0 \neq 0$ and $\theta_1 = (\beta_0, 0, \mathbf{P}_1)$. It is straightforward to see that θ_1 is an admissible parameter because

[3] That is, $\mathbb{B} = \left\{\beta \in \mathbb{R}^k : \|\beta\| = 1\right\}$.

$0 \in \Gamma$ and under $\mathbf{P}_1$ Assumption 1 is satisfied. Then, under Model $\mathbb{P}_{\theta_1}$, the observation y_t is generated by

$$y_t = sgn\left(\beta_0' x_t - u_t + \gamma_0' z_t\right),$$

which is identical to the observation y_t generated under Model $\mathbb{P}_{\theta_0}$. Therefore, under Assumptions 1 and 2, the parameter γ_0 of the stationary component is not identified.

Before moving on, we introduce a result that is helpful in analyzing the weak limit of the objective function that will be introduced in the next section. Let

$$V_n^0(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} v_t,$$

and

$$S_n^0(r) = \left(Z_n^0(r)', U_n^0(r)\right)' = \frac{s_{[nr]}}{\sqrt{n}} = \left(\frac{z_{[nr]}}{\sqrt{n}}', \frac{u_{[nr]}}{\sqrt{n}}\right)'.$$

Assume that $S_n^0(r)$ and $V_n^0(r)$ are stochastic processes on $D[0,1]^{m+1}$ and $D[0,1]^k$, respectively, where $D[0,1]^l$ is the $l-$fold Cartesian product of the space $D[0,1]$ that is the set of cadlag functions on the interval $[0,1]$, with the uniform topology.

Lemma 3 *Under Assumption 1, there exists a probability space* $(\Omega^*, \mathcal{F}^*, \mathbb{P}^*)$ *supporting sequences of random vectors* $V_n(r)$ *and* $S_n(r) = \left(Z_n(r)', U_n(r)\right)'$ *such that*

(a) $V_n(\cdot) \overset{d}{=} V_n^0(\cdot)$ *and*

$$\sup_{0 \le r \le 1} \|V_n(r) - V(r)\| \to 0 \ a.s.,$$

where $V(r)$ *is a Brownian motion in* $(\Omega, \mathcal{F}, \mathbb{P})$ *with covariance matrix* Σ_v,

(b) $S_n(\cdot) \overset{d}{=} S_n^0(\cdot)$ *with*

$$\sup_{0 \le r \le 1} \|S_n(r)\| \to 0 \ a.s.$$

3 Maximum Score Estimation of β_0

The maximum score estimator originally proposed by Manski (1975) is a binary analog of the least absolute deviation estimator of a linear median regression model. Assuming that the median of the error term in the equation for the latent variable is zero, Manski (1985) proved that the maximum score estimator is strongly consistent. The main purpose of this section is to investigate the identification of β_0 and to show the consistency of the maximum score estimator of the nonstationary binary choice model (1).

First, let $\left(\hat{\beta}, \hat{\gamma}\right)$ denote the maximum score estimator that maximizes the following sample score function,

$$Q_n^0(\beta, \gamma) = \frac{1}{n} \sum_{t=1}^{n} sgn\left(\beta' x_t + \gamma' z_t\right) y_t, \tag{4}$$

over the parameter set $\mathbb{B} \times \Gamma$. Define

$$Q_n(\beta, \gamma) = \int_0^1 sgn\left(\beta' V_n(r) + \gamma' Z_n(r)\right) sgn\left(\beta_0' V_n(r) + \gamma_0' Z_n(r) - U_n(r)\right) dr$$

and

$$Q(\beta) = \int_0^1 sgn\left(\beta' V(r)\right) sgn\left(\beta_0' V(r)\right) dr,$$

where $V(r)$ is the Brownian motion in Lemma 3. Notice that $Q(\beta)$ does not depend on the parameter γ. Rewriting

$$Q_n^0(\beta,\gamma) = \int_0^1 sgn\left(\beta' V_n^0(r) + \gamma' Z_n^0(r)\right) sgn\left(\beta_0' V_n^0(r) + \gamma_0' Z_n^0(r) - U_n^0(r)\right) dr,$$

we can notice that the two objective functions, Q_n and Q_n^0, have identical functional forms as functionals of $\left(\beta', \gamma', V_n(r)', S_n(r)'\right)'$ and $\left(\beta', \gamma', V_n^0(r)', S_n^0(r)'\right)'$, respectively. It follows, then, by Lemma 3 that

$$Q_n^0(\cdot,\cdot) \overset{d}{=} Q_n(\cdot,\cdot).$$

3.1 Uniform Limit of the Objective Function $Q_n(\beta,\gamma)$

In view of Lemma 3, it seems that $Q(\beta)$ would be a natural candidate for the (uniform) limit of $Q_n(\beta,\gamma)$. The main difficulty we would face in proving this is that the function $sgn(x)$ is not continuous. In this case, the objective function $Q_n(\beta,\gamma)$ does not satisfy the regularity conditions for Theorem 3.1 of Park and Phillips (2001) that establishes the uniform convergence for a sample average of a regular function of a partial sum process.

First, find the finite dimensional limit of the objective function $Q_n(\beta,\gamma)$. Define $f: D[0,1]^5 \to \mathbb{R}$ to be

$$f(x_1(r), x_2(r), x_3(r), x_4(r), x_5(r)) = \int_0^1 sgn(x_2(r) + x_4(r))\, sgn(x_1(r) + x_3(r) - x_5(r))\, dr.$$

Let C_β and C_0 denote the sets of all the continuous time paths on [0,1] of Brownian motions $\beta' V(r)$ and $\beta_0' V(r)$, respectively. Then, C_β and C_0 are subsets of $C[0,1]$, the set of all the continuous functions on $[0,1]$, and they are separable with respect to the uniform topology endowed on the $D[0,1]$.

Let $X_n(r) = \left(\beta_0' V_n(r), \beta' V_n(r), \gamma_0' Z_n(r), \gamma' Z_n(r), U_n(r)\right)'$ and $X(r) = \left(\beta_0' V(r), \beta' V(r), \mathbf{0}, \mathbf{0}, \mathbf{0}\right)$, where $\mathbf{0}$ is the null function on [0,1]. From Lemma 3, we have

$$X_n(r) \overset{a.s.}{\to} X(r)$$

uniformly in $r \in [0,1]$. Also, by definition,

$$\mathbb{P}^*\left(X(r) \in C_0 \times C_\beta \times \mathbb{O} \times \mathbb{O} \times \mathbb{O}\right) = 1,$$

where $\mathbb{O} = \{\mathbf{0}\}$.

For any sequence $(x_{1n}(r), x_{2n}(r), x_{3n}(r), x_{4n}(r), x_{5n}(r))$ in $D[0,1]^5$ that converges to $(x_1(r), x_2(r), \mathbf{0}, \mathbf{0}, \mathbf{0})$ in $C_0 \times C_\beta \times \mathbb{O} \times \mathbb{O} \times \mathbb{O}$ uniformly in $r \in [0,1]$, by modifying Lemma 12 in the Appendix, we may deduce that the functional f is continuous at all the points in $C_0 \times C_\beta \times \mathbb{O} \times \mathbb{O} \times \mathbb{O}$. Then, by applying the continuous mapping theorem, the finite dimensional convergence of $Q_n(\beta,\gamma)$ to $Q(\beta)$ follows. Summarizing this, we have the following lemma:

Lemma 4 *Suppose that Assumptions 1 and 2 hold. For any K-tuple $((\beta_1,\gamma_1), ..., (\beta_K,\gamma_K))$,*

$$\left(Q_n(\beta_1,\gamma_1), ..., Q_n(\beta_K,\gamma_K)\right)' \overset{a.s.}{\to} \left(Q(\beta_1), ..., Q(\beta_K)\right)'.$$

To establish consistency of an extremum estimator, we need a convergence stronger than the finite dimensional convergence shown in Lemma 4. Before we introduce the main result, the uniform convergence of the objective function, we introduce another useful lemma. Define

$$T_n(\beta,\gamma,M) = Q_n(\beta,\gamma)\,1\left\{\sup_{r\in[0,1]}\left\|\left(V_n(r)',Z_n(r)'\right)\right\| \le M\right\}$$

and

$$T(\beta,M) = Q(\beta)\,1\left\{\sup_{r\in[0,1]}\|V(r)\| \le M\right\}.$$

Lemma 5 *Suppose that Assumptions 1 and 2 hold. For any given $M>0$, as $n\to\infty$*

$$\sup_{(\beta,\gamma)\in\mathbb{B}\times\Gamma}|T_n(\beta,\gamma,M)-T(\beta,M)| \xrightarrow{a.s.} 0.$$

Now we establish the uniform convergence of $Q_n(\beta,\gamma)$ to $Q(\beta)$. Since $|Q_n(\beta,\gamma)|, |Q(\beta)| \le 1$, we have

$$\begin{aligned}
&\sup_{(\beta,\gamma)\in\mathbb{B}\times\Gamma}|Q_n(\beta,\gamma)-Q(\beta)| \\
\le\;&\sup_{(\beta,\gamma)\in\mathbb{B}\times\Gamma}|T_n(\beta,\gamma,M)-T(\beta,M)| \\
&+1\left\{\sup_{r\in[0,1]}\left\|\left(V_n(r)',Z_n(r)'\right)\right\| > M\right\} + 1\left\{\sup_{r\in[0,1]}\|V(r)\| > M\right\}.
\end{aligned}$$

Since $\sup_{r\in[0,1]}\left\|\left(V_n(r)',Z_n(r)'\right)\right\| = O_{a.s.}(1)$, for any given $\varepsilon>0$, we can choose a constant M such that

$$\limsup_{n\to\infty}\mathbb{P}^*\left\{\sup_{r\in[0,1]}\left\|\left(V_n(r)',Z_n(r)'\right)\right\| > M\right\} \le \frac{\varepsilon^2}{6} \tag{5}$$

$$\mathbb{P}^*\left\{\sup_{r\in[0,1]}\|V(r)\| > M\right\} \le \frac{\varepsilon^2}{6}. \tag{6}$$

By the Markov inequality and from (5) and (6), we deduce that

$$\begin{aligned}
&\limsup_{n\to\infty}\mathbb{P}^*\left\{\sup_{(\beta,\gamma)\in\mathbb{B}\times\Gamma}|Q_n(\beta,\gamma)-Q(\beta)| > \varepsilon\right\} \\
\le\;&\limsup_{n\to\infty}\mathbb{P}^*\left\{\sup_{(\beta,\gamma)\in\mathbb{B}\times\Gamma}|T_n(\beta,\gamma,M)-T(\beta,M)| > \frac{\varepsilon}{3}\right\} \\
&+\frac{3}{\varepsilon}\limsup_{n\to\infty}\mathbb{P}^*\left\{\sup_{r\in[0,1]}\left\|\left(V_n(r)',Z_n(r)'\right)\right\| > M\right\} + \frac{3}{\varepsilon}\limsup_{n\to\infty}\mathbb{P}^*\left\{\sup_{r\in[0,1]}\|V(r)\| > M\right\} \\
\le\;&\varepsilon.
\end{aligned}$$

This shows that $Q_n(\beta,\gamma)$ converges in probability to $Q(\beta)$ uniformly in β and γ. Summarizing this, we have the following theorem.

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

$$\sup_{(\beta,\gamma)\in\mathbb{B}\times\Gamma}|Q_n(\beta,\gamma)-Q(\beta)| \xrightarrow{p} 0.$$

The main reason that the limit function Q depends only on β and not on γ, is that the score signal from the nonstationary variables dominates that from the stationary variables (see Lemma 3).

3.2 Identification of β_0

Lemma 7 *(a) $Q(\beta) < 1$ a.s. if $\beta \neq \beta_0$. (b) $Q(\beta)$ is continuous in β.*

Lemma 7 shows that with probability one, $Q(\beta)$ takes a value strictly less than one if $\beta \neq \beta_0$. On the other hand, when $\beta = \beta_0$,

$$Q(\beta_0) = \int_0^1 sgn\left(\beta_0' V(r)\right)^2 dr = 1. \tag{7}$$

Therefore, with probability one, the continuous function $Q(\beta)$ has a unique maximum at the true parameter β_0. Furthermore, since the parameter set $\mathbb{B}$ is compact, for any $\delta > 0$, we can find an $\varepsilon > 0$ such that

$$\sup_{\beta \in \mathbb{B} \text{ s.t. } \|\beta - \beta_0\| > \delta} Q(\beta) < 1 - \varepsilon \text{ a.s.} \tag{8}$$

From (7) and (8), we deduce that under Assumptions 1 and 2, the coefficient β_0 of the nonstationary variable x_t is identified.

This identification result together with the nonidentification of γ_0 in the previous section can be compared to some of the well known results in the literature. In the parametric nonstationary binary choice model studied by Park and Phillips (2000) where the distribution of the model is known, all the coefficients of the nonstationary and the stationary components are identified and consistently estimable (see Remark 4 on page 1257 of Park and Phillips, 2000). However, when the distribution of the model is unknown and it could be heterogeneous, under the regularity conditions in the paper, we can identify only the normalized coefficient of the nonstationary regressors, β_0.

On the other hand, when the data are from random samples, Manski (1985)[4] found that β_0 is identified under the assumption of quantile independence of u_t and some restrictions on the distribution of the regressors. The former restrictions exclude the regressors whose distribution support is degenerated or finite. Compared to these, the coefficient of the nonstationary covariates β_0 is identified without any quantile restriction on the conditional distribution of u_t. Crucial conditions used in identifying β_0 are that u_t does not have stochastic trends (*i.e.*, y_t^* and x_t are cointegrated), there is no-cointegration relation in x_t, and the error process generating x_t, regressor z_t, and the error u_t should satisfy the moment conditions in Assumption 1[5]. The condition of no cointegration relation in x_t corresponds to the non-degeneracy condition in Assumption 2(a) of Manski (1985). In view of the functional central limit theorem in Lemma 3(a), the distribution of the standardized regressor $\frac{x_t}{\sqrt{n}}$ has an unbounded continuous support in the limit.

3.3 Consistency of $\hat{\beta}$

Let $\hat{\beta}^*$ and $\hat{\gamma}^*$ maximize $Q_n(\beta, \gamma)$ over the compact parameter set $\mathbb{B} \times \Gamma$. Then, from (8), by the definition of $\left(\hat{\beta}^*, \hat{\gamma}^*\right)$, and by Theorem 6, we have

$$\mathbb{P}^* \left\{ \left\| \hat{\beta}^* - \beta_0 \right\| > \delta \right\}$$

[4] Manski (1988) discusses more detailed conditions for identification of the binary choice model.

[5] Compared to this, notice that Manski (1985) imposes no restriction on the moments of the regressors

$$
\begin{aligned}
&\leq \quad \mathbb{P}^*\left\{Q\left(\beta_0\right)-Q_n\left(\hat{\beta}^*, \hat{\gamma}^*\right)+Q_n\left(\hat{\beta}^*, \hat{\gamma}^*\right)-Q\left(\hat{\beta}^*\right)>\varepsilon\right\} \\
&\leq \quad \mathbb{P}^*\left\{Q\left(\beta_0\right)-Q_n\left(\beta_0, \gamma_0\right)+Q_n\left(\hat{\beta}^*, \hat{\gamma}^*\right)-Q\left(\hat{\beta}^*\right)>\varepsilon\right\} \\
&\leq \quad \mathbb{P}^*\left\{\sup_{(\beta,\gamma)\in\mathbb{B}\times\Gamma}\left|Q_n\left(\beta,\gamma\right)-Q\left(\beta\right)\right|>\varepsilon\right\} \\
&\rightarrow \quad 0.
\end{aligned}
$$

Therefore, it follows that in the probability space $(\Omega^*, \mathcal{F}^*, \mathbb{P}^*)$,

$$\hat{\beta}^* \xrightarrow{p} \beta_0.$$

Recall that $Q_n^0(\cdot,\cdot) \overset{d}{=} Q_n(\cdot,\cdot)$ and notice that $\hat{\beta}^* \overset{d}{=} \hat{\beta}$.[6] Therefore, on the original probability space, we have

$$\hat{\beta} \xrightarrow{p} \beta_0.$$

The following theorem summarizes this.

Theorem 8 *Under Assumptions 1 and 2, the maximum score estimator $\hat{\beta}$ is consistent.*

4 Concluding Remarks

This paper studies the maximum score estimator of a nonstationary binary choice model where explanatory variables include both stationary variables and nonstationary variables. There are two major findings of the paper. First, we find a set of restrictions under which the coefficient of the nonstationary explanatory variables are identified. Second, we show that the maximum score estimator of the coefficient of the nonstationary variables is consistent.

There are several important extensions we may consider for future projects. First, it is important to find sufficient conditions for the identification of the stationary component coefficients γ_0 and develop a consistent estimation procedure. Having a consistent estimate of γ_0 as well as a consistent estimate of β_0 makes the study of the structure analysis of the nonstationary model (1) more useful (for example, see Manski, 1988). Second, in order to perform a test for restrictions on the parameters, we need to derive the limiting distribution of the maximum score estimator. This is a considerably harder problem mainly because the score objective function is non-differentiable. To cope with this difficulty, smoothing the objective function as in Horowitz (1992) would be a natural extension.

5 Appendix: Technical Proofs

5.1 Appendix A: Useful Results

Lemma 9 *Let $W(r)$ be a standard Brownian motion. Define $A_t^+ = \int_0^t 1\{W(r) \geq 0\}\, dr$ and $A_t^- = \int_0^t 1\{W(r) < 0\}\, dr$. Then, the laws of A_1^+ and A_t^- are the Arcsine law on [0,1] whose density is $\frac{1}{\pi\sqrt{x(1-x)}}$ on [0,1] with respect to the Lebesgue measure.*

[6] We have $\hat{\beta}^* \overset{d}{=} \hat{\beta}$ because the functional forms of Q_n and Q_n^0 are identical and in consequence, the two argmaxes $\left(\hat{\beta}^*, \hat{\gamma}^*\right)$ and $\left(\hat{\beta}, \hat{\gamma}\right)$ have identical functional forms in $\left(V_n(r)', S_n(r)'\right)'$ and $\left(V_n^0(r)', S_n^0(r)'\right)'$, respectively.

Proof See Theorem 2.7 in Revuz and Yor (1999). ■

Lemma 10 *Suppose that $B_1(r)$ and $B_2(r)$ are two standard Brownian motions that are independent of each other. Then,*

$$-1 < \int_0^1 sgn(B_1(r))\, sgn(B_2(r))\, dr < 1$$

almost surely.

Proof

Define $A_{11}^+ = \int_0^1 1\{0 \le B_1(r)\}\, dr$ and $A_{21}^+ = \int_0^1 1\{0 \le B_2(r)\}\, dr$.[7] First, suppose that

$$\int_0^1 sgn(B_1(r))\, sgn(B_2(r))\, dr = 1$$

with a positive probability. Then, with a positive probability it follows that $sgn(B_1(r)) = sgn(B_2(r))$ for almost all $0 \le r \le 1$ in Lebesgue measure, which implies that

$$A_{11}^+ = A_{21}^+ \tag{9}$$

with a positive probability. However, by Lemma 9 and by the independence of $B_1(r)$ and $B_2(r)$, the joint density function of A_{11}^+ and A_{21}^+ is the product of the marginal densities of A_{11}^+ and A_{21}^+ that is continuous with respect to the product of the Lebesgue measures. So, the event that

$$A_{11}^+ = A_{21}^+$$

occurs with zero probability, which contradicts to (9). Therefore, $\int_0^1 sgn(B_1(r))\, sgn(B_2(r))\, dr < 1$ almost surely. Next, using similar arguments, we can show that $-1 < \int_0^1 sgn(B_1(r))\, sgn(B_2(r))\, dr$ almost surely. In consequence, we have

$$-1 < \int_0^1 sgn(B_1(r))\, sgn(B_2(r))\, dr < 1$$

almost surely, as required. ■

The implication of the lemma is that two independent Brownian motion sample paths do not stay on the same side, nor on the opposite sides for all $r \in (0,1]$.

Lemma 11 *Let $W(r)$ be a standard Brownian motion. For any $\varepsilon > 0$, $W(r)$ changes sign infinitely many times in the time interval $[0, \varepsilon]$.*

Proof See Problem 7.18 on page 94 of Karatzas and Shreve (1991). ■

Lemma 12 *Let $B(r)$ be a Brownian motion with continuous paths on time interval [0,1]. Let $C_0[0,1]$ be the collection of all the time paths of $B(r)$. Then, for any $y_n(r) \in D[0,1]$ and $y(r) \in C_0[0,1]$ with*

$$\sup_{r \in [0,1]} |y_n(r) - y(r)| \to 0, \tag{10}$$

we have

$$\int_0^1 sgn(y_n(r))\, dr \to \int_0^1 sgn(y(r))\, dr$$

as $n \to \infty$.

[7] Notice that $\int_0^1 1\{B_1(r) = 0\}\, dr = \int_0^1 1\{B_2(r) = 0\}\, dr = 0$ by Proposition 3.12 of Revuz and Yor (1999). Thus, it also holds that $A_{11}^+ = \int_0^1 1\{0 < B_1(r)\}\, dr$ and $A_{21}^+ = \int_0^1 1\{0 < B_2(r)\}\, dr$.

Proof First, notice that the limit of

$$\lim_{\eta\to 0}\frac{1}{2\eta}\int_0^1 1\{|B(r)|\le\eta\}\,dr$$

exists because the local time of a Brownian motion is well defined. From this, for any $\varepsilon>0$, we can choose $\eta>0$ such that

$$\int_0^1 1\{|y(r)|\le\eta\}\,dr<\frac{\varepsilon}{2}.$$

Also, from (10), we can choose n_0 such that whenever $n\ge n_0$,

$$\sup_{r\in[0,1]}|y_n(r)-y(r)|<\frac{\eta}{2}. \tag{11}$$

Notice that

$$\begin{aligned}
&\left|\int_0^1 sgn(y_n(r))\,dr-\int_0^1 sgn(y(r))\,dr\right| \\
\le\;& \int_0^1 |sgn(y_n(r))-sgn(y(r))|\,dr \qquad (12)\\
\le\;& \int_0^1 |sgn(y_n(r))-sgn(y(r))|\,1\{|y(r)|>\eta\}\,dr \\
&+2\int_0^1 1\{|y(r)|\le\eta\}\,dr.
\end{aligned}$$

If $n\ge n_0$, then,

$$\int_0^1 |sgn(y_n(r))-sgn(y(r))|\,1\{|y(r)|>\eta\}\,dr=0$$

due to (11). Also, we have

$$\int_0^1 1\{|y(r)|\le\eta\}\,dr<\frac{\varepsilon}{2}.$$

Therefore, whenever $n\ge n_0$, we have

$$\left|\int_0^1 sgn(y_n(r))\,dr-\int_0^1 sgn(y(r))\,dr\right|<\varepsilon,$$

and we complete the proof. ■

Let $V_n(r)$, $Z_n(r)$, and $V(r)$ be the processes in Lemma 3. Suppose that $\Omega_0^*\subset\Omega^*$ is a set with $\mathbb{P}^*(\Omega_0^*)=1$ in which $V_n(r)\to V(r)$ and $Z_n(r)\to 0$ uniformly in r and the local time of a Brownian motion exists. Define $\alpha=(\beta',\gamma')'$ and $\tilde V_n(r)=\left(V_n(r)',Z_n(r)'\right)'$. The following two lemmas hold for any fixed $\omega^*\in\Omega_0^*$.

Lemma 13 *Suppose $\bar\alpha=\left(\bar\beta',\bar\gamma'\right)'\in\mathbb{B}\times\Gamma$ and $\varepsilon>0$ are given. Fix $\omega^*\in\Omega_0^*$. Then, we can choose $\kappa(\omega^*,\bar\alpha,\varepsilon)>0$ such that*

$$\limsup_{n\to\infty}\int_0^1 1\left\{\left|\bar\alpha'\tilde V_n(r)(\omega^*)\right|<\kappa(\omega^*,\bar\alpha,\varepsilon)\right\}dr<\varepsilon.$$

Proof. From the existence of the local time of Brownian motion $\bar{\beta}' V(r)$, for the given $\omega^*, \bar{\alpha}$, and $\varepsilon > 0$, we can choose $2\kappa(\omega^*, \bar{\alpha}, \varepsilon)$ such that

$$\int_0^1 1\left\{\left|\bar{\beta}' V(r)(\omega^*)\right| < 2\kappa(\omega^*, \bar{\alpha}, \varepsilon)\right\} dr < \varepsilon.$$

Let $M_\gamma = \sup_{\gamma \in \Gamma} \|\gamma\|$. Since the parameter set Γ is compact on $\mathbb{R}^m$, M_γ is finite. Recalling that $\|\bar{\beta}\| = 1$, we have

$$\begin{aligned}
& \left|\bar{\alpha}' \tilde{V}_n(r)(\omega^*) - \bar{\beta}' V(r)(\omega^*)\right| \\
= & \left|\bar{\beta}' V_n(r)(\omega^*) + \bar{\gamma}' Z_n(r)(\omega^*) - \bar{\beta}' V(r)(\omega^*)\right| \\
\leq & \left|\bar{\beta}' V_n(r)(\omega^*) - \bar{\beta}' V(r)(\omega^*)\right| + |\bar{\gamma}' Z_n(r)(\omega^*)| \\
\leq & \|V_n(r)(\omega^*) - V(r)(\omega^*)\| + \sup_{\gamma \in \Gamma} \|\gamma\| \|Z_n(r)(\omega^*)\| \\
\leq & \|V_n(r)(\omega^*) - V(r)(\omega^*)\| + M_\gamma \|Z_n(r)(\omega^*)\| \\
\to & \ 0
\end{aligned}$$

uniformly in r. So, we can choose $n_0(\omega^*, \bar{\alpha}, \varepsilon)$ such that

$$\sup_{r \in [0,1]} \left|\bar{\alpha}' \tilde{V}_n(r)(\omega^*) - \bar{\beta}' V(r)(\omega^*)\right| < \kappa(\omega^*, \bar{\alpha}, \varepsilon)$$

whenever $n \geq n_0(\omega^*, \bar{\alpha}, \varepsilon)$. In consequence, we have

$$\begin{aligned}
& \int_0^1 1\left\{\left|\bar{\alpha}' \tilde{V}_n(r)(\omega^*)\right| < \kappa(\omega^*, \bar{\alpha}, \varepsilon)\right\} dr \\
\leq & \int_0^1 1\left\{\left|\bar{\beta}' V(r)(\omega^*)\right| < 2\kappa(\omega^*, \bar{\alpha}, \varepsilon)\right\} dr < \varepsilon
\end{aligned} \tag{13}$$

whenever $n \geq n_0(\omega^*, \bar{\alpha}, \varepsilon)$, and we have the required result. ■

Lemma 14 *Suppose that $M > 0$ and $\varepsilon > 0$ are given. For any fixed $\omega^* \in \Omega_0^*$, it is possible to choose $n_1(\omega^*, \varepsilon)$ and $\delta(\omega^*, \varepsilon, M)$ such that whenever $n \geq n_1(\omega^*, \varepsilon)$ and $\left\|(\beta', \gamma') - (\bar{\beta}', \bar{\gamma}')\right\| < \delta(\omega^*, \varepsilon, M)$, we have*

$$\left|Q_n(\beta, \gamma) - Q_n(\bar{\beta}, \bar{\gamma})\right| 1\left\{\sup_{r \in [0,1]} \left\|(V_n(r)', Z_n(r)')\right\| \leq M\right\} < 4\varepsilon.$$

Proof. Recall the notation $\alpha = (\beta', \gamma')'$ and $\tilde{V}_n(r) = (V_n(r)', Z_n(r)')'$.
Step 1: First we fix $\bar{\alpha} = (\bar{\beta}', \bar{\gamma}')' \in \mathbb{B} \times \Gamma$. Notice that

$$\begin{aligned}
& \left|Q_n(\beta, \gamma) - Q_n(\bar{\beta}, \bar{\gamma})\right| 1\left\{\sup_{r \in [0,1]} \left\|(V_n(r)', Z_n(r)')\right\| \leq M\right\} \\
\leq & \left(\int_0^1 \left|sgn\left(\alpha' \tilde{V}_n(r)\right) - sgn\left(\bar{\alpha}' \tilde{V}_n(r)\right)\right| dr\right) 1\left\{\sup_{r \in [0,1]} \left\|\tilde{V}_n(r)\right\| \leq M\right\}.
\end{aligned}$$

Now choose $n_1 = n_0(\omega^*, \bar{\alpha}, \varepsilon)$ and $\kappa_0 = \kappa(\omega^*, \bar{\alpha}, \varepsilon)$ in Lemma 13. Then, for the fixed $\omega^* \in \Omega_0^*$, if $n \geq n_1$,

$$
\begin{aligned}
& \left(\int_0^1 \left| sgn\left(\alpha'\tilde{V}_n(r)\right) - sgn\left(\bar{\alpha}'\tilde{V}_n(r)\right)\right| dr\right) 1\left\{\sup_{r\in[0,1]} \left\|\tilde{V}_n(r)\right\| \leq M\right\} \\
\leq & \left(\int_0^1 \left| sgn\left(\alpha'\tilde{V}_n(r)\right) - sgn\left(\bar{\alpha}'\tilde{V}_n(r)\right)\right| 1\left\{\left|\bar{\alpha}'\tilde{V}_n(r)\right| \geq \kappa_0\right\} dr\right) 1\left\{\sup_{r\in[0,1]} \left\|\tilde{V}_n(r)\right\| \leq M\right\} \\
& +2\int_0^1 1\left\{\left|\bar{\alpha}'\tilde{V}_n(r)\right| < \kappa_0\right\} dr \\
\leq & \left(\int_0^1 \left| sgn\left(\alpha'\tilde{V}_n(r)\right) - sgn\left(\bar{\alpha}'\tilde{V}_n(r)\right)\right| 1\left\{\left|\bar{\alpha}'\tilde{V}_n(r)\right| \geq \kappa_0\right\} dr\right) 1\left\{\sup_{r\in[0,1]} \left\|\tilde{V}_n(r)\right\| \leq M\right\} + 2\varepsilon,
\end{aligned}
$$

where the first inequality holds since $|sgn(x) - sgn(y)| \leq 2$ and the last inequality holds by Lemma 13.

Choose

$$
\delta(\omega^*, \bar{\alpha}, \varepsilon, M) < \frac{\kappa_0}{2M}.
$$

Suppose that $\|\alpha - \bar{\alpha}\| < \delta(\omega^*, \bar{\alpha}, \varepsilon, M)$. Then, for the fixed $\omega^* \in \Omega_0^*$, $n \geq n_1$ implies that

$$
\begin{aligned}
& \sup_{r\in[0,1]} \left|\alpha'\tilde{V}_n(r) - \bar{\alpha}'\tilde{V}_n(r)\right| 1\left\{\sup_{r\in[0,1]} \left\|\tilde{V}_n(r)\right\| \leq M\right\} \\
\leq & \|\alpha - \bar{\alpha}\| \sup_{r\in[0,1]} \left\|\tilde{V}_n(r)\right\| 1\left\{\sup_{r\in[0,1]} \left\|\tilde{V}_n(r)\right\| \leq M\right\} \\
\leq & \delta(\omega^*, \bar{\alpha}, \varepsilon, M) \sup_{r\in[0,1]} \left\|\tilde{V}_n(r)\right\| 1\left\{\sup_{r\in[0,1]} \left\|\tilde{V}_n(r)\right\| \leq M\right\} \\
< & \frac{\kappa_0}{2},
\end{aligned}
$$

and, in consequence, if $n \geq n_1$,

$$
\left(\int_0^1 \left| sgn\left(\alpha'\tilde{V}_n(r)\right) - sgn\left(\bar{\alpha}'\tilde{V}_n(r)\right)\right| 1\left\{\left|\bar{\alpha}'\tilde{V}_n(r)\right| \geq \kappa_0\right\} dr\right) 1\left\{\sup_{r\in[0,1]} \left\|\tilde{V}_n(r)\right\| \leq M\right\} = 0
$$

for the fixed $\omega^* \in \Omega_0^*$ because $\alpha'\tilde{V}_n(r)$ and $\bar{\alpha}'\tilde{V}_n(r)$ have the same sign when $\left|\bar{\alpha}'\tilde{V}_n(r)\right| \geq \kappa_0$ with $n > n_1$.

Thus, if $n \geq n_0(\omega^*, \bar{\alpha}, \varepsilon)$ and $\|\alpha - \bar{\alpha}\| < \delta(\omega^*, \bar{\alpha}, \varepsilon, M)$,

$$
|Q_n(\alpha) - Q_n(\bar{\alpha})| 1\left\{\sup_{r\in[0,1]} \left\|\tilde{V}_n(r)\right\| \leq M\right\} < 2\varepsilon. \tag{14}
$$

Step 2: For each $\alpha = (\beta', \gamma')' \in \mathbb{B} \times \Gamma$, we choose $n_1(\omega^*, \alpha, \varepsilon)$ and $\delta(\omega^*, \alpha, \varepsilon, M)$. Let $\mathbb{S}(\alpha, r)$ denote the open ball centered at α with radius r. Since the parameter set $\mathbb{B} \times \Gamma$ is compact, we can choose a finite number of $\alpha_l's$, say L, such that

$$
\cup_{l=1}^{L} \mathbb{S}\left(\alpha_l, \frac{\delta(\omega^*, \alpha_l', \varepsilon, M)}{2}\right) \supset \mathbb{B} \times \Gamma.
$$

For notational simplicity, write $\delta(\omega^*, \alpha_l, \varepsilon, M) = \delta_l$. Set

$$\delta(\omega^*, \varepsilon, M) = \min_{1 \le l \le L} \left\{ \frac{\delta_1}{2}, ..., \frac{\delta_L}{2} \right\}$$

and

$$n_1(\omega^*, \varepsilon) = \max_{1 \le l \le L} \{n_0(\omega^*, \alpha_1, \varepsilon), ..., n_0(\omega^*, \alpha_L, \varepsilon)\}.$$

Notice that for any $\|\alpha - \bar{\alpha}\| \le \delta(\omega^*, \varepsilon, M)$, we can find an α_l such that

$$\|\alpha - \alpha_l\| \le \delta(\omega^*, \alpha_l, \varepsilon, M)$$

and

$$\|\alpha_l - \bar{\alpha}\| \le \delta(\omega^*, \alpha_l, \varepsilon, M).$$

Therefore, for fixed $\omega^* \in \Omega_0^*$, if $n \ge n_1(\omega^*, \varepsilon)$ and $\|\alpha - \bar{\alpha}\| \le \delta(\omega^*, \varepsilon, M)$, we have

$$\begin{aligned} & |Q_n(\alpha) - Q_n(\bar{\alpha})| 1\left\{ \sup_{r\in[0,1]} \left\|\tilde{V}_n(r)\right\| \le M \right\} \\ \le\ & |Q_n(\alpha) - Q_n(\alpha_l)| 1\left\{ \sup_{r\in[0,1]} \left\|\tilde{V}_n(r)\right\| \le M \right\} \\ & + |Q_n(\bar{\alpha}) - Q_n(\alpha_l)| 1\left\{ \sup_{r\in[0,1]} \left\|\tilde{V}_n(r)\right\| \le M \right\} \\ \le\ & 4\varepsilon, \end{aligned}$$

where the last inequality holds by (14) and we complete the proof. ■

5.2 Appendix B: Proofs of Main Results

Proof of Lemma 3

The proof of Part (a) is quite similar to the proof of Lemma 1(c) in Park and Phillips (2000). First, by Theorem 3.4 of Phillips and Solo (1992), we have

$$V_n^0(r) \Rightarrow V(r)$$

in $D[0,1]^k$ with the uniform metric on the original probability space. Also, for any $\varepsilon > 0$,

$$\begin{aligned} \mathbb{P}\left\{ \sup_{1\le t\le n} \frac{\|s_t\|}{\sqrt{n}} > \varepsilon \right\} &\le \frac{1}{n\varepsilon^2} \sum_{t=1}^{n} E\left[\|s_t\|^2 1\left\{\|s_t\|^2 > n\varepsilon^2\right\}\right] \\ &\le \frac{1}{\varepsilon^2} \sup_{1\le t\le n} E\left[\|s_t\|^2 1\left\{\|s_t\|^2 > n\varepsilon^2\right\}\right] \\ &\to 0 \text{ as } n \to \infty \end{aligned}$$

because

$$\sup_{1\le t\le n} E\left[\|s_t\|^2 1\left\{\|s_t\|^\delta > n^{\frac{\delta}{2}}\varepsilon^\delta\right\}\right] \le \frac{\sup_{1\le t\le n} E\|s_t\|^{2+\delta}}{n^{\frac{\delta}{2}}\varepsilon^\delta} 0 \text{ as } n \to \infty$$

by Assumption 1(iv). Thus,

$$S_n^0(r) \to_p 0$$

in $D[0,1]^{m+1}$ with the uniform metric on the original probability space. From this, we have the following joint limit

$$\begin{pmatrix} V_n^0(r) \\ S_n^0(r) \end{pmatrix} \Rightarrow \begin{pmatrix} V(r) \\ 0 \end{pmatrix}.$$

The required result follows by the representation theorem in Pollard (1984, pp 71-72). ■

Proof of Lemma 7
Part (a).

For notational convenience, write

$$V_\beta(r) = \beta' V(r)$$

and

$$V_0(r) = \beta_0' V(r).$$

Recall that $\beta_0 \neq 0$ and Σ_v is positive definite (see model (1) and Assumption 1). For $\beta \neq \beta_0$, using Lemma 3.1 in Phillips (1989), we decompose $V_\beta(r)$ as

$$V_\beta(r) = \frac{\beta'\Sigma_v\beta_0}{\beta_0'\Sigma_v\beta_0} V_0(r) + \Sigma_{\beta\beta.\beta_0}^{1/2} W(r), \tag{15}$$

where $W(r)$ is a standard Wiener process in the probability space $(\Omega^*, \mathcal{F}^*, \mathbb{P}^*)$ that is independent of $V_0(r)$ and $\Sigma_{\beta\beta.\beta_0} = \beta'\Sigma_v\beta - \frac{(\beta'\Sigma_v\beta_0)^2}{\beta_0'\Sigma_v\beta_0} > 0$.

Case 1: When $\beta'\Sigma_v\beta_0 = 0$.
In this case, $V_0(r)$ is independent of $V_\beta(r)$. Then, by Lemma 10, we have

$$Q(\beta) = \int_0^1 sgn\left(V_\beta(r) V_0(r)\right) dr < 1,$$

and we have the required result. ■

Case II: When $\beta'\Sigma_v\beta_0 < 0$.
Using the decomposition in (15), we may write

$$\begin{aligned} Q(\beta) &= \int_0^1 sgn\left(V_\beta(r) V_0(r)\right) dr \\ &= -\int_0^1 sgn\left\{V_0(r)^2 + \gamma_\beta V_0(r) W(r)\right\} dr, \end{aligned}$$

where $\gamma_\beta = \left(\frac{\beta'\Sigma_v\beta_0}{\beta_0'\Sigma_v\beta_0}\right)^{-1} \Sigma_{\beta\beta.\beta_0}^{1/2} < 0$, and the last equality holds because $\beta'\Sigma_v\beta_0 < 0$. To have the required result, it is enough to show that

$$\int_0^1 sgn\left\{V_0(r)^2 + \gamma_\beta V_0(r) W(r)\right\} dr > -1 \tag{16}$$

almost surely.

For this, notice that

$$sgn\left\{V_0(r)^2 + \gamma_\beta V_0(r) W(r)\right\} = -1$$

if and only if,

$$
\begin{aligned}
0 &< V_0(r) < -\gamma_\beta W(r) \;\text{ when } W(r) > 0 \\
-\gamma_\beta W(r) &< V_0(r) < 0 \;\text{ when } W(r) < 0.
\end{aligned}
$$

Then, inequality (16) follows if we show that

$$
\begin{aligned}
& \int_0^1 1\left\{0 < V_0(r) < -\gamma_\beta W(r)\right\} 1\{W(r) > 0\}\, dr \\
& + \int_0^1 1\left\{-\gamma_\beta W(r) < V_0(r) < 0\right\} 1\{W(r) < 0\}\, dr \\
< \;& 1 \text{ with probability one.}
\end{aligned}
$$

The above inequality follows because

$$
\begin{aligned}
& \int_0^1 1\left\{0 < V_0(r) < -\gamma_\beta W(r)\right\} 1\{W(r) > 0\}\, dr \\
& + \int_0^1 1\left\{-\gamma_\beta W(r) < V_0(r) < 0\right\} 1\{W(r) < 0\}\, dr \\
\leq \;& \int_0^1 1\{0 < V_0(r)\} 1\{W(r) > 0\}\, dr + \int_0^1 1\{0 > V_0(r)\} 1\{W(r) < 0\}\, dr \\
< \;& 1 \text{ almost surely,}
\end{aligned}
$$

where the last inequality holds by Lemma 10, and we have the required result. ■

Case III: When $\beta' \Sigma_v \beta_0 > 0$.
Again, using the decomposition of $V_\beta(r)$ in (15), we write

$$
\begin{aligned}
Q(\beta) &= \int_0^1 sgn\left(V_\beta(r) V_0(r)\right) dr \\
&= \int_0^1 sgn\left(V_0(r)\left(\frac{1}{\gamma_\beta} V_0(r) + W(r)\right)\right) dr,
\end{aligned}
$$

where $\gamma_\beta = \left(\frac{\beta' \Sigma_v \beta_0}{\beta_0' \Sigma_v \beta_0}\right)^{-1} \Sigma_{\beta\beta.\beta_0}^{1/2} > 0$. Observe that

$$
Q(\beta) = 1
$$

if and only if, for all $r \in [0,1]$,

$$
W(r) \geq -\frac{1}{\gamma_\beta} V_0(r) \;\; \Leftrightarrow \;\; V_0(r) \geq 0. \tag{17}
$$

Now we show that (17) does not occur with probability one.

For this, first notice that the set

$$
Z = \left\{ r \in [0,1] : \; W(r) = -\frac{1}{\gamma_\beta} V_0(r) \right\}
$$

is almost surely a non-empty closed set without an isolation point and has zero Lebesgue measure. (Apply Proposition 3.12 on page 109 of Revuz and Yor to Brownian motion $W(r) + \frac{1}{\gamma_\beta} V_0(r)$.)

Choose $r^0 \in Z \cap (0,1)$. Without loss of generality, we assume that $V_0\left(r^0\right) > 0$ because the set $\{r \in [0,1] : V_0(r) = 0\}$ has zero Lebesgue measure. Since the sample path of $V_0(r)$ is continuous, we can choose an interval around r^0 with length 2ε such that $V_0\left(r^0\right) > 0$ for all $r \in \left(r^0 - \varepsilon, r^0 + \varepsilon\right)$.

Now suppose that (17) is true with probability one. Then, $W(r) + \frac{1}{\gamma_\beta} V_0(r) \geq 0$ for $r \in \left(r^0 - \varepsilon, r^0 + \varepsilon\right)$ with probability one. This cannot occur with probability one because $W\left(r^0\right) + \frac{1}{\gamma_\beta} V_0\left(r^0\right) = 0$ by definition and, from Lemma 11, $W(r) + \frac{1}{\gamma_\beta} V_0(r)$ changes sign infinitely often in the interval $\left(r^0 - \varepsilon, r^0 + \varepsilon\right)$ with probability one. Therefore, (17) does not occur with probability one and so $Q(\beta) < 1$ almost surely. ■

Part (b). Notice that for any sequence $\beta_n \to \beta$, it follows that

$$
\begin{aligned}
& \left|Q\left(\beta_n\right) - Q(\beta)\right| \\
= & \left|\int_0^1 sgn\left(\beta_0' V(r)\right) sgn\left(\beta_n' V(r)\right) dr - \int_0^1 sgn\left(\beta_0' V(r)\right) sgn\left(\beta' V(r)\right) dr\right| \\
\leq & \; 2 \int_0^1 \left|sgn\left(\beta_n' V(r)\right) - sgn\left(\beta' V(r)\right)\right| dr \\
\to & \; 0,
\end{aligned}
$$

where the last convergence holds by the convergence result of (12) in the proof of Lemma 12. ■

Proof of Lemma 5

Notice from the uniform convergence of $\left(V_n(r), Z_n(r)\right) \overset{a.s.}{\to} \left(V(r), \mathbf{0}\right)$, we have

$$
\sup_{r \in [0,1]} \left\|\left(V_n(r)', Z_n(r)'\right)\right\| \overset{a.s.}{\to} \sup_{r \in [0,1]} \left\|V(r)\right\|.
$$

Combining this with Lemma 4, for any fixed $M > 0$, we may deduce the finite dimensional convergence of $T_n(\beta, \gamma, M)$ to $T(\beta, M)$. Also, the truncated process $T_n(\beta, \gamma, M)$ is asymptotically uniformly equicontinuous in (β, γ) by Lemma 14. Using the conventional arguments with the assumption that the parameter set is compact, the required result follows. ■

References

[1] Canavagh, C.L. (1987): Limiting Behavior of Estimators Defined by Optimization, *Mimeo*, Harvard University.

[2] Guerre, E. and H.R. Moon (2002): A Note on the Nonstationary Binary Choice Logit Model, *Economics Letters, 76, 267–271.*

[3] Horowitz, J. (1992): A Smoothed Maximum Score Estimator for the Binary Response Model, *Econometrica*, 60, 505–531.

[4] Horowitz, J. (1998): *Semiparametric Methods in Econometrics*, Lecture Notes in Statistics, Vol. 131, Springer, New York and Heidelberg.

[5] Karatzas, I. and S. Shreve (1991): *Brownian Motion and Stochastic Calculus*, 2^{nd} Ed., New York, Springer-Verlag.

[6] Kim, J. and D. Pollard (1990): Cube Root Asymptotics, *Annals of Statistics*, 18, 191–219.

[7] Manski, C. (1975): Maximum Score Estimation of the Stochastic Utility Model of Choice, *Journal of Econometrics*, 3, 205-228.

[8] Manski, C. (1985): Semiparametric Analysis of Discrete Response, *Journal of Econometrics*, 27, 313–333.

[9] Manski, C. (1988): Identification of Binary Response Models, Journal of American Statistical Association, 83, 729–738.

[10] Park, J. and P.C.B. Phillips (2000): Nonstationary Binary Choice, *Econometrica*, 68, 1249–1280.

[11] Park, J. and P.C.B. Phillips (2001): Nonlinear Regressions with Integrated Time Series, *Econometrica*, 69, 117–162.

[12] Phillips, P.C.B. (1989): Partially Identified Econometric Models, *Econometric Theory*, 5, 181–240.

[13] Phillips, P.C.B. and V. Solo (1992): Asymptotics for Linear Processes, *Annals of Statistics*, 20, 971–1001.

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

[15] Revuz, Y. and M. Yor (1999): *Continuous Martingales and Brownian Motion*, 3^{rd} Ed., New York, Springer-Verlag.