Dynamic time series binary choice*

Robert de Jong† Tiemen Woutersen ‡

January 30, 2004

Abstract

This paper considers dynamic time series binary choice models. It shows in a time series setting the validity of the dynamic probit likelihood procedure when lags of the dependent binary variable are used as regressors, and it establishes the asymptotic validity of Horowitz' smoothed maximum score estimation of dynamic binary choice models with lags of the dependent variable as regressors. The latent error is explicitly allowed to be correlated. It turns out that no long-run variance estimator is needed for the validity of the smoothed maximum score procedure in the dynamic time series framework. One novel aspect of this paper is a proof that weak dependence properties hold for dynamic binary choice models with correlated errors.

1 Introduction

For a dynamic linear time series model

$$y_n = \sum_{j=1}^{p} \rho_j y_{n-j} + \gamma' x_n + u_n, \tag{1}$$

$n = 1, \ldots, N$, it is well-known that a sufficient condition for consistency as $N \to \infty$ of the least squares estimator is that $E(u_n|y_{n-1}, \ldots, y_{n-p}, x_n) = 0$, and that even if u_n is weakly

*We thank James Davidson, Lung-Fei Lee, Benedikt Pötscher, Jim Stock and Jeff Wooldridge for helpful discussions.

†Department of Economics, Ohio State University, email dejong@econ.ohio-state.edu.

‡Department of Economics, University of Western Ontario and MIT, email twouters@uwo.ca.

dependent, consistency can be proven as long as this condition holds, without assumption of normality on u_n. In this paper, we analyze maximum likelihood estimation of the dynamic probit model of order p, and maximum score estimation of dynamic binary choice models of order p, and we explicitly allow the error to be correlated. We define the dynamic binary choice model of order p as

$$y_n = I(\sum_{j=1}^{p} \rho_j y_{n-j} + \gamma' x_n + u_n > 0). \tag{2}$$

For conditional maximum likelihood estimation of the dynamic probit model, the key condition that is needed will turn out to be

$$E(y_n|x_n, y_{n-1}, y_{n-2}, \ldots) = \Phi(\sum_{j=1}^{p} \rho_j y_{n-j} + \gamma' x_n), \tag{3}$$

while in the smoothed maximum score setting, we will need the condition

$$\text{Median}(u_n|y_{n-1}, \ldots, y_{n-p}, x_n) = 0. \tag{4}$$

Therefore, this paper attempts to analyze the dynamic time series binary choice model at a level of generality that is comparable to the level of generality at which linear dynamic time series models can be analyzed.

Manski (1975) uses the sign function to develop the first semiparametric estimator for the binary choice model. Coslett (1983) and Ichimura (1993) derive alternative estimators for the binary choice model. Imbens (1992) and Matztkin (1992) also develop estimators for the semiparametric binary choice model. Finally, in his seminal paper, Horowitz (1992) smooths the sign function of Manski (1975, 1985) and derives an estimator that is asymptotically normally distributed. However, all these estimators assume that one has a random sample. Thus, none of these estimators allows for lagged dependent explanatory variables. In this paper we consider the binary choice model in a time series setting and we allow for lagged dependent variables as explanatory variables. For the semiparametric case, we only impose a median assumption. Thus, we allow the variance (and other moments of the error distribution) to depend on lagged error terms, lagged dependent variables as well as regressors. Moreover, the median assumption allows for heterogeneity that is caused by random coefficients, e.g. a data generating process whose parameters are random and symmetrically distributed around $(\rho, \gamma')'$.

Ruud (1981) and Poirier and Ruud (1988) have considered the probit model with correlated errors. Robinson (1982) considered the tobit model with correlated errors. An example of an empirical paper that focuses on the dynamic probit methodology is Eichengreen, Watson,

and Grossman (1985). However, no formal stationarity properties for dynamic probit models are derived in these papers, nor anywhere else in the literature as far as the authors are aware. Potential applications include finance models concerning the likelihood of a financial transaction in a given time period as well as models concerning labor market participation decisions in which the relative importance of wealth versus welfare effects are studied.
The setup of this paper is as follows. In Section 2, the weak dependence properties of y_n are analyzed. Section 3 of this paper will analyze the dynamic probit procedure when lagged values of y_n have been included among the regressors and normality of u_n is assumed. In Section 4, we consider consistency of the smoothed maximum score estimator of the dynamic time series binary choice model. The smoothed maximum score estimator was first suggested in Horowitz (1992). Section 5 establishes asymptotic normality of the smoothed maximum score estimator[1].

2 Properties of the dynamic time series binary choice model

A big analytical complication in the analysis below is to show that y_n satisfies the appropriate "fading memory" property when generated through a general dynamic binary choice model with regressors and possibly correlated errors. For the analysis of the smoothed maximum score estimator, this "fading memory" property that is proven for y_n needs to be strong enough to allow a proof of an equivalent of the Hoeffding inequality, and in addition, it needs to allow for a proof of a central limit theorem (CLT) for a function of y_n and x_n that depends on N in a situation where no martingale difference CLT can be applied. For a proof of validity of the dynamic probit model, the "fading memory" property only needs to support laws of large numbers and uniform laws of large numbers.
The "fading memory" property that we will prove for y_n is that of *near epoch dependence*. The idea of the proof is similar to that of proofs for showing fading memory properties of processes y_n of the form

$$y_n = f(y_{n-1}) + \varepsilon_n, \tag{5}$$

where $f(.)$ is such that $|f(x) - f(y)| \leq L|x - y|$ for some $L < 1$. Functions $f(.)$ satisfying this condition are called contraction mappings. Such proofs can be found in Bierens (1981) and Pötscher and Prucha (1997), for example. Pötscher and Prucha (1997, Section 6.4) contains a thorough discussion of these types of results, but the approach in the proof of

[1] In addition, some corrections to Horowitz' proof of the validity of the smoothed maximum score procedure are provided.

this paper is somewhat different from the approaches discussed there. The differences are that the $f(.)$ function in the dynamic binary choice case is not continuous, depends on ε_n, is not strictly less than 1, and depends on more than one lagged value of y_n. These problems are essentially solved by smoothing the response function by the expectations operator, by using the fact that y_n is a binary random variable, and by the use of the appropriate metric on the arguments of the $f(.)$ function.
Near epoch dependence of random variables y_n on a base process of random variables v_n is defined as follows:

Definition 1 *Random variables y_n are called* near epoch dependent on v_n *if*

$$\sup_{n\in\mathbb{Z}} E|y_n - E(y_n|v_{n-m}, v_{n-m+1}, \ldots, v_n)|^2 = \nu(m)^2 \to 0 \quad as \quad m \to \infty. \tag{6}$$

The base process v_n needs to satisfy a condition such as strong or uniform mixing or independence. For the definitions of strong (α-) and uniform (ϕ-) mixing see e.g. Gallant and White (1988, p. 23) or Pötscher and Prucha (1997, p. 46). The near epoch dependence condition then functions as a device that allows approximation of y_n by a function of finitely many mixing or independent random variables v_n. Note also that for strictly stationary (y_n, v_n), the "sup" in the above definition can be removed, because in that case

$$E|y_n - E(y_n|v_{n-m}, v_{n-m+1}, \ldots, v_n)|^2 \tag{7}$$

does not depend on n. The reader is referred to Gallant and White (1988) for a detailed account of the near epoch dependence condition. See also Pötscher and Prucha (1997) for a more up-to-date treatment of dependence concepts such as near epoch dependence.
For establishing near epoch dependence of y_n, we have the following result. Define S as the set of all 2^p possible p -vectors s such that its elements s_i are 0 or 1, and define

$$\Phi = \{\phi : \phi = \sum_{i=1}^{p} \rho_i s_i, s \in S\}. \tag{8}$$

Let ϕ_{min} denote the smallest element of Φ, and let ϕ_{max} denote the largest element.

Theorem 1 *Assume that y_n is generated as $y_n = I(\sum_{j=1}^{p} \rho_j y_{n-j} + \eta_n > 0)$. Let η_n be strong mixing and strictly stationary. Assume that there is some $\delta > 0$ for which there exists a positive integer K such that*

$$P(\phi_{max} + \max_{i=1,\ldots,p} \eta_{n-i} > 0 | y_{n-p-K}, y_{n-p-K-1}, \ldots)$$

$$-P(\phi_{min} + \min_{i=1,\ldots,p} \eta_{n-i} > 0 | y_{n-p-K}, y_{n-p-K-1}, \ldots) < 1 - \delta \qquad \text{almost surely,} \tag{9}$$

and the near epoch dependence sequence $v(.)$ satisfies $v(m) \leq C_1 \exp(-C_2 m)$, for positive constancts C_1 and C_2. Then (i) y_n is near epoch dependent on η_n; (ii) (y_n, η_n) is strictly stationary.

Note that if $\eta_n = \gamma' x_n + u_n$ for strong mixing and strictly stationary $(x_n', u_n)'$, clearly η_n is mixing as well. This observation will be used below.
The proof of the above result is substantially easier for the case where η_n is i.i.d., only one lagged y_n is used as regressor and no other regressors are included. In that case, we can write

$$y_n = y_{n-1} I(\rho_1 + \eta_n > 0) + (1 - y_{n-1}) I(\eta_n > 0), \tag{10}$$

implying that

$$\begin{aligned}
\nu_m &\equiv \sup_{n \in \mathbb{Z}} E|y_n - E(y_n|\eta_{n-m}, \ldots, \eta_n)|^2 \\
&= \sup_{n \in \mathbb{Z}} E|(I(\rho_1 + \eta_n > 0) - I(\eta_n > 0))(y_{n-1} - E(y_{n-1}|\eta_{n-m}, \ldots, \eta_{n-1})|^2 \\
&= |P(\rho_1 + \eta_n > 0) - P(\eta_n > 0)| \sup_{n \in \mathbb{Z}} E|y_{n-1} - E(y_{n-1}|\eta_{(n-1)-(m-1)}, \ldots, \eta_{n-1})|^2 \\
&= |P(\rho_1 + \eta_n > 0) - P(\eta_n > 0)| \nu_{m-1},
\end{aligned} \tag{11}$$

which implies that the $\nu(m)$ sequence decays exponentially under the condition of Equation (9). The proof of Theorem 1 should be viewed as an extension to the above reasoning.
The fact that y_n is a 0/1-valued near epoch dependent random variable can now be exploited to show that (y_n, x_n) is also strong mixing. Note that this is an observation that apparently has not been made in the literature before. The result is as follows:

Theorem 2 *Suppose that $y_n = f(v_n, v_{n-1}, \ldots)$ is a sequence of 0/1-valued random variables that is near epoch dependent on v_n with near epoch dependence coefficients $\nu(m)$, where $v_n = (x_n', u_n)'$ is strong mixing with mixing coefficients $\alpha(m)$. Then $(y_n, x_n')'$ is strong mixing with strong mixing coefficients $C(\nu(m) + \alpha(m))$ for some $C > 0$.*

The mixing property of (y_n, x_n) will be used in the proofs for consistency and asymptotic normality of the next sections.

3 The dynamic probit model

This section examines the behavior of the dynamic probit model estimator that results from including lagged y_n among the regressors. Let $\beta = (\rho', \gamma')'$ denote the true parameter and let $b = (r', c')'$, $\rho, r \in \mathbb{R}^p$ and $\gamma, c \in \mathbb{R}^q$, and let R and Γ denote the parameter spaces for r and c respectively, and let $B = R \times \Gamma$. We assume normality of the errors so that the likelihood conditional on $y_1, ..., y_p$ has the following form,

$$L_N(b) = (N-p)^{-1} \sum_{n=p+1}^{N} l_n(b)$$

$$= (N-p)^{-1} \sum_{n=p+1}^{N} [y_n \log(\Phi(\sum_{j=1}^{p} r_j y_{n-j} + c' x_n)) + (1-y_n) \log(1 - \Phi(\sum_{j=1}^{p} r_j y_{n-j} + c' x_n))]. \quad (12)$$

Given the result of Theorem 2, it is now straightforward to find standard conditions under which the maximum likelihood estimator b_N^{ML} is consistent.

Assumption A

1. x_n is a sequence of strictly stationary strong mixing random variables with α-mixing numbers $\alpha(m)$, where $x_n \in \mathbb{R}^q$ for $q \geq 0$ and $\gamma \in \mathbb{R}^q$, and the second absolute moment of x_n exits. The distribution of x_n is not contained in any linear subspace of $\mathbb{R}^q$.

2. u_n is independently normally distributed with mean zero and variance one.

3. $y_n = I(\sum_{i=1}^{p} \rho_i y_{n-i} + \gamma' x_n + u_n > 0)$.

4. $\beta \in B$, where B is a compact subset of $\mathbb{R}^{p+q}$.

Theorem 3 *Under Assumption A, $b_N^{ML} \xrightarrow{p} \beta$. If in addition the strong mixing coefficients satisfy $\alpha(m) \leq Cm^{-\eta}$ for positive constants C and η and $E|l_n(b)|^{1+\delta} < \infty$ for some $\delta > 0$ and all $b \in B$, then $b_N^{ML} \xrightarrow{as} \beta$.*

Let $I = -E(\partial/\partial b)(\partial/\partial b')l_n(\beta)$. For asymptotic normality, we need an additional assumption.

Assumption B

1. β is in the interior of B.

Theorem 4 *Under Assumptions A and B,* $N^{1/2}(b_N^{ML} - \beta) \xrightarrow{d} N(0, I^{-1})$.

Under the above Assumptions A and B, it also follows that the usual estimators of I, using either the outer product or Hessian approach, will both be weakly consistent for I.
Note that given the weak dependence property of Theorem 2, it is also possible to set forth conditions such that for weakly dependent u_n with arbitrary distribution, $N^{1/2}(b_N^{ML} - \beta^*) \xrightarrow{d} N(0, J)$ for some matrix J and a β^* that uniquely minimizes the objective function. Here of course β^* does not necessarily equal the true parameter value β. However, in order to show that the probit objective function is uniquely minimized at β, we need that a first order condition of the type

$$E(y_n - \Phi(\sum_{i=1}^{p} \rho_i y_{n-i} + \gamma' x_n))w(y_{n-1}, \ldots, y_{n-p}, x_n) = 0 \tag{13}$$

holds for some function $w(., \ldots, .)$. This condition is implied by

$$E(y_n | y_{n-1}, \ldots, x_n) = \Phi(\sum_{i=1}^{p} \rho_i y_{n-i} + \gamma' x_n), \tag{14}$$

and the latter condition is equivalent to assuming that u_n is i.i.d. and standard normal if lagged values of y_n are included.

4 Consistency of the smoothed maximum score estimator

The smoothed maximum score estimator is defined as $\text{argmin}_{b \in B} S_N(b, \sigma_N)$, where

$$S_N(b, \sigma_N) = (N-p)^{-1} \sum_{n=p+1}^{N} (2I(y_n = 1) - 1)K((\sum_{j=1}^{p} r_j y_{n-j} + c' x_n)/\sigma_N) \tag{15}$$

and σ_N is a bandwidth-type sequence such that $\sigma_N \to 0$ as $N \to \infty$, where $K(.)$ is a function such that $K(-\infty) = 0$ and $K(\infty) = 1$. This objective function is a smoothed version of the maximum score objective function

$$S_N^*(b) = (N-p)^{-1} \sum_{n=p+1}^{N} (2I(y_n = 1) - 1)I(\sum_{j=1}^{p} r_j y_{n-j} + c' x_n \geq 0). \tag{16}$$

In addition, let $S(b) = ES_N^*(b)$. This notation is justified because we will use conditions under which (y_n, x_n) will be proven to be strictly stationary. See Manski (1985) and Pollard (1984) for more information and results regarding the maximum score estimator. While Horowitz' maximum score estimator can reach an asymptotic efficiency bound (see Horowitz (1992)), Pollard (1984) showed that the maximum score estimator in general is consistent of order $N^{-1/3}$.
The following five assumptions are needed for the proof of our consistency result:

Assumption 1 *$v_n = (x_n', u_n)'$ is a sequence of strictly stationary strong mixing random variables with α-mixing numbers $\alpha(m)$, where $x_n \in \mathbb{R}^q$ for $q \geq 1$ and $\gamma \in \mathbb{R}^q$, and*

$$y_n = I(\sum_{i=1}^{p} \rho_i y_{n-i} + \gamma' x_n + u_n > 0). \tag{17}$$

Note that by Theorem 1 and the discussion following that theorem, $(y_n, x_n)'$ is strictly stationary. This justifies the formulation of the assumptions below in their current forms. Define $\tilde{x}_n = (y_{n-1}, \ldots, y_{n-p}, x_{n2}, \ldots, x_{nq})$.

Assumption 2 *The support of the distribution of x_n is not contained in any proper linear subspace of $\mathbb{R}^q$. (b) $\delta < P(y_n = 1|x_{n1}, \tilde{x}_n) < 1 - \delta$ almost surely for some $\delta > 0$. (c) $\gamma_1 \neq 0$, and for almost every $\tilde{x}_n$, the distribution of x_{n1} conditional on $\tilde{x}_n$ has everywhere positive density with respect to Lebesgue measure.*

Assumption 3 $\text{Median}(u_n|x_n, y_{n-1}, \ldots, y_{n-p}) = 0$ almost surely.

Assumption 4 *$|\gamma_1| = 1$, and $\tilde{\beta} = (\rho_1, \ldots, \rho_p, \gamma_2, \ldots, \gamma_q)$ is contained in a compact subset $\tilde{B}$ of $\mathbb{R}^{p+q-1}$.*

Assumption 5 *For ϕ_{max} and ϕ_{min} as defined before, for some $\delta > 0$ there exists a positive integer K such that*

$$\begin{aligned} &P(\phi_{max} + \max_{i=1,\ldots,p}(\gamma' x_{n-i} + u_{n-i}) > 0 | y_{n-p-K}, y_{n-p-K-1}, \ldots) \\ &\quad - P(\phi_{min} + \min_{i=1,\ldots,p}(\gamma' x_{n-i} + u_{n-i}) > 0 | y_{n-p-K}, y_{n-p-K-1}, \ldots) < 1 - \delta. \end{aligned} \tag{18}$$

For normalization, we need to set $|c_1| = 1$ when calculating the minimizer of $S_N(b, \sigma_N)$. Therefore, the estimator b_N needs to be defined as

$$b_N = \text{argmin}_{b:|c_1|=1} S_N(b, \sigma_N). \tag{19}$$

The following result shows the consistency of b_N:

Theorem 5 *Under Assumptions 1,2,3,4 and 5, $b_N \xrightarrow{p} \beta$. If in addition the strong mixing coefficients satisfy $\alpha(m) \leq Cm^{-\eta}$ for positive constants C and η, then $b_N \xrightarrow{as} \beta$.*

5 Asymptotic normality

Define, analogously to Horowitz (1992), $\tilde{b} = (r_1, \ldots, r_p, c_2, \ldots, c_q)$, and let

$$T_N(b, \sigma_N) = \partial S_N(b, \sigma_N)/\partial \tilde{b}, \tag{20}$$

$$Q_N(b, \sigma_N) = \partial^2 S_N(b, \sigma_N)/\partial \tilde{b} \partial \tilde{b}'. \tag{21}$$

Also, define

$$z_n = \sum_{j=1}^{p} \rho_j y_{n-j} + \gamma' x_n, \tag{22}$$

and let $p(z_n|\tilde{x}_n)$ denote the density of z_n given $\tilde{x}_n$, let $P(.)$ denote the distribution of $\tilde{x}_n$, let $F(.|z_n, \tilde{x}_n)$ denote the cumulative distribution of u_n conditional on z_n and $\tilde{x}_n$. For each positive integer i, define

$$F^{(i)}(-z, x, \tilde{x}) = \partial^i F(-z|z, \tilde{x})/\partial z^i \tag{23}$$

and

$$\alpha_A = \int_{-\infty}^{\infty} v^h K'(v) dv \tag{24}$$

$$\alpha_D = \int_{-\infty}^{\infty} K'(v)^2 dv. \tag{25}$$

Also analogously to Horowitz (1992), define

$$A = -2\alpha_A \sum_{i=1}^{h} \{[i!(h-i)!]^{-1} E[F^{(i)}(0, 0, \tilde{x}_n) p^{(h-i)}(0|\tilde{x}_n)\tilde{x}_n]\}, \tag{26}$$

$$D = \alpha_D E[\tilde{x}_n \tilde{x}_n' p(0|\tilde{x}_n)], \tag{27}$$

$$Q = 2E[\tilde{x}_n \tilde{x}_n' F^{(1)}(0|0, \tilde{x}_n) p(0|\tilde{x}_n)]. \tag{28}$$

We need to strengthen the fading memory conditions of Assumption 1 in order to establish asymptotic normality:

Assumption 1' *$v_n = (x_n', u_n)'$ is a sequence of strictly stationary strong mixing random*

variables with α-mixing numbers $\alpha(m)$ such that $\alpha(m) \leq Cm^{-(2r-2)/(r-2)-\eta}$ for some $\eta > 0$, where $x_n \in \mathbb{R}^q$ for $q \geq 1$ and $\gamma \in \mathbb{R}^q$, and

$$y_n = I(\sum_{i=1}^{p} \rho_i y_{n-i} + \gamma' x_n + u_n > 0). \tag{29}$$

The following assumption is the analogue of Horowitz' Assumption 5, which is the assumption below for $r = 4$. It appears that Horowitz' truncation argument is in error, but that his argument is correct for bounded data. This explains the presence here of a condition here that is stronger than that of Horowitz.

Assumption 6 *For all vectors γ such that $|\gamma| = 1$, $E|\gamma'\tilde{x}|^r < \infty$ for some $r > 4$.*

The assumption below is needed in lieu of Horowitz' Assumption 6.

Assumption 7 *For some sequence $m_N \geq 1$,*

$$\sigma_N^{-3(p+q-1)}\sigma_N^{-2}N^{1/r}\alpha(m_N) + \sigma_N^{-2(p+q-1)/\beta}N^{2/r}\alpha(m_N)$$
$$+|\log(Nm_N)|(N^{1-4/r}\sigma_N^4 m_N^{-2})^{-1} \to 0 \quad \text{as} \quad N \to \infty. \tag{30}$$

For the case of independent (x_n, u_n), $\alpha(m) = 0$ for $m \geq 1$, and we can set $m_N = 1$ for that case. The first condition then becomes

$$(\log(N))(N^{1-2/r}\sigma_N^4)^{-1} \to 0 \quad \text{as} \quad N \to \infty, \tag{31}$$

implying that for bounded data, we can set $r = \infty$ and obtain Horowitz' condition

$$(\log(N))(N\sigma_N^4)^{-1} \to 0 \quad \text{as} \quad N \to \infty. \tag{32}$$

The following assumptions are identical to Horowitz' Assumptions 7-11:

Assumption 8 *(a) $K(.)$ is twice differentiable everywhere, $|K(.)|$ and $K''(.)$ are uniformly bounded, and each of the following integrals over $(-\infty, \infty)$ is finite: $\int [K'(v)]^4 dv$, $\int [K''(v)]^2 dv$, $\int |v^2 K''(v)| dv$. (b) For some integer $h \geq 2$ and each integer i $(1 \leq i \leq h)$, $\int |v^i K'(v)| dv < \infty$, and*

$$\int_{-\infty}^{\infty} v^i K'(v) dv = \begin{cases} 0 & \text{if } i < h, \\ d & (\text{nonzero}) \text{ if } i{=}h. \end{cases} \tag{33}$$

(c) For any integer i between 0 and h, any $\eta > 0$, and any sequence $\{\sigma_N\}$ converging to 0,

$$\lim_{N\to\infty} \sigma_N^{i-h} \int_{|\sigma_N v|>\eta} |v^i K'(v)| dv = 0 \tag{34}$$

and

$$\lim_{N\to\infty} \sigma_N \int_{|\sigma_N v|>\eta} |K''(v)| dv = 0. \tag{35}$$

Assumption 9 *For each integer i such that $1 \leq i \leq h-1$, all z in a neighborhood of 0, almost every $\tilde{x}_n$, and some $M < \infty$, $p^{(i)}(z_n|\tilde{x}_n)$ exists and is a continuous function of z_n satisfying $|p^{(i)}(z_n|\tilde{x}_n)| < M$. In addition, $p(z_n|\tilde{x}_n) < M$ for all z and almost every $\tilde{x}$.*

Assumption 10 *For each integer i such that $1 \leq i \leq h$, all z_n in a neighborhood of 0, almost every $\tilde{x}_n$, and some $M < \infty$, $F^{(i)}(-z_n, z_n, \tilde{x}_n)$ exists and is a continuous function of z_n satisfying $|F^{(i)}(-z_n, z_n, \tilde{x}_n)| < M$.*

Assumption 11 *$\tilde{\beta}$ is an interior point of $\tilde{B}$.*

Assumption 12 *The matrix Q is negative definite.*

In addition to the above equivalents to Horowitz' assumptions, we will also need the following two assumptions. The first assumption is needed to assure proper behavior of covariance terms.

Assumption 13 *The conditional joint density $p(z_n, z_{n-j}|x_n, x_{n-j})$ exists for all $j \geq 1$ and is continuous at $(z_n, z_{n-j}) = (0, 0)$ for all $j \geq 1$.*

The next condition on $K''(.)$ is needed to formally show a uniform law of large numbers for the second derivative of the objective function.

Assumption 14 *$K''(.)$ satisfies, for some $\beta \in (0, 1]$ and all $x, y \in \mathbb{R}$,*

$$|K''(x) - K''(y)| \leq L|x - y|^\beta. \tag{36}$$

To prove asymptotic normality, we need an inequality in the spirit of Hoefding's inequality but for random variables with unbounded support. We derive such an inequality in the Appendix as Lemma 10. The inequality of Lemma 10 also allows for martingale difference sequences so that it covers both the random sample case of Horowitz (1992) as well as the dynamic case.
Our asymptotic normality result now is the following. This result, of course, is nearly identical to Horowitz' in the non-dynamic cross-section case.

Theorem 6 *Let Assumptions 1' and Assumptions 2-14 hold for some $h \geq 2$. Then*

1. *If $N\sigma_N^{2h+1} \to \infty$ as $N \to \infty$, $\sigma_N^{-h}(\tilde{b}_N - \tilde{\beta}) \xrightarrow{p} -Q^{-1}A$.*
2. *If $N\sigma_N^{2h+1}$ has a finite limit λ as $N \to \infty$,*

$$(N\sigma_N)^{1/2}(\tilde{b}_N - \tilde{\beta}) \xrightarrow{d} N(-\lambda^{1/2}Q^{-1}A, Q^{-1}DQ^{-1}). \tag{37}$$

In order to estimate the matrices A, D and Q, we need an additional result, the analogue of Horowitz' (1992) Theorem 3.

Theorem 7 *Let b_N be a consistent smoothed maximum score estimator based on σ_N such that $\sigma_N = O(n^{-1/(2h+1)})$. For $b \in \{-1, 1\} \times \tilde{B}$, define*

$$t_n(b, \sigma) = (2I(y_n = 1) - 1)(\tilde{x}_n/\sigma_N)K'(b'x_n/\sigma). \tag{38}$$

Let σ_N^* be such that $\sigma_N^* = O(N^{-\delta/(2h+1)})$, where $0 < \delta < 1$. Then: (a) $\hat{A}_N \equiv (\sigma_N^*)^{-h}T_N(b_N, \sigma_N^*)$ converges in probability to A; (b) the matrix

$$\hat{D}_N \equiv (\sigma_N/N)\sum_{n=1}^{N} t_n(b_N, \sigma_N)t_n(b_N, \sigma_N)' \tag{39}$$

converges in probability to D; (c) $Q_N(b_N, \sigma_N)$ converges in probability to Q.

References

Andrews, D.W.K. (1987), Consistency in nonlinear econometric models: a generic uniform law of large numbers, *Econometrica* 55, 1465-1471.

Andrews, D.W.K. (1988), Laws of large numbers for dependent non-identically distributed random variables, *Econometric Theory* 4, 458-467.

Azuma, K. (1967), Weighted sums of certain dependent random variables, *Tokohu Mathematical Journal* 19, 357-367.

Bierens, H.J. (1981) *Robust methods and asymptotic theory in nonlinear econometrics*. New York: Springer-Verlag.

Bierens, H. J. (2004) *Introduction to the mathematical and statistical foundations of Econometrics*. Cambridge University Press, forthcoming, available at

http://econ.la.psu.edu/~hbierens/CHAPTER7.PDF.

Cosslett, S. R. (1983), Distribution-free maximum likelihood estimator of the binary choice model, *Econometrica* 51, 765-782.

Davidson, J. (1994) *Stochastic limit theory.* Oxford: Oxford University Press.

de Jong, R.M. (1995), Laws of large numbers for dependent heterogeneous processes, *Econometric Theory* 11, 347-358.

de Jong, R.M. (1997), Central limit theorems for dependent heterogeneous random variables, *Econometric Theory* 13, 353-367.

Eichengreen, B., Watson, M. and R. Grossman (1985), Bank rate policy under the interwar gold standard: a dynamic probit model, *Economic Journal* 95, 725-745.

Gallant, A.R. and H. White (1988) *A unified theory of estimation and inference for nonlinear dynamic models.* New York: Basil Blackwell.

Hall, P and C. Heyde (1980) *Martingale limit theory and its application.* New York: Academic Press.

Horowitz, J. (1992), A smoothed maximum score estimator for the binary response model, *Econometrica* 60, 505-531.

Ichimura, I. (1993), Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, *Journal of Econometrics* 58, 71-120.

Imbens, G. W. (1992), An efficient method of moment estimator for discrete choice models with choice-based sampling, *Econometrica* 60, 1187-1214.

Manski, C.F. (1985), Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator, *Journal of Econometrics* 27, 313-333.

Manski, C.F. (1985), Semiparametric analysis of discrete response: asymptotic properties of the maximum score estimator, *Journal of Econometrics* 27, 313-333.

Matzkin, R.L. (1992), Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models, *Econometrica* 60, 239-270.

McLeish, D. L. (1974), Dependent central limit theorems and invariance principles, *Annals of Probability* 2, 620-628.

Newey, W. K., and D. McFadden (1994), Large sample estimation and hypothesis testing, in *Handbook of Econometrics,* Vol. 4, ed. by R. F. Engle and D. MacFadden. Amsterdam:

North-Holland.

Poirier and Ruud (1988), Probit with dependent observations, *Review of Economic Studies* 55, 593-614.

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

Pötscher, B.M. and I.R. Prucha (1997) *Dynamic nonlinear econometric models*. Berlin: Springer-Verlag.

Robinson, P.M. (1982), On the asymptotic properties of estimators of models containing LDV, *Econometrica* 50, 27-41.

Ruud, P. (1981), Conditional minimum distance estimation and autocorrelation in limited dependent variable models, Chapter 3 of Ph.D. thesis, Department of Economics, MIT.

White, H. (2001) *Asymptotic theory for Econometricians*. New York: Academic Press.

Wooldridge, J. (1994), Estimation and inference for dependent processes, in *Handbook of Econometrics*, volume 4, ed. by R. F. Engle and D. MacFadden. Amsterdam: North-Holland.

Proofs

Proof of Theorem 1:

The dynamic binary choice model of order p can be written as

$$y_n = I(\sum_{i=1}^{p} \rho_i y_{n-i} + \eta_n > 0) = g(y_{n-1}, y_{n-2}, \ldots, y_{n-p}, \eta_n).$$

This $g(., \ldots, .)$ satisfies, for all 0-1 valued $y_1, y_2, \ldots, y_{n-p}$ and $\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_{n-p}$,

$$|g(y_1, y_2, \ldots, y_p, \eta_n) - g(\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_p, \eta_n)| \leq L(\eta_n) \max_{j=1,\ldots,p} |y_j - \tilde{y}_j|,$$

where

$$L(\eta_n) = \sup_{\phi, \phi' \in \Phi} |I(\phi + \eta_n > 0) - I(\phi' + \eta_n > 0)|.$$

The idea of the proof is to show that the process y_n can be approximated arbitrarily well by using a function of a finite number of η_n - this is the content of the near epoch dependence concept. We do this by using for our approximation $\hat{y}_n^m$ the y that would have resulted if the process had been started up using 0 values for the y_n and η_n that occurred m periods

or longer ago. Formally, for all n define $\hat{y}_n^m = 0$ for $m \leq 0$. Then for all $m \geq 1$ recursively define

$$\hat{y}_n^m = g(\hat{y}_{n-1}^{m-1}, \hat{y}_{n-2}^{m-2}, \ldots, \hat{y}_{n-p}^{m-p}, \eta_n).$$

Note that by construction, $\hat{y}_n^m = f_m(\eta_n, \eta_{n-1}, \ldots, \eta_{n-m})$. Define $\max_{j \in A} c_j = 0$ if A is empty. Then for these approximators $\hat{y}_n^m$ we have

$$\max_{j=1,\ldots,p} |y_{n-j+1} - \hat{y}_{n-j+1}^{m-j+1}|$$

$$= \max(|g(y_{n-1}, y_{n-2}, \ldots, y_{n-p}, \eta_n) - g(\hat{y}_{n-1}^{m-1}, \hat{y}_{n-2}^{m-2}, \ldots, \hat{y}_{n-p}^{m-p}, \eta_n)|, \max_{j=2,\ldots,p} |y_{n-j+1} - \hat{y}_{n-j+1}^{m-j+1}|)$$

$$\leq \max(L(\eta_n) \max_{j=1,\ldots,p} |y_{n-j} - \hat{y}_{n-j}^{m-j}|, \max_{j=2,\ldots,p} |y_{n-j+1} - \hat{y}_{n-j+1}^{m-j+1}|)$$

$$\leq \max(L(\eta_n)|y_{n-p} - \hat{y}_{n-p}^{m-p}|, \max_{j=2,\ldots,p} |y_{n-j+1} - \hat{y}_{n-j+1}^{m-j+1}|)$$

$$\leq \max(L(\eta_n)|y_{n-p} - \hat{y}_{n-p}^{m-p}|, L(\eta_{n-1}) \max_{j=1,\ldots,p} |y_{n-j-1} - \hat{y}_{n-j-1}^{m-j-1}|, \max_{j=3,\ldots,p} |y_{n-j+1} - \hat{y}_{n-j+1}^{m-j+1}|)$$

$$\leq \max(L(\eta_n)|y_{n-p} - \hat{y}_{n-p}^{m-p}|, L(\eta_{n-1})|y_{n-p-1} - \hat{y}_{n-p-1}^{m-p-1}|,$$

$$L(\eta_{n-1})|y_{n-p} - \hat{y}_{n-p}^{m-p}|, \max_{j=3,\ldots,p} |y_{n-j+1} - \hat{y}_{n-j+1}^{m-j+1}|)$$

$$\leq \max_{j=1,\ldots,p} L(\eta_{n-j+1}) \max_{j=1,\ldots,p} |y_{n-p-j+1} - \hat{y}_{n-p-j+1}^{m-p-j+1}|,$$

and because $0 \leq L(u) \leq 1$, we also have by repeating this reasoning K times, for all $K \geq 1$,

$$\max_{j=1,\ldots,p} |y_{n-j+1} - \hat{y}_{n-j+1}^{m-j+1}|$$

$$\leq \max_{j=1,\ldots,p} L(\eta_{n-j+1}) \max_{j=1,\ldots,p} |y_{n-pK-j+1} - \hat{y}_{n-pK-j+1}^{m-pK-j+1}|.$$

Next, note that by assumption there exists a positive integer $\tilde{K}$ such that, for some $\delta > 0$,

$$|E(\max_{j=1,\ldots,p} L(\eta_{n-j+1})|y_{n-p\tilde{K}}, y_{n-p\tilde{K}-1}, \ldots)|$$

$$= |P(\phi_{max} + \max_{j=1,\ldots,p} \eta_{n-j+1} > 0|y_{n-p\tilde{K}}, y_{n-p\tilde{K}-1}, \ldots)$$

$$-P(\phi_{min} + \min_{j=1,\ldots,p} \eta_{n-j+1} > 0 | y_{n-p\tilde{K}}, y_{n-p\tilde{K}-1}, \ldots)| < 1 - \delta$$

for some $\delta > 0$. Therefore, for $m \geq p\tilde{K}$,

$$\chi_m \stackrel{def}{=} \sup_{n\in\mathbb{Z}} E \max_{j=1,\ldots,p} |y_{n-j+1} - \hat{y}_{n-j+1}^{m-j+1}|^2$$

$$\leq \sup_{n\in\mathbb{Z}} E \left(E(\max_{j=1,\ldots,p} L(\eta_{n-j+1}) | y_{n-p\tilde{K}}, y_{n-p\tilde{K}-1}, \ldots) \max_{j=1,\ldots,p} |y_{n-p\tilde{K}-j+1} - \hat{y}_{n-p\tilde{K}-j+1}^{m-p\tilde{K}-j+1}|^2 \right)$$

$$\leq (1-\delta) \sup_{n\in\mathbb{Z}} E \max_{j=1,\ldots,p} |y_{n-p\tilde{K}-j+1} - \hat{y}_{n-p\tilde{K}-j+1}^{m-p\tilde{K}-j+1}|^2$$

$$= (1-\delta) \sup_{n\in\mathbb{Z}} E \max_{j=1,\ldots,p} |y_{n-j+1} - \hat{y}_{n-j+1}^{m-p\tilde{K}-j+1}|^2$$

$$= (1-\delta)\chi_{m-p\tilde{K}}.$$

Because $\chi_j \leq 1$ for all $j \geq 0$, it now follows that

$$\chi_m \leq (1-\delta)^{[m/(p\tilde{K})]},$$

where $[x]$ denotes the integer part of x. Now, because the conditional expectation is the best possible L_2-approximation,

$$\sup_{n\in\mathbb{Z}} E|y_n - E(y_n|\eta_n, \eta_{n-1}, \ldots, \eta_{n-m})|^2$$

$$\leq \sup_{n\in\mathbb{Z}} E \max_{j=1,\ldots,p} |y_{n-j+1} - \hat{y}_{n-j+1}^{m-j+1}|^2$$

$$\leq (1-\delta)^{[m/(2\tilde{K})]} \leq C_1 \exp(-C_2 m)$$

for positive constants C_1 and C_2 depending on δ and on the $\alpha(.)$ sequence.
To show that (y_n, η_n) is strictly stationary, note that $\hat{y}_n^m = f_m(\eta_{n-m}, \ldots, \eta_n)$ by construction, where $f_m(., \ldots, .)$ does not depend on n or N. This then implies that (y_n, η_n) is strictly stationary. □

Proof of Theorem 2:

The definition of the strong mixing coefficients is

$$\sup_{n\in\mathbb{Z}} \sup_{F,G}\{|P((x_n, y_n) \in F, (x_{n+m}, y_{n+m}) \in G) - P((x_n, y_n) \in F)P((x_{n+m}, y_{n+m}) \in G)|\},$$

see for example White (2001, page 47). Because y_n is a 0/1-valued random variable, there are only four possibilities for the possible values of the (y_n, y_{n-m}) pair.
For the case $y_n = 1$, $y_{n+m} = 1$, we now have, defining $\mathcal{F}_n = \sigma(v_n, v_{n-1}, \ldots)$,

$$|EI((x_n, 1) \in F)y_n I((x_{n+m}, 1) \in G)y_{n+m} - EI((x_n, 1) \in F)y_n EI((x_{n+m}, 1) \in G)y_{n+m}|$$

$$= |EI((x_n, 1) \in F)y_n[E(I((x_{n+m}, 1) \in G)y_{n+m}|\mathcal{F}_n) - EI((x_{n+m}, 1) \in G)y_{n+m}]|$$

$$\leq E|E((I(x_{n+m}, 1) \in G)y_{n+m}|\mathcal{F}_n) - E(I((x_{n+m}, 1) \in G)y_{n+m})|,$$

and convergence to zero with m of the last expression constitutes the L_1-mixingale condition for $I((x_n, 1) \in G)y_n$. Now $I((x_n, 1) \in G)y_n$ is a sequence that is bounded and near epoch dependent on v_n, implying that it is an L_1-mixingale, which in turn implies that

$$E|E(I((x_n, 1) \in G)y_n|\mathcal{F}_{n-m}) - E(I((x_n, 1) \in G)y_n)|$$

$$\leq C(\nu(m) + \alpha(m)).$$

The cases $y_n = 1$, $y_{n+m} = 0$; $y_n = 0$, $y_{n+m} = 1$; and $y_n = 0$, $y_{n-m} = 0$ are analogous, which then proves the result. □

For the proof of Theorem 3, we need the following two lemmas.

Lemma 1 *Under the conditions of Theorem 3,* $E\sup_{b\in B}|l_n(b)| < \infty$.

Proof of Lemma 1:

Define $w_n = (y_{n-1}, \ldots, y_{n-p}, x_n')'$ and note that $Ew_n w_n'$ exists by Assumption A1. Existence of $Ew_n w_n'$ and the probit specification imply the result. The reasoning is similar to the result for cross-section probit, see Newey and McFadden (1994, page 2125, Example 1.2). □

Lemma 2 *Under the conditions of Theorem 3, (i)* $E(w_n w_n')$ *is positive definite and (ii)* $El_n(b)$ *is uniquely minimized at* $b = \beta$.

Proof of Lemma 2:

The assumptions of Theorem 1 are satisfied so that $(x_n', y_n)'$ is strongly stationary. The assumption that distribution of x_n is not contained in any linear subspace of $\mathbb{R}^q$ implies that the distribution of w_n is not contained in any proper linear subspace of $\mathbb{R}^{p+q}$ (see the proof of Lemma 3 for a proof of the general case). This implies that $Ew_n w_n'$ is nonsingular so that $Ew_n w_n'$ is positive definite. Let $b \neq \beta$ so that $E[(x_n'(b-\beta))^2] = (b-\beta)'Ew_n w_n'(b-\beta) > 0$, implying that $w_n'(b-\beta) \neq 0$ on a set with positive probability, implying that $w_n'b \neq w_n'\beta$ on a set with positive probability. Both $\Phi(z)$ and $\bar{\Phi}(z) = 1 - \Phi(z)$ are strictly monotonic, and therefore $w_n'b \neq w_n'\beta$ implies that both $\Phi(w_n'b) \neq \Phi(w_n'\beta)$ and $\bar{\Phi}(w_n'b) \neq \bar{\Phi}(w_n'\beta)$. Thus, the density

$$p(y_n|w_n, b) = \Phi(w_n'\beta)^{y_n}\bar{\Phi}(w_n'b)^{1-y_n} \neq p(y_n|w_n, \beta)$$

on a set with positive probability. Lemma 1 stated that $E\sup_{b\in B}|l_n(b)| < \infty$, and therefore both conditions of Newey and McFadden (1994, Lemma 2.2) are satisfied. Therefore, $El_n(b)$ is uniquely minimized at $b = \beta$.

□

Proof of Theorem 3:

It is easily seen from Lemma 1 and Lemma 2 that all the conditions of Theorem A1 of Wooldridge (1994) are satisfied, except for the condition of uniform convergence in probability of $L_N(b)$. Note that Wooldridge's Theorem A1 can be extended to include a strong convergence result if instead of uniform convergence in probability of $L_N(b)$, uniform almost sure convergence $L_N(b)$ is assumed. To show this uniform convergence, we use the generic uniform law of large numbers of Andrews (1987). To show a weak or strong uniform law of large numbers, this theorem requires compactness of the parameter space, and in addition it needs to be verified that the summands $q_n(w_n, b)$ are such that $q_n(w_n, b)$, $q_n^*(w_n, b) = \sup\{q_n(w_n, \tilde{b}) : \tilde{b} \in B, |\tilde{b} - b| < \rho\}$ and $q_{n*}(w_n, b) = \inf\{q_n(w_n, \tilde{b}) : b \in B, |\tilde{b} - b| < \rho\}$ are well-defined and satisfy a (respectively weak or strong) law of large numbers, and that for all $b \in B$,

$$\lim_{\rho\to 0}\sup_{n\in\mathbb{Z}}|N^{-1}\sum_{n=1}^{N} Eq_n^*(w_n, b) - Eq_{*n}(w_n, b)| = 0.$$

The latter condition follows from stationarity of (y_n, x_n), continuity, and the envelope condition of Assumption A. In addition, $q_n(w_n, b)$, $q_n^*(w_n, b)$ and $q_{n*}(w_n, b)$ are well-defined

and strong mixing random variables, implying that weak law of large numbers for uniformly integrable mixingales of Andrews (1988) applies. Alternatively we can apply the strong law of large numbers of Theorem 4 of de Jong (1995), from which it follows that if $\alpha(m) + \nu(m) \leq Cm^{-\eta}$ for some positive constants C and η, these variables will satisfy a strong law of large numbers. This is because under the condition that $E|l_n(b)|^{1+\delta} < \infty$, the summands will be an $L_{1+\delta/2}$-mixingale. □

Lemma 3 *Under the conditions of Theorem 4,*

$$(N-p)^{1/2}(\partial L_N(b)/\partial b)|_{b=\beta} \xrightarrow{d} N(0, I).$$

Proof:

Note that by assumption, $E((\partial L_n(b)/\partial b)|_{b=\beta}|w_n) = 0$ so that $E(\partial L_n(b)/\partial b)|_{b=\beta} = 0$. Moreover, $(\partial L_n(b)/\partial b)|_{b=\beta}$ is a martingale difference sequence that is strong mixing and strictly stationary. In particular, the version of Bierens (2004, Theorem 7.11) of a central limit theorem of McLeish (1974) yields asymptotic normality. Applying the information matrix equality yields the result. □

Proof of Theorem 4:

We prove Theorem 4 by checking the conditions of Newey and McFadden (1994, theorem 3.1). Consistency was shown in Theorem 3. Condition (i) was assumed. Condition (ii), twice differentiability of the log likelihood, follows from the probit specification. Condition (iii) was shown in Lemma 3. Note that stationarity and strong mixing imply ergodicity, see White (2001, theorem 3.34). Condition (iv) then follows from the probit specification and reasoning similar to Newey and McFadden, page 2147, example 1.2. Nonsingularity follows from the probit specification and $Ew_n w_n'$ being positive definite so that condition (iv) is satisfied. □

For the proof of Theorem 5, we need the following lemmas.

Lemma 4 *For all $a \in \mathbb{R}$, if $0 \leq z_n \leq 1$ and (z_n, x_n) is strictly stationary and strong mixing, then*

$$\sup_{b \in B} |N^{-1} \sum_{n=1}^{N} (z_n I(b' x_n \leq a) - E z_n I(b' x_n \leq a))| \xrightarrow{p} 0.$$

In addition, if $\alpha(m) \leq C m^{-\eta}$ for positive constants C and η, the convergence is almost surely.

Proof of Lemma 4:

We will apply the generic uniform law of large numbers of the Theorem of Andrews (1987). It requires compactness of the parameter space B (which is assumed), and in addition it needs to be verified that the summands $q_n(w_n, b)$ are such that $q_n(w_n, b)$, $q_n^*(w_n, b) = \sup\{q_n(w_n, \tilde{b}) : \tilde{b} \in B, |\tilde{b} - b| < \rho\}$ and $q_{n*}(w_n, b) = \inf\{q_n(w_n, \tilde{b}) : b \in B, |\tilde{b} - b| < \rho\}$ are well-defined and satisfy a (respectively weak or strong) law of large numbers, and for all $b \in B$,

$$\lim_{\rho \to 0} \sup_{n \in \mathbb{Z}} |N^{-1} \sum_{n=1}^{N} E q_n^*(w_n, b) - E q_{*n}(w_n, b)| = 0.$$

To show the last result, note that (z_n, x_n) is strictly stationary under the conditions of the theorem, and therefore

$$\lim_{\rho \to 0} \sup_{n \in \mathbb{Z}} |N^{-1} \sum_{n=1}^{N} E q_n^*(w_n, b) - E q_{*n}(w_n, b)|$$

$$= \lim_{\rho \to 0} \sup_{n \in \mathbb{Z}} |E z_n I(\sup_{\tilde{b}: |b - \tilde{b}| < \rho} b' x_n < a) - E z_n I(\inf_{\tilde{b}: |b - \tilde{b}| < \rho} b' x_n < a)|$$

$$\leq \limsup_{K \to \infty} \lim_{\rho \to 0} \sup_{n \in \mathbb{Z}} |E z_n (I(b' x_n < a + \rho |x_n|) - I(b' x_n < a - \rho |x_n|)) I(|x_n| \leq K)$$

$$+ \limsup_{K \to \infty} \lim_{\rho \to 0} \sup_{n \in \mathbb{Z}} |E z_n (I(b' x_n < a + \rho |x_n|) - I(b' x_n < a - \rho |x_n|)) I(|x_n| > K))|$$

$$\leq \limsup_{K \to \infty} \lim_{\rho \to 0} |P(b' x_n < a + \rho K) - P(b' x_n < a - \rho K)| + \limsup_{K \to \infty} P(|x_n| > K) = 0,$$

because x_{1n} has a continuous distribution. Furthermore, note that $q_n(z_n, b)$,

$$q_n^*(w_n, b) = z_n I(\sup_{\tilde{b}: |b - \tilde{b}| < \rho} b' x_n < a)$$

and

$$q_{*n}(w_n, b) = z_n I(\inf_{\tilde{b}:|b-\tilde{b}|<\rho} b' x_n < a)$$

are well-defined and strong mixing random variables, implying that weak law of large numbers for mixingales of Andrews (1988) applies; or alternatively we can apply the strong law of large numbers of Theorem 4 of de Jong (1995), from which it follows that if $\alpha(m) + \nu(m) \leq Cm^{-\eta}$ for some positive constants C and η, these variables will satisfy a strong law of large numbers (note that because of boundedness of the summands, the summands are L_2-mixingales). □

Lemma 5 *Under Assumptions 1,2, 3, 4 and 5,*

$$\sup_{b\in B} |S_N(b, \sigma_N) - ES_N(b, \sigma_N)| \xrightarrow{p} 0.$$

In addition, if $\alpha(m) \leq Cm^{-\eta}$ for positive constants C and η, the convergence is almost surely.

Proof of Lemma 5:

First note that Horowitz' proof of his Lemma 4 (i.e. $\sup_{b\in B} |S_N(b, \sigma_N) - S_N^*(b)| \xrightarrow{as} 0$) goes through as it stands, except for the proof of uniform convergence of the term in his Equation (A4), which uses a uniform law of large numbers for i.i.d. random variables. To show that

$$\sup_{b\in B} |N^{-1} \sum_{n=1}^{N} (I(|\sum_{j=1}^{p} \rho_j y_{n-j} + c' x_n| < \alpha) - EI(|\sum_{j=1}^{p} \rho_j y_{n-j} + c' x_n| < \alpha))|$$

satisfies a strong or weak law of large numbers, we can use Lemma 4. To do so, note that

$$N^{-1} \sum_{n=1}^{N} (I(|\sum_{j=1}^{p} \rho_j y_{n-j} + c' x_n| < \alpha)$$

$$= \sum_{j_1=0}^{1} \cdots \sum_{j_p=0}^{1} N^{-1} \sum_{n=1}^{N} I(y_{n-1} = j_1) \ldots I(y_{n-p} = j_p) I(|\sum_{i=1}^{p} \rho_i j_i + c' x_n| < \alpha)$$

and note that $I(y_{n-1} = j_1) \dots I(y_{n-p} = j_p)$ is strong mixing, because it is the product of strong mixing random variables. It now only remains to be proven that

$$\sup_{b \in B} |S_N^*(b) - S(b)| \xrightarrow{p} 0 \quad \text{or} \quad \xrightarrow{as} 0,$$

which Horowitz shows by referring to Manski (1985). This can be shown by noting that

$$S_N^*(b) = N^{-1} \sum_{n=1}^{N} (2I(y_n = 1) - 1) I(b' x_n \geq 0)$$

$$= 2N^{-1} \sum_{n=1}^{N} y_n I(b' x_n \geq 0) - N^{-1} \sum_{n=1}^{N} I(b' x_n \geq 0),$$

and by Lemma 4, both terms satisfy a (weak or strong) uniform law of large numbers. □

Lemma 6 *Under Assumptions 1,2, 3 and 4,* $S(b) \leq S(\beta)$ *with equality holding only if* $b = \beta$.

Proof of Lemma 6:

This result follows by noting that all conditions from Lemma 3 of Manski (1985) are satisfied, except that we need to show that the distribution of $(y_{n-1}, \dots, y_{n-p}, x_n')'$ is not contained in any proper linear subspace of $\mathbb{R}^{p+q}$. To show this, note that under the assumptions,

$$\delta < p(y_n | x_n) < 1 - \delta,$$

$$\delta < p(y_n | y_{n-1}, x_n) < 1 - \delta,$$

$$\dots$$

$$\delta < p(y_n | y_{n-1}, \dots, x_n) < 1 - \delta.$$

The first of the above equations implies that y_n is not a deterministic function of x_n, so y_n cannot be a linear function of x_n. Therefore, since by assumption the distribution of x_n is not contained in any proper linear subspace of $\mathbb{R}^q$, the distribution of $(y_n, x_n')'$ is not contained in any proper linear subspace of $\mathbb{R}^{q+1}$. From the second of the above equations, it follows likewise that y_n cannot be a linear function of y_{n-1} and x_n, and that therefore the distribution of $(y_n, x_n')'$ is not contained in any proper linear subspace of $\mathbb{R}^{q+2}$. Continuing this reasoning, it follows that the distribution of $(y_n, y_{n-1}, \dots, y_{n-p}, x_n')'$ is not contained in any proper linear subspace of $\mathbb{R}^{p+q+1}$. Therefore certainly, the distribution of $(y_{n-1}, \dots, y_{n-p}, x_n')'$ is not contained in any proper linear subspace of $\mathbb{R}^{p+q}$. □

Proof of Theorem 5 :

The proof of the theorem now follows from Theorem A1 of Wooldridge (1994) and the results of Lemma 5 and Lemma 6. □

Let $z_n = \sum_{j=1}^{p} \rho_j y_{n-j} + \gamma' x_n$. The following lemma shows that Horowitz' Lemma 5 holds as it stands in our setting:

Lemma 7 *Under Assumptions 1' and Assumptions 2-14,*

$$\lim_{N\to\infty} E[\sigma_N^{-h} T_N(\beta, \sigma_N)] = A;$$

$$\lim_{N\to\infty} \mathrm{Var}[(N\sigma_N)^{1/2} T_N(\beta, \sigma_N)] = D.$$

Proof of Lemma 7:

The only adjustment to Horowitz' Lemma 5 that needs to be made is to show that the covariance terms in $\mathrm{Var}[(N\sigma_N)^{1/2} T_N(\beta, \sigma_N)]$ are asymptotically negligible. To prove this, we show that for all vectors γ such that $|\gamma| = 1$,

$$\lim_{N\to\infty} \sigma_N \sum_{m=1}^{\infty} |\mathrm{cov}(\gamma'(\tilde{x}_n/\sigma_N)K'(z_n/\sigma_N), \gamma'(\tilde{x}_{n-m}/\sigma_N)K'(z_{n-m}/\sigma_N))| = 0.$$

By the covariance inequality for mixingales (see Davidson (1994, p. 212, Corollary 14.3)),

$$\sigma_N \mathrm{cov}(\gamma'(\tilde{x}_n/\sigma_N)K'(z_n/\sigma_N), \gamma'(\tilde{x}_{n-m}/\sigma_N)K'(z_{n-m}/\sigma_N))$$

$$\leq \sigma_N C\alpha(m)^{1-2/r} (E|\gamma'(\tilde{x}_n/\sigma_N)K'(z_n/\sigma_N)|^r)^{1/r} (E|\gamma'(\tilde{x}_{n-m}/\sigma_N)K'(z_{n-m}/\sigma_N)|^r)^{1/r}$$

$$= \sigma_N^{-1} C\alpha(m)^{1-2/r} \Big(\int |\gamma'\tilde{x}|^r |K'(z/\sigma_N)|^r p(z|\tilde{x}) dz dP(\tilde{x}) \Big)^{2/r}$$

$$= C\alpha(m)^{1-2/r} \sigma_N^{2/r-1} \Big(\int |\gamma'\tilde{x}|^r |K'(\zeta)|^r p(\sigma_N\zeta|\tilde{x}) d\zeta dP(\tilde{x}) \Big)^{2/r}$$

by substituting $\zeta = z/\sigma_N$. The last term is smaller than $C'\sigma_N^{2/r-1}\alpha(m)^{1-2/r}$ for some constant C'. Also, note that by Horowitz' arguments, under the conditions of the theorem,

$$\sigma_N E\gamma'(\tilde{x}_{n-m}/\sigma_N)K'(z_{n-m}/\sigma_N) = O(\sigma_N),$$

implying that

$$\sigma_N \text{cov}(\gamma'(\tilde{x}_n/\sigma_N)K'(z_n/\sigma_N), \gamma'(\tilde{x}_{n-m}/\sigma_N)K'(z_{n-m}/\sigma_N))$$

$$= O(\sigma_N) + \sigma_N E(\sigma_N \gamma'(\tilde{x}_n/\sigma_N)K'(z_n/\sigma_N)\gamma'(\tilde{x}_{n-m}/\sigma_N)K'(z_{n-m}/\sigma_N))$$

$$= O(\sigma_N) + \sigma_N^{-1} \int \gamma'\tilde{x}_n K'(z_n/\sigma_N)\gamma'\tilde{x}_{n-m}K'(z_{n-m}/\sigma_N)dP(x_n, x_{n-m}, z_n, z_{n-m})$$

$$= O(\sigma_N) + \sigma_N^{-1} \int \gamma'\tilde{x}_n K'(z_n/\sigma_N)\gamma'\tilde{x}_{n-m}K'(z_{n-m}/\sigma_N)dp(z_n, z_{n-m}|x_n, x_{n-m})dz_n dz_{n-m}dP(x_n, x_{n-m})$$

$$= O(\sigma_N) + \sigma_N \int\int K'(\zeta_n)K'(\zeta_{n-m})p(\sigma_N\zeta_n, \sigma_N\zeta_{n-m}|x_n, x_{n-m})d\zeta_n d\zeta_{n-m}\gamma'\tilde{x}_{n-m}\gamma'\tilde{x}_n dP(x_n, x_{n-m})$$

$$= O(\sigma_N)$$

under the assumptions of the theorem. Therefore for any $\beta \in (0, 1)$,

$$\sum_{m=1}^{\infty} |\text{cov}(\gamma'(\tilde{x}_n/\sigma_N)K'(z_n/\sigma_N), \gamma'(\tilde{x}_{n-m}/\sigma_N)K'(z_{n-m}/\sigma_N))|$$

$$\leq C \sum_{m=1}^{\infty} (\sigma_N)^{\beta}(\alpha(m)^{(1-2/r)}\sigma_N^{2/r-1})^{1-\beta},$$

and by choosing $\beta = (r-2)/(2r-2) + \eta$ and $\eta > 0$ small enough, the last term can be bounded by

$$C(\sum_{m=1}^{\infty} \alpha(m)^{(r-2)/(2r-2)-\eta(r-2)/r})\sigma_N^{(2r-2)\eta/r} = O(\sigma_N^{(2r-2)\eta/r}) = o(1),$$

where the finiteness of the summation follows from the assumptions. □

Horowitz' Lemma 6 now holds as follows:

Lemma 8 *Under Assumptions 1' and Assumptions 2-14, (a) If $N\sigma_N^{2h+1} \to \infty$ as $N \to \infty$, $\sigma_N^{-h}T_N(\beta, \sigma_N) \xrightarrow{p} A$. (b) If $N\sigma_N^{2h+1}$ has a finite limit λ as $N \to \infty$, $(N\sigma_N)^{1/2}T_N(\beta, \sigma_N) \xrightarrow{d} N(\lambda^{1/2}A, D)$.*

Proof of Lemma 8:

The modification of Horowitz (1992) that is needed is to show that for all vectors γ such that $|\gamma| = 1$,

$$(\sigma_N/N)^{1/2}\gamma' \sum_{n=1}^{N}(t_{Nn} - Et_{Nn}) \xrightarrow{d} N(0, \gamma' D\gamma),$$

where

$$t_{Nn} = (2y_n - 1)(\tilde{x}_n/\sigma_N)K'(z_n/\sigma_N).$$

Since t_{Nn} is strong mixing, Theorem 2 of de Jong (1997) for strong mixing arrays can now be applied to show this result under the conditions of the lemma. Note that the condition $\alpha(m) \leq Cm^{-r/(r-2)-\eta}$ from that theorem follows from the assumptions of the lemma. □

For reproving Horowitz' Lemma 7 for the case of strong mixing data, we need the following lemmas:

Lemma 9 (Azuma(1967)) *If η_n is a martingale difference sequence with respect to $\mathcal{F}_n$ and $|\eta_n| \leq C$, then*

$$P(|N^{-1}\sum_{n=1}^{N}\eta_n| > \delta) \leq 2\exp(-N\delta^2/C^2).$$

Proof of Lemma 9:

See Azuma (1967). □

An m_N-fold application of the above lemma now gives the following result:

Lemma 10 *If $\mathcal{F}_n$ is a sequence of sigma-fields such that $\eta_n - E(\eta_n|\mathcal{F}_{n-1})$ is a martingale difference sequence with respect to $\mathcal{F}_n$, then for any integer-valued sequence m_N such that $m_N \geq 1$,*

$$P(|N^{-1}\sum_{n=1}^{N}(\eta_n - E(\eta_n|\mathcal{F}_{n-m_N}))| > \delta) \leq 2m_N\exp(-\delta^2/(m_N^2 C_N^2)).$$

Proof of Lemma 10:

Obviously

$$N^{-1}\sum_{n=1}^{N}(\eta_n - E(\eta_n|\mathcal{F}_{n-m_N})) = \sum_{j=0}^{m_N-1} N^{-1}\sum_{n=1}^{N}(E(\eta_n|\mathcal{F}_{n-j}) - E(\eta_n|\mathcal{F}_{n-j-1})),$$

and therefore

$$P(|N^{-1}\sum_{n=1}^{N}(\eta_n - E(\eta_n|\mathcal{F}_{n-m_N}))| > \delta)$$

$$\leq \sum_{j=0}^{m_N-1} P(|N^{-1}\sum_{n=1}^{N}(E(\eta_n|\mathcal{F}_{n-j}) - E(\eta_n|\mathcal{F}_{n-j-1}))| > \delta/m_N)$$

$$\leq 2m_N \exp(-\delta^2/(m_N^2 C_N^2))$$

by m_N applications of Lemma 9. □

As in Horowitz (1992), define

$$g_{Nn}(\theta) = (2I(y_n = 1) - 1)\tilde{x}_n K'(z_n/\sigma_N + \theta'\tilde{x}_n) -$$
$$E(2I(y_n = 1) - 1)\tilde{x}_n K'(z_n/\sigma_N + \theta'\tilde{x}_n).$$

The following result is now the analogue[2] of Horowitz' Lemma 7.

Lemma 11 *If (y_n, x_n) is strong mixing with strong mixing sequence $\alpha(m)$, and there exists a sequence $m_N \geq 1$ such that*

$$\sigma_N^{-3(p+q-1)}\sigma_N^{-2}N^{1/r}\alpha(m_N) + (\log(Nm_N))(N^{1-2/r}\sigma_N^4 m_N^{-2})^{-1} \to 0,$$

then

$$\sup_{\theta\in\Theta_N} |(N\sigma_N^2)^{-1}\sum_{n=1}^{N}(g_{Nn}(\theta) - Eg_{Nn}(\theta))| \xrightarrow{p} 0.$$

Note that the second part of Horowitz' Lemma 7 will hold without modification. Also note that the case of i.i.d. (y_n, x_n) is a special case, because then $\alpha(m) = 0$ for $m \geq 1$, and we could set $m_N = 1$ for that case.

[2]Note that Horowitz' Lemma 7 only holds for bounded regressors, and that the truncation argument at the start of Lemma 8 appears to be in error. Horowitz does not explicitly consider the remainder statistic containing the summation elements for which $|\tilde{x}_n|$ exceeds a. Horowitz' Lemma 9 appears to have a similar problem in its proof. Therefore, Lemma 11 also serves to correct this aspect of Horowitz' proof.

Proof of Lemma 11:

Consider

$$g_{Nn}^{C_N}(\theta) = (2I(y_n = 1) - 1)\tilde{x}_n K'(z_n/\sigma_N + \theta'\tilde{x}_n)I(|\tilde{x}_n| \leq C_N)$$
$$-E(2I(y_n = 1) - 1)\tilde{x}_n K'(z_n/\sigma_N + \theta'\tilde{x}_n)I(|\tilde{x}_n| \leq C_N),$$

and note that obviously,

$$g_{Nn}(\theta) - Eg_{Nn}(\theta) = (g_{Nn}^{C_N}(\theta) - Eg_{Nn}^{C_N}(\theta))$$
$$+(g_{Nn}(\theta) - g_{Nn}^{C_N}(\theta) - Eg_{Nn}(\theta) + Eg_{Nn}^{C_N}(\theta)). \tag{40}$$

Now define $C_N = \eta^{-1/r} N^{1/r} (E|\tilde{x}_n|^r)^{1/r}$ for any $\eta > 0$. Then because $C_N \to \infty$ as $N \to \infty$, following the reasoning as in the proof of (A16) of Horowitz (1992, page 525-526), it follows that

$$\sup_{\theta \in \Theta} |Eg_{Nn}(\theta) - Eg_{Nn}^{C_N}(\theta)| \to 0. \tag{41}$$

In addition,

$$P(\sup_{\theta \in \Theta} |\sum_{n=1}^{N} (g_{Nn}(\theta) - g_{Nn}^{C_N}(\theta))| = 0) \leq P(\exists n : |\tilde{x}_n| > C_N) \leq NE|\tilde{x}_n|^r C_N^{-r} \leq \eta, \tag{42}$$

and we can choose η arbitrarily small. For the case $r = \infty$, it is trivial that these two terms disappear asymptotically for some constant C_N not depending on N. To deal with the first part of Equation (40), note that

$$g_{Nn}(\theta) - Eg_{Nn}(\theta) = (g_{Nn}(\theta) - E(g_{Nn}(\theta)|\mathcal{F}_{n-m_N})) + (E(g_{Nn}(\theta)|\mathcal{F}_{n-m_N}) - Eg_{Nn}(\theta)). \tag{43}$$

To deal with the first part of the right-hand side of Equation (43), we can copy the argument on page 525 of Horowitz (1992), except that now, by Lemma 10,

$$\sum_{i=1}^{\Gamma_N} P((N\sigma_N^2)^{-1} |\sum_{n=1}^{N} (g_{Nn}(\theta_{Ni}) - Eg_{Nn}(\theta_{Ni}))| > \varepsilon/2)$$
$$\leq 2\Gamma_N m_N \exp(-\varepsilon^2 4^{-1} N\sigma_N^4 C_N^{-2} m_N^{-2}).$$

Since $\Gamma_N = O(\sigma_N^{-3(p+q-1)})$, this term will converge to zero if

$$(\log(Nm_N))(N\sigma_N^4 C_N^{-2} m_N^{-2})^{-1} \to 0, \tag{44}$$

which is assumed. For dealing with the second part of the right-hand side of Equation (43), note since $g_{Nn}(\theta)$ is strong mixing, it is also an L_1-mixingale (see for example Davidson (1994, p. 249, Example 16.3), implying that

$$E|E(g_{Nn}^{C_N}(\theta)|\mathcal{F}_{n-m_N}) - Eg_{Nn}^{C_N}(\theta)| \leq 6C_N\alpha(m_N).$$

Using Horowitz' reasoning of page 525, it now suffices to show that for all $\varepsilon > 0$,

$$\sum_{i=1}^{\Gamma_N} P((N\sigma_N^2)^{-1}|\sum_{n=1}^{N} E(g_{Nn}^{C_N}(\theta_{Ni})|\mathcal{F}_{n-m_N}) - Eg_{Nn}^{C_N}(\theta_{Ni})| > \varepsilon) \to 0.$$

By the Markov inequality,

$$\sum_{i=1}^{\Gamma_N} P((N\sigma_N^2)^{-1}|\sum_{n=1}^{N} E(g_{Nn}^{C_N}(\theta_{Ni})|\mathcal{F}_{n-m_N}) - Eg_{Nn}^{C_N}(\theta_{Ni})| > \varepsilon)$$

$$\leq \sum_{i=1}^{\Gamma_N} \varepsilon^{-1}\sigma_N^{-2}N^{-1}\sum_{n=1}^{N} E|E(g_{Nn}^{C_N}(\theta)|\mathcal{F}_{n-m_N}) - Eg_{Nn}^{C_N}(\theta)|$$

$$= O(\sigma_N^{-3(p+q-1)}\sigma_N^{-2}C_N\alpha(m_N)) = o(1)$$

by assumption. □

Lemma 12 *Under Assumptions 1' and Assumptions 2-14,* $(\tilde{b}_N - \tilde{\beta})/\sigma_N \xrightarrow{p} 0$.

Proof of Lemma 12:

This follows from Lemma 11 and the reasoning[3] of Horowitz' (1992) Lemma 8. □

The following lemma corresponds[4] to Horowitz' Lemma 9.

[3]It should be noted that the truncation argument that is employed in Horowitz' (1992) proof of his Lemma 8 appears to be incorrect as it stands. This is because the conditioning on the event C_γ does not appear relevant; while Horowitz' $\tilde{x}$ stands for a random variable distributed identically to any $\tilde{x}_n$, the conditioning should be with respect to every $\tilde{x}_n$, $n = 1, \ldots, N$, in order for this argument to work. However, unless $\tilde{x}_n$ is almost surely bounded, such a conditioning set C_γ would depend on N, and will not have the desired property that $\limsup_{\gamma\to\infty} \limsup_{N\to\infty} P(C_\gamma) = 0$.

[4]Note that Horowitz' conditioning on X_N appears to be in error, and note that when Horowitz uses his Lemma 8 in the proof of his Theorem 2, a uniform law of large numbers appears to be needed rather than the result of his Lemma 8.

Lemma 13 *Let* $\{\beta_N\} = \{\beta_{N1}, \tilde{\beta}_N\}$ *be such that* $(\beta_N - \beta)/\sigma_N \xrightarrow{p} 0$ *as* $N \to \infty$. *Then under Assumptions 1' and Assumptions 2-14,*

$$Q_N(\beta_N, \sigma_N) \xrightarrow{p} Q.$$

Proof of Lemma 13:

Remember that

$$Q_N(\beta_N, \sigma_N) = [\sigma_N^{-2} N^{-1} \sum_{n=1}^{N} (2y_n - 1)\tilde{x}_n \tilde{x}_n' K''((\sum_{j=1}^{p} r_j y_{n-j} + c' x_n)/\sigma_N)]_{b=\beta_N}.$$

Since $P(b_1 = \beta_1) \to 1$ and by the assumption that $(\beta_N - \beta)/\sigma_N \xrightarrow{p} 0$ as $N \to \infty$, it suffices to show that for all $\eta > 0$ and any vector γ such that $|\gamma| = 1$,

$$\sup_{|\tilde{\theta}| \le \eta} |N^{-1} \sum_{n=1}^{N} r_{nN}(\tilde{\theta}) - E r_{nN}(\tilde{\theta})| \equiv \sup_{|\tilde{\theta}| \le \eta} |\sigma_N^{-2} N^{-1} \sum_{n=1}^{N} (2y_n - 1)(\gamma' \tilde{x}_n)^2 K''(z_n/\sigma_N + \tilde{\theta}' \tilde{x}_n)$$

$$- E(2y_n - 1)(\gamma' \tilde{x}_n)^2 K''(z_n/\sigma_N + \tilde{\theta}' \tilde{x}_n)| \xrightarrow{p} 0. \tag{45}$$

Note that Horowitz (1992) shows the continuity of $E r_{nN}(\tilde{\theta})$ in $\tilde{\theta}$ uniformly in N. To show the result of Equation (45), note that

$$P(\sup_{|\tilde{\theta}| \le \eta} |N^{-1} \sum_{n=1}^{N} r_{nN}(\tilde{\theta}) I(|r_{nN}(\tilde{\theta})| > C_N) = 0)$$

$$\le \sum_{n=1}^{N} P((\gamma' \tilde{x}_n)^2 > C_N) \le N E|\gamma' \tilde{x}_n|^r C_N^{-r/2}$$

and the last term can be made smaller than ε by choosing $C_N^{-r/2} = N^{-1} \varepsilon (E|\gamma' \tilde{x}_n|^r)^{-1}$. In addition, it is easily verified that

$$\sup_{|\tilde{\theta}| \le \eta} |N^{-1} \sum_{n=1}^{N} E(r_{nN}(\tilde{\theta}) I(|r_{nN}(\tilde{\theta})| > C_N)) \to 0.$$

Because of these two results, it suffices to show uniform convergence to zero in probability of

$$R_N(\tilde{\theta}) = N^{-1} \sum_{n=1}^{N} (r_{nN}(\tilde{\theta}) I(|r_{nN}(\tilde{\theta})| \le C_N) - E r_{nN}(\tilde{\theta}) I(|r_{nN}(\tilde{\theta})| \le C_N)$$

$$+C_N I(|r_{nN}(\tilde{\theta})| > C_N) - EC_N I(|r_{nN}(\tilde{\theta})| > C_N)).$$

Now note that since $\tilde{\theta} \in \mathbb{R}^{p+q-1}$, we can cover the parameter space $\{\tilde{\theta} : |\tilde{\theta}| \leq \eta\}$ with $O(\sigma_N^{-2(p+q-1)/\beta})$ balls of size $\sigma_N^{2/\beta}$ and with centers $\tilde{\theta}_j$. Now note that, by Assumption 14,

$$\sup_{N\geq 1} E \sup_{|\tilde{\theta}-\tilde{\theta}'|<\delta\sigma_N^{2/\beta}} |R_N(\tilde{\theta}) - R_N(\tilde{\theta}')|$$

$$\leq \sup_{N\geq 1} E(\gamma'\tilde{x}_n)^2 L \sup_{|\tilde{\theta}-\tilde{\theta}'|<\delta\sigma_N^{2/\beta}} |\tilde{\theta} - \tilde{\theta}'|^\beta \sigma_N^{-2} \to 0 \qquad \delta \to 0.$$

Using Lemma 11 and following the same reasoning as in the proof of that lemma, we can now argue

$$\limsup_{n\to\infty} P(\sup_{|\tilde{\theta}|\leq\eta} |R_N(\tilde{\theta}) - ER_N(\tilde{\theta})| > \varepsilon)$$

$$\leq \limsup_{n\to\infty} P(\max_j |R_N(\tilde{\theta}_j) - ER_N(\tilde{\theta}_j)| > \varepsilon/2)$$

$$\leq \limsup_{n\to\infty} \sum_j P(|R_N(\tilde{\theta}_j) - ER_N(\tilde{\theta}_j)| > \varepsilon/2)$$

$$= O(\sigma_N^{-2(p+q-1)/\beta}[2m_N \exp(-N\varepsilon^2/(4\sigma_N^4 C_N^2 m_N^2)) + \varepsilon^{-1} C_N \alpha(m_N)])$$

and because $C_N = O(N^{2/r})$, the last term converges to 0 if

$$\sigma_N^{-2(p+q-1)/\beta} N^{2/r} \alpha(m_N) + (m_N^{-2}\sigma_N^{-4} N^{1-4/r})^{-1} \log(Nm_N) \to 0,$$

which is assumed. □

Proof of Theorem 6:

This proof is identical to the proof of Horowitz' Theorem 2, where we need to use our Lemma 12 and Lemma 13 instead of Horowitz' Lemma 8 and Lemma 9. □

Proof of Theorem 7:

Part (a) now follows exactly[5] as in Horowitz' proof of his Theorem 3, where our Lemma 12 and Lemma 13 replace Horowitz' Lemma 8 and Lemma 9. Part (c) follows from Lemma 13. □

[5]To show part (b), note that it appears as though Horowitz (1992, page 530) neglects to provide a formal uniform law of large numbers result of the type

$$\sup_{|\tilde{\theta}|\leq\eta} |\sigma_N^{-1} N^{-1} \sum_{n=1}^{N} (\gamma'\tilde{x}_n)^2 K'(z_n/\sigma_N + \tilde{\theta}'\tilde{x}_n) - E((\gamma'\tilde{x}_n)^2 K'(z_n/\sigma_N + \tilde{\theta}'\tilde{x}_n)| \xrightarrow{p} 0$$

for all γ such that $|\gamma| = 1$. Under the conditions of our theorem, this result can be proven analogously to the proof of Lemma 13, using the same C_N and ball size sequences. Note that $K'(.)$ is Lipschitz-continuous with $\beta = 1$, since $K''(.)$ is assumed to exist and to be uniformly bounded.