Econometric Theory, **22**, 2006, 721–742. Printed in the United States of America.
DOI: 10.1017/S0266466606060336

A STUDY OF A SEMIPARAMETRIC BINARY CHOICE MODEL WITH INTEGRATED COVARIATES

EMMANUEL GUERRE
LSTA Université Paris 6
and
CREST

HYUNGSIK ROGER MOON
University of Southern California

This paper studies a semiparametric nonstationary binary choice model. Imposing a spherical normalization constraint on the parameter for identification purposes, we find that the maximum score estimator and smoothed maximum score estimator are at least $\sqrt{n}$-consistent. Comparing this rate to the convergence rate of the parametric maximum likelihood estimator (MLE), we show that when a normalization restriction is imposed on the parameter, the Park and Phillips (2000, *Econometrica* 68, 1249–1280) parametric MLE converges at a rate of $n^{3/4}$ and its limiting distribution is a mixed normal. Finally, we show briefly how to apply our estimation method to a nonstationary single-index model.

1. INTRODUCTION

Discrete choice econometric models have been widely used in estimating models that involve discrete economic decision problems. When the economic decision is affected by fundamental macroeconomic or financial variables, many of which are known to show nonstationary characteristics, one may need to consider a discrete choice model with nonstationary explanatory variables. Examples of recent applications of this approach include Hu and Phillips (2004a, 2004b), where the authors estimate models of the federal fund rate targeting by the Federal Reserve and of overnight target rate adjustments by the Bank of Canada, respectively.

The first draft of the paper was written while Guerre was visiting the economics department of the University of Southern California. We thank Peter C.B. Phillips, a co-editor, and three anonymous referees for helpful comments and John Dolfin for proofreading. Guerre thanks the economics department of the University of Southern California for its hospitality during his visit. Moon appreciates financial support of the University of Southern California faculty development award. Address correspondence to Hyungsik Roger Moon, KAP 300, Department of Economics, University of Southern California, Los Angeles, CA 90089, USA; e-mail: moonr@usc.edu.

© 2006 Cambridge University Press 0266-4666/06 $12.00

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press

This paper studies a simple binary choice model in which the observable binary sequence y_t is generated by

$$y_t = sgn(y_t^*), \qquad y_t^* = \beta_0' x_t + u_t, \qquad t = 1, \ldots, n, \tag{1.1}$$

where sgn is the sign function (i.e., $sgn(x) = 1$ if $x \geq 0$ and -1 otherwise), x_t is a p-vector of I(1) observable explanatory variables, and u_t is an unobservable variable whose distribution is *unknown*.

When the regressor x_t and the error term u_t in the binary choice model (1.1) are stationary (mostly assumed to be independent and identically distributed (i.i.d.)), and a parametric distribution of u_t is assumed, it is well known that under regularity conditions the maximum likelihood estimator (MLE) of β_0 is $\sqrt{n}$-consistent and asymptotically normal. When the distribution of u_t is not known, various semiparametric estimation methods have been proposed. See, for example, Horowitz (1998) for a survey of these developments and asymptotic properties of the various semiparametric estimators of β_0 in (1.1) with i.i.d. variables.

The study of the nonstationary binary choice model is quite new. When the distribution function of u_t is known and the true parameter β_0 is different from zero, Park and Phillips (2000) show that the MLE of β_0 converges at a rate of $n^{1/4}$ and the limiting distribution of the MLE is a mixed normal. Hu and Phillips (2004b) extend Park and Phillips' nonstationary binary choice model to a multinomial setup and develop an asymptotic theory. When the true parameter β_0 is zero, Guerre and Moon (2002) show that the MLE has a convergence rate of order n. When the distribution of u_t in the nonstationary binary choice model is not known, very little is known about the asymptotic properties of the semiparametric estimators. Recently, Moon (2004) has shown that Manski's maximum score estimator (MSE) (see Manski, 1975, 1985) of β_0 in model (1.1) is consistent but does not derive any consistency rate, a challenging issue in view of the history of this estimator in the i.i.d. framework. An additional concern with the MSE is that standard numerical routines cannot be used to compute the estimator which does not have a limit normal distribution, because of a discontinuous objective function. A standard solution to that issue is the smoothed maximum score estimator (SMSE) introduced in Horowitz (1992).

The main contribution of the paper is to find *a convergence rate* of the MSE and SMSE of the nonstationary binary choice model (1.1). As is well known, β_0 is identified up to scale, and we focus on the estimation of its direction; that is, we restrict ourselves to parameters in the unit sphere to achieve identification. We find that, for such a normalization, the MSE and SMSE can be $\sqrt{n}$-consistent at least, a surprisingly fast rate at first sight, which holds under fairly general conditions for the error term u_t, allowing us in particular to relax standard exogeneity assumptions. This can be explained at an intuitive level by the fact that integrated covariates have an $\sqrt{n}$ divergence rate by a functional central limit theorem (FCLT) and dominate the error terms u_t. Hence the latent

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press

variable y_t^* reduces to $\beta_0' x_t$ asymptotically and the maximum score can detect departures from β_0 within the inverse of the divergence rate $\sqrt{n}$ of the covariate. Another distinctive feature of the integrated framework is that the $\sqrt{n}$-consistency rate of the MSE and the SMSE can be derived with similar techniques. This contrasts with the i.i.d. case where different methods are used to show that the exact convergence rates of the MSE and the SMSE are $n^{1/3}$ and close to $\sqrt{n}$, respectively (e.g., Kim and Pollard, 1990; Horowitz, 1993), which is slower than ours. Second, we compare the $\sqrt{n}$-convergence rate of the semiparametric estimator to the convergence rate of the parametric MLE of the normalized parameter. In the parametric case where the error term has a fully specified distribution such as the standard normal or the logistic, the scale and direction of β_0 are both identified. In such a framework where the parameters are not normalized, Park and Phillips (2000) have shown that the MLE rate is $n^{1/4}$ with an asymptotic mixed normal distribution. We build on Park and Phillips (2000) to show that the MLE of the normalized β_0 converges with a rate $n^{3/4}$. Third, we discuss briefly how to apply our estimation method to a single-index model.

The paper is organized as follows. In Section 2, we introduce the regularity conditions assumed in the paper. In Section 3 we define the MSE and the SMSE. In Section 4 we present the main result on the lower bound of the convergence rate. In Section 5 we compare the main result with the i.i.d. semiparametric results and the Park and Phillips (2000) result. As an application, we also discuss how to apply the estimation method to a single-index model. Section 6 concludes the paper. The Appendix contains all the technical proofs. In what follows, $\|\beta\| = (\sum_{j=1}^{p} \beta_j^2)^{1/2}$ is the standard euclidean norm, so that the unit sphere $\mathbb{S}_p$, which contains β_0, is the set of β with $\|\beta\| = 1$. We use $\to_p$ and $\Rightarrow$ to denote convergence in probability and distribution, respectively.

2. ASSUMPTIONS

First, we introduce an assumption about the integrated explanatory variables x_t. It corresponds to Assumption 1 of Park and Phillips (2000), except that these authors required $r > 8$.

Assumption X. Let $x_t = x_{t-1} + v_t$ with $x_0 = 0$ and where $v_t = \Pi(L)e_t = \sum_{i=0}^{\infty} \Pi_i e_{t-i}$, with $\Pi(1)$ nonsingular and $\sum_{i=0}^{\infty} i\|\Pi_i\| < \infty$. The innovations $\{e_t\}_{t\in\mathbb{Z}}$ are i.i.d. with mean 0 and $E\|e_t\|^r < \infty$ for some $r > 2$, have a distribution that is absolutely continuous with respect to the Lebesgue measure, and have a characteristic function $\varphi_e(t)$ that satisfies $\lim_{\|t\|\to+\infty} \|t\|^{\kappa}\varphi(t) = 0$ for some $\kappa > 0$.

Thanks to a multivariate strong approximation result due to Einhmal (1989) and using the Beveridge–Nelson decomposition of $\{x_t\}_{t\geq 1}$ as in Akonom (1993),

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press

one can reconstruct the variables jointly with a Brownian motion $\{W(t)\}_{t\in\mathbb{R}^+}$ with full-rank covariance matrix Σ such that

$$\max_{1\le t\le n}\|x_t - W(t)\| = o_p(n^{1/r}). \tag{2.1}$$

Because $\max_{1\le t\le n}\|W(t)/\sqrt{n}\| \overset{d}{=} \max_{1\le t\le n}\|W(t/n)\| \le \sup_{r\in[0,1]}\|W(r)\|$, the exact order of $\max_{1\le t\le n}\|x_t\|$ is $\sqrt{n}$.

Next we introduce two different assumptions that deal with the tail behavior and dependence structure of the error term u_t. The main results in Section 4 will be derived under each assumption.

Assumption U1. There exists a sequence $r_n \le Cn^{1/2-\kappa}$, for some $C > 0$ and κ in $(0,\frac{1}{2})$, such that $\max_{1\le t\le n}|u_t| = O_p(r_n)$.

Assumption U2. For all t, the distribution of $|u_t|$ given $x_t, x_{t-1}, \ldots, x_1$ only depends upon $\beta_0' x_t$, that is, $P(|u_t| < u|x_t, x_{t-1}, \ldots, x_1) = F(u;\beta_0' x_t)$, and $\int(1 - F(x;x))\,dx < \infty$.

Assumption U1 corresponds to Assumption 1(iii) in Moon (2004), $\max_{t\in\mathbb{N}^*} E|u_t|^{2+\kappa} < \infty$ for some $\kappa > 0$, yielding that $\max_{1\le t\le n}|u_t| = O_p(n^{1/(2+\kappa)})$ by the Markov inequality. The rate r_n of Assumption U1 is such that $\max_{1\le t\le n}|u_t| = o_p(1)\max_{1\le t\le n}\|x_t\|$. Under Assumption U1, the distribution error term u_t could depend upon time, being, for instance, affected by structural breaks. It also allows that u_t can include some omitted stationary variables z_t, that is, that $u_t = \gamma_0' z_t + \tilde{u}_t$ where $\|z_t\|$ and $\tilde{u}_t$ verify Assumption U1. As noted in Moon (2004), Assumption U1 does not impose any specific form for the conditional distribution of u_t given $\{x_t\}_{t\ge1}$, so that endogeneity and heteroskedasticity are possible.

Assumption U2 restricts the distribution of u_t given $x_t, \ldots, x_1$ to be time homogeneous. The error term u_t can still depend upon the regressor x_t, but only through the index $\beta_0' x_t$, allowing again for endogeneity and heteroskedasticity. For example, consider an error term $u_t = m(\beta_0' x_t) + s(\beta_0' x_t)\eta_t$, where the i.i.d. η_t's are independent of the x_t's. Notice that u_t satisfies Assumption U2 as soon as $\int P(|m(x) + s(x)\eta| \ge |x|)\,dx < \infty$. By the triangular inequality, this holds provided $\int P(|m(x)| + s(x)|\eta| \ge |x|)\,dx < \infty$, a condition that relates the tails of η_t to the divergence rate of $(|x| - |m(x)|)/s(x)$.

Notice that the two assumptions are not nested. The dependence of the u_t's on the time index t and on the covariate x_t is not restricted by Assumption U1, whereas it is under Assumption U2. On the other hand Assumption U2 sometimes allows fatter tails of the distribution of u_t than under Assumption U1, and $\max_{1\le t\le n}|u_t|$ can diverge faster than $\sqrt{n}$. For example, consider an i.i.d. u_t that is independent of $\{x_t\}_{t\ge1}$. Under Assumption U2, u_t has finite first moment because

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press

$$\int (1 - F(x;x))\,dx = \int_0^\infty P(|u_t| \geq x)\,dx = E|u_t| < \infty,$$

but the second moment of u_t may not be finite, so that $\max_{1\leq t\leq n}|u_t|$ can diverge faster than $\sqrt{n}$.

3. MAXIMUM SCORE AND SMOOTHED MAXIMUM SCORE ESTIMATORS

Our semiparametric estimation of β_0 is based upon the SMSE proposed by Horowitz (1992). Define

$$\hat{\beta}_n = \arg\max_{\beta\in\mathbb{S}_p} Q_n(\beta;h_n), \quad \text{where } Q_n(\beta;h_n) = \frac{1}{n}\sum_{t=1}^{n} y_t K\left(\frac{\beta' x_t}{h_n}\right), \tag{3.1}$$

$(K(\cdot)+1)/2$ is a cumulative distributive function, and $h_n > 0$ is a bandwidth. Here $K(x/h_n)$ can be viewed as a suitable approximation of $sgn(x)$. In the particular case of $(K(x)+1)/2 = 1(x>0)$ that gives $K(x) = sgn(x)$, $Q_n(\beta;h_n)$ does not depend upon h_n and is the objective function of the MSE, so that the MSE and SMSE can be studied in the same framework. For $K(x) = sgn(x)$, we write

$$Q_n(\beta) = Q_n(\beta;0) = \frac{1}{n}\sum_{t=1}^{n} y_t\, sgn(\beta' x_t),$$

and then $\hat{\beta}_n$ is the maximum score estimator of Manski (1975). Our last assumption is about $K(\cdot)$ and h_n Note that we allow for a diverging bandwidth, which must remain negligible with respect to the divergence rate $\sqrt{n}$ of the covariates x_t.

Assumption K. The function $(K(\cdot)+1)/2$ from $\mathbb{R}$ to $\mathbb{R}^+$ is a cumulative distribution function with median 0 and $\int_{-\infty}^{\infty}|K(x) - sgn(x)|\,dx < \infty$. The bandwidth sequence verifies $n^{-1/2+\kappa}/C \leq h_n \leq Cn^{1/2-\kappa}$ for some $C > 1$ and κ in $(0,\frac{1}{2}]$.

As in Park and Phillips (2000), it is mathematically convenient to split $\hat{\beta}_n$ into a component $\hat{\alpha}_{1n} = \beta_0'\hat{\beta}_n$ colinear to β_0 and an orthogonal component α_2. To this end, consider a rotation matrix $R = (R_1, R_2)$ with $R_1 = \beta_0$. Set

$$\alpha = R'\beta = [\alpha_1, \alpha_2']', \qquad \hat{\alpha}_n = R'\hat{\beta}_n, \tilde{x}_t = R'x_t = [x_{1t}, x_{2t}']',$$

$$V(t) = R'W(t) = [V_1(t), V_2(t)']', \tag{3.2}$$

so that $\alpha_0 = R'\beta_0 = [1,0]'$, $x_{1t} = \beta_0' x_t = \alpha_0\tilde{x}_t$, $\beta' x_t = \alpha'\tilde{x}_{1t} = \alpha_1 x_{1t} + \alpha_2' x_{2t}$.

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press

4. MAIN RESULTS

Moon (2004) establishes consistency of the MSE. The following theorem establishes consistency of the SMSE.

THEOREM 1. *Under Assumptions X, K, and U1 or U2, $\hat{\beta}_n$ is a consistent estimator of β_0.*

The proof of Theorem 1 follows from suitable modifications of Moon (2004) to cope with the SMSE and is not reported here. We now give a preliminary result that is crucial to derive some lower bounds for the consistency rate of the MSE and SMSE. Define

$$\tilde{Q}_n(\beta;h_n) = \frac{1}{n}\sum_{t=1}^{n} sgn(\beta_0' x_t)K\left(\frac{\beta' x_t}{h_n}\right). \tag{4.1}$$

This objective function is associated with the "idealistic" deterministic model $y_t = sgn(\beta_0' x_t)$, derived from (1.1) by removing the error term u_t.

PROPOSITION 1. *Assume that Assumptions X and K hold. Then, under Assumption U1,*

$$\sup_{\beta\in\mathbb{S}_p} |Q_n(\beta;h_n) - \tilde{Q}_n(\beta;h_n)| = O_p\left(\frac{r_n}{\sqrt{n}}\right) = o_p(1),$$

and under Assumption U2,

$$\sup_{\beta\in\mathbb{S}_p} |Q_n(\beta;h_n) - \tilde{Q}_n(\beta;h_n)| = O_p\left(\frac{1}{\sqrt{n}}\right).$$

Proposition 1 indicates that the error term u_t of the auxiliary binary model (1.1) can be neglected when studying the smoothed score function $Q_n(\beta;h_n)$. Indeed, under Assumption U1, the integrated covariates x_t, $1 \le t \le n$, diverge faster than the maximum of error terms $\max_{1\le t\le n}|u_t|$, so that setting the u_t's to 0 intuitively yields a suitable approximation $\tilde{Q}_n(\beta;h_n)$ for $Q_n(\beta;h_n)$. That this still holds under Assumption U2 comes from the fact that the difference $Q_n(\beta;h_n) - \tilde{Q}_n(\beta;h_n)$ can be bounded by an empirical mean of the nonlinear transformation of $\beta' x_t$, which can be studied through local time limit results developed in Park and Phillips (1999, 2001). This gives the better order $1/\sqrt{n}$ for the uniform norm of $Q_n(\beta;h_n) - \tilde{Q}_n(\beta;h_n)$ in Proposition 1. Proposition 1, together with suitable lower bounds for $\tilde{Q}_n(\beta;h_n) - \tilde{Q}_n(\beta_0;h_n)$ given by Lemma 1 and Proposition 2 in the Appendix, shows that $\tilde{Q}_n(\beta;h_n)$ is sensitive to small departures of β from β_0. As a consequence, the next results show that the MSE and SMSE can achieve an $\sqrt{n}$-rate under Assumption U2.

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press

THEOREM 2. *Assume that Assumptions X and K hold.*

(i) Under Assumption U1, $(\sqrt{n}/\max(r_n, h_n))\hat{\alpha}_{2n} = O_p(1)$ *and* $(n/\max^2(r_n, h_n))(\hat{\alpha}_{1n} - 1) = O_p(1)$.

(ii) Under Assumption U2, $(\sqrt{n}/\max(1, h_n))\hat{\alpha}_{2n} = O_p(1)$ *and* $(n/\max^2(1, h_n))(\hat{\alpha}_{1n} - 1) = O_p(1)$.

As a consequence, we immediately get the following corollary.

COROLLARY 3. *Assume that Assumptions X, U2, and K hold with* $h_n = h > 0$. *Then* $\hat{\beta}_n$ *is an* $\sqrt{n}$*-consistent estimate of* β_0.

Remark 1. Note that Theorem 2 always gives a faster rate for $\hat{\alpha}_{1n} - 1$ than for $\hat{\alpha}_{2n}$. This is actually due to the spherical normalization that imposes that $\hat{\alpha}_{1n}^2 + \|\hat{\alpha}_{2n}\|^2 = 1$. Indeed, because $\hat{\alpha}_{1n} = 1 + o_p(1)$ by Theorem 1, $\hat{\alpha}_{1n} - 1 = (\hat{\alpha}_{1n}^2 - 1)/(\hat{\alpha}_{1n} + 1) = \|\hat{\alpha}_{2n}\|^2/[2 + o_p(1)] = o_p(1)\|\hat{\alpha}_{2n}\|$, so that $\hat{\alpha}_{1n} - 1$ goes to 0 faster than $\hat{\alpha}_{2n}$. However, such a result cannot be used in practice because $\hat{\alpha}_{1n} = \beta_0'\hat{\beta}_n$ depends upon the unknown parameter β_0. This means that the relevant rate from a statistical point of view is the rate derived for $\hat{\beta}_n$, which is the slowest one as seen from the corollary.

Remark 2. Theorem 2 holds under rather weak conditions on the error terms u_t of (1.1), which can depend on t and on the x_t's. Moon (2004) has already noted that the usual error conditions of the i.i.d. setup can be relaxed for consistency in the presence of integrated covariates (see Theorem 1 in our paper). See also Chang and Park (2004) and de Jong (2002) for similar results in the context of nonlinear regression models with integrated covariates.

Remark 3. Smoothing does not seem to improve the convergence rates of the SMSE with respect to the MSE in Theorem 2. However, the rates given in Theorem 2 are only lower bounds, meaning that the MSE and SMSE can converge faster. In a simple setup we discuss the issue of the exact rate of the SMSE, whose limit behavior can be studied with a standard Taylor expansion of the objective function. To simplify the discussion, take $h = 1$ and assume that x_t is bivariate, that $\beta_0' = (1,0) = \alpha_0'$ so that the rotation matrix in (3.2) is identity, and that $x_t = (x_{1t}, x_{2t})'$. We use an angular representation of β, $\beta' = (\cos\varphi, \sin\varphi)$. The first and second derivatives of the objective function with respect to φ are, because $\beta' x_t = x_{1t}\cos\varphi + x_{2t}\sin\varphi$,

$$Q_n^{(1)}(\beta;1) = \frac{\partial Q_n(\beta;1)}{\partial\varphi} = \frac{1}{n}\sum_{t=1}^{n} y_t(x_{2t}\cos\varphi - x_{1t}\sin\varphi)K'(\beta' x_t),$$

$$Q_n^{(2)}(\beta;1) = \frac{\partial^2 Q_n(\beta;1)}{\partial^2\varphi} = \frac{1}{n}\sum_{t=1}^{n} y_t[(x_{2t}\cos\varphi - x_{1t}\sin\varphi)^2 K''(\beta' x_t) - (x_{1t}\cos\varphi + x_{2t}\sin\varphi)K'(\beta' x_t)].$$

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press

Let $\hat{\varphi}_n$ be the SMSE of φ, so that a Taylor expansion gives $Q^{(1)}(\hat{\beta}_n;1) = 0 = Q^{(1)}(\beta_0;1) + (\hat{\varphi}_n - \varphi_0)Q^{(2)}(\bar{\beta}_n;1)$ where $\bar{\beta}_n$ is a unit vector of angle $\bar{\varphi}_n$ between φ_0 and $\hat{\varphi}_n$. Hence, if $Q^{(2)}(\bar{\beta}_n;1) = Q^{(2)}(\beta_0;1)(1 + o_p(1))$ as could be expected from similar expansion in Park and Phillips (2000) and Chang and Park (2003),

$$\hat{\varphi}_n - \varphi_0 = \frac{Q_n^{(1)}(\beta_0;1)}{Q_n^{(2)}(\beta_0;1)}(1 + o_p(1)), \quad \text{with}$$

$$Q_n^{(1)}(\beta_0;1) = \frac{1}{n}\sum_{t=1}^{n} y_t x_{2t} K'(x_{1t}),$$

$$Q_n^{(2)}(\beta_0;1) = \frac{1}{n}\sum_{t=1}^{n} y_t [x_{2t}^2 K''(x_{1t}) - x_{1t} K'(x_{1t})].$$

Under Assumption U2, the limit behavior of these sums can be derived from Park and Phillips (1999, 2000, 2001). Introduce $\mu(x_1) = E[y_t | x_{1t} = x_1] = 1 - 2P(u_t \geq -x_1)$, $\sigma^2(x_1) = V[y_t | x_{1t} = x_1] = 4P(u_t \geq -x_1)P(u_t < -x_1)$, consider a Brownian motion $B(\cdot)$ independent of $W'(\cdot) = (W_1(\cdot), W_2(\cdot))$, and let $L_1(\cdot, r)$ be the local time of $\{W_1(t)\}_{t\in[0,r]}$. Then

$$\begin{bmatrix} \frac{1}{n}\sum_{t=1}^{n} \mu(x_{1t}) x_{2t} K'(x_{1t}) \\ \frac{1}{n^{1/4}}\sum_{t=1}^{n} (y_t - \mu(x_{1t})) x_{2t} K'(x_{1t}) \\ \frac{1}{n^{3/2}}\sum_{t=1}^{n} y_t [x_{2t}^2 K''(x_{1t}) - x_{1t} K'(x_{1t})] \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \int \mu(x_1) K'(x_1)\, dx_1 \int_0^1 W_2(r)\, dL_1(0,r) \\ \left(\int \sigma^2(x_1) K'^2(x_1)\, dx_1 \int_0^1 W_2^2(r)\, dL_1(0,r) \right)^{1/2} B(1) \\ \int \mu(x_1) K''(x_1)\, dx_1 \int_0^1 W_2^2(r)\, dL_1(0,r) \end{bmatrix}, \tag{4.2}$$

where the first two terms correspond to a decomposition of $Q_n^{(1)}(\beta_0;1)$ given by $y_t = \mu(x_{1t}) + y_t - \mu(x_{1t})$. Hence, we expect that the distribution of $\sqrt{n}(\hat{\varphi}_n - \varphi_0)$ should be close to

$$\frac{\int \mu(x_1) K'(x_1)\, dx_1 \int_0^1 W_2(r)\, dL_1(0,r) + n^{-1/4}\left(\int \sigma^2(x_1) K'^2(x_1)\, dx_1 \int_0^1 W_2^2(r)\, dL_1(0,r) \right)^{1/2} B(1)}{\int \mu(x_1) K''(x_1)\, dx_1 \int_0^1 W_2^2(r)\, dL_1(0,r)},$$

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press

which does not asymptotically vanish if $\int \mu(x_1)K'(x_1)\,dx_1 \neq 0$, in which case $\sqrt{n}$ is the exact rate of the SMSE $\hat{\beta}_n$. On the other hand, $\int \mu(x_1)K'(x_1)\,dx_1 = 0$ if the distribution of u_t given x_{1t} is symmetric, an assumption that is quite natural in the i.i.d. case. In this case, a faster rate is expected, which could be the MLE rate $n^{-3/4}$. However, deriving such a rate would necessitate studying the first sum of (4.2), which writes as $\sum_{t=1}^{n} f(x_{1t})x_{2t}$ with $\int f(x_1)\,dx_1 = 0$, a case not covered by Park and Phillips (1999, 2000, 2001). Results showing that $\sum_{t=1}^{n} f(x_{1t}) = O_p(n^{1/4})$ can be found in Borodin and Ibragimov (1995) but are limited to the case of a random walk. Because x_{2t} must be standardized by $\sqrt{n}$, one can expect that $\sum_{t=1}^{n} f(x_{1t})x_{2t}$ is of order $n^{3/4}$, in which case the exact rate of the SMSE would be $n^{3/4}$.

5. DISCUSSION

5.1. Comparison with i.i.d. Semiparametric Models

There are several remarks we would like to highlight when comparing our results to the i.i.d. case. First, we know that the convergence rates of the MSE and the SMSE in the i.i.d. binary choice model are $n^{1/3}$ (Kim and Pollard, 1990) and close to $\sqrt{n}$ (Horowitz, 1992), respectively. When the regressors are nonstationary, according to our finding, the convergence rates of the MSE and the SMSE are faster than those of the i.i.d. binary choice model. This is due to the diverging nature of integrated covariates as discussed after Proposition 1. Second, note that the bandwidth of Corollary 3 is kept constant, which is different from the i.i.d. setting. Indeed, considering $h_n \to \infty$ as studied in Phillips and Park (1998) or $h_n \to 0$ as is usual in the i.i.d. setting will not improve the rate derived from Theorem 2(ii). Because the covariates diverge, taking a constant h means a certain selection of the observations, and the usefulness of a vanishing h_n deserves further investigation.

5.2. Comparison with Parametric Nonstationary Binary Choice Model

In this section we compare the result in Theorem 2 and Corollary 3 with the asymptotic results of the parametric nonstationary binary choice model in Park and Phillips (2000). For this, we first summarize the main results in Park and Phillips. The parametric nonstationary binary choice model investigated by Park and Phillips is

$$y_t = sgn(y_t^*), \qquad y_t^* = \theta_0' x_t - \eta_t, \qquad \theta_0 \in \mathbb{R}^p \tag{5.1}$$

where the regressor x_{t+1} is adapted to some filtration $\mathcal{F}_t$, with respect to which the error term η_t is measurable. Given $x_t, \ldots, x_0$, it is assumed that the error term η_t has a known standardized cumulative distribution function F, as, for instance, the standard normal (leading to the probit model) and the standard

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press

logistic (leading to the logit model). The covariate x_t is integrated and obeys Assumption 1 in Park and Phillips, that is, Assumption X with $r > 8$. Also, following Park and Phillips, we assume that F satisfies Assumption 2 in Park and Phillips, which is used to derive the asymptotic distribution of the maximum likelihood (ML) estimator $\hat{\theta}_n$. Finally, we assume that, as in the logit and probit cases, $\log F(x)$ and $\log(1 - F(x))$ are concave to ensure uniqueness of the MLE and that $\theta_0 \neq 0$. Let R be as in (3.2) and $\gamma = (\gamma_1, \gamma_2')' = R'\theta$. Then, $\gamma_0 = ((\theta_0'\theta_0)^{1/2}, 0)'$. Let $\hat{\theta}_n$ be the MLE of model (5.1). Denote $\hat{\gamma}_n = (\hat{\gamma}_{1n}, \hat{\gamma}_{2n}')' = R'\hat{\theta}_n$. According to Theorem 2 in Park and Phillips, $\hat{\gamma}_n \to_p \gamma_0$ and

$$\begin{pmatrix} n^{1/4}(\hat{\gamma}_{1n} - (\theta_0'\theta_0)^{1/2}) \\ n^{3/4}\hat{\gamma}_{2n} \end{pmatrix} \Rightarrow MN(0, Q) = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}, \quad \text{say, with } Q = [q_{ij}]_{1 \leq i,j \leq 2}, \tag{5.2}$$

where MN denotes a mixed normal distribution with a random covariance matrix Q defined in equation (18) of Park and Phillips. This gives

$$n^{1/4}(\hat{\theta} - \theta_0) \Rightarrow MN(0, R_1 R_1'(q_{11} - q_{12} q_{22}^{-1} q_{21})^{-1}).$$

We now turn to the comparison of estimation rates in the parametric and semiparametric binary models. Note that our semiparametric model uses a different identifying condition from the parametric model of Park and Phillips (2000). In Park and Phillips, it is assumed that η_t has a *standardized* conditional known distribution, whereas in model (1.1) the scale of the error term is unknown, so that it is assumed that β_0 is in the unit sphere S_p, that is, that $\beta_0'\beta_0 = 1$.[1] A fair comparison of the MSE and MLE rates must, therefore, compare the estimation rates of β_0 and $\theta_0/(\theta_0'\theta_0)^{1/2}$, respectively.

Notice that the MLE of the normalized parameter $\theta_0/(\theta_0'\theta_0)^{1/2}$ is $\hat{\theta}_n/(\hat{\theta}_n'\hat{\theta}_n)^{1/2}$. Premultiplying the rotating matrix R', we write $R'\hat{\theta}_n/(\hat{\theta}_n'RR'\hat{\theta}_n)^{1/2} = (\hat{\gamma}_{1n}^2 + \hat{\gamma}_{2n}'\hat{\gamma}_{2n})^{-1/2}(\hat{\gamma}_{1n}, \hat{\gamma}_{2n}')'$. Comparing this to the notation in Section 3, $(\hat{\gamma}_{1n}^2 + \hat{\gamma}_{2n}'\hat{\gamma}_{2n})^{-1/2}\hat{\gamma}_{1n}$ and $(\hat{\gamma}_{1n}^2 + \hat{\gamma}_{2n}'\hat{\gamma}_{2n})^{-1/2}\hat{\gamma}_{2n}$ correspond to $\hat{\alpha}_{1n}$ and $\hat{\alpha}_{2n}$ in (3.2), respectively.

From (5.2), we have $\hat{\gamma}_{1n} \to_p (\theta_0'\theta_0)^{1/2} > 0$. Therefore, $\hat{\gamma}_{1n} > 0$ with probability approaching one. Define $A_{1n} = n^{1/4}(\hat{\gamma}_{1n} - (\theta_0'\theta_0)^{1/2})$, $A_{2n} = n^{3/4}\hat{\gamma}_{2n}$, which are both $O_p(1)$. Because $(1 + x)^{-1/2} = 1 - (x/2) + O(x^2)$ and $(1 + x)^{-2} = 1 - 2x + O(x^2)$, we have

$$\begin{aligned} \frac{\hat{\gamma}_{1n}}{(\hat{\gamma}_{1n}^2 + \hat{\gamma}_{2n}'\hat{\gamma}_{2n})^{1/2}} &= \left(1 + \frac{\hat{\gamma}_{2n}'\hat{\gamma}_{2n}}{\hat{\gamma}_{1n}^2}\right)^{-1/2} \\ &= 1 - \frac{1}{2}\frac{n^{-3/2}A_{2n}'A_{2n}}{\theta_0'\theta_0[1 + n^{-1/4}A_{1n}(\theta_0'\theta_0)^{-1/2}]^2} + O_p(n^{-3}) \\ &= 1 - \frac{1}{2}\frac{n^{-3/2}A_{2n}'A_{2n}}{\theta_0'\theta_0} + O_p(n^{-7/4}), \end{aligned}$$

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press

which leads to

$$n^{3/2}\left(\frac{\hat{\gamma}_{1n}}{(\hat{\gamma}_{1n}^2+\hat{\gamma}_{2n}'\hat{\gamma}_{2n})^{1/2}}-1\right)=\frac{1}{2}\,\frac{A_{2n}'A_{2n}}{(\theta_0'\theta_0)^{1/2}}+o_p(1). \tag{5.3}$$

Also, we have $\hat{\gamma}_{2n}(\hat{\gamma}_{1n}^2+\hat{\gamma}_{2n}'\hat{\gamma}_{2n})^{-1/2}=n^{-3/4}A_{2n}(\theta_0'\theta_0)^{-1/2}+O_p(n^{-1})$, which leads to

$$n^{3/4}\,\frac{\hat{\gamma}_{2n}}{(\hat{\gamma}_{1n}^2+\hat{\gamma}_{2n}'\hat{\gamma}_{2n})^{1/2}}=\frac{A_{2n}}{(\theta_0'\theta_0)^{1/2}}+o_p(1). \tag{5.4}$$

Finally, from (5.3) and (5.4), we deduce that

$$n^{3/4}\left(\frac{\hat{\theta}_n}{(\hat{\theta}_n'\hat{\theta}_n)^{1/2}}-\frac{\theta_0}{(\theta_0'\theta_0)^{1/2}}\right)=R_2\,\frac{A_{2n}}{(\theta_0'\theta_0)^{1/2}}+o_p(1)$$
$$\Rightarrow MN(0,(\theta_0'\theta_0)^{-1}R_2(q_{22}-q_{21}q_{11}^{-1}q_{12})^{-1}R_2').$$

Summarizing this gives the following theorem.

THEOREM 4. *Assume Assumptions 1 and 2 in Park and Phillips (2000). Then, the MLE* $\hat{\theta}_n/(\hat{\theta}_n'\hat{\theta}_n)^{1/2}$ *of the normalized parameter* $\theta_0/(\theta_0'\theta_0)^{1/2}$ *is* $n^{3/4}$*-consistent, and its limit distribution is*

$$n^{3/4}\left(\frac{\hat{\theta}_n}{(\hat{\theta}_n'\hat{\theta}_n)^{1/2}}-\frac{\theta_0}{(\theta_0'\theta_0)^{1/2}}\right)\Rightarrow MN(0,(\theta_0'\theta_0)^{-1}R_2(q_{22}-q_{21}q_{11}^{-1}q_{12})^{-1}R_2').$$

This result complements Park and Phillips (2000) by showing that the MLE of the standardized direction $\theta_0/(\theta_0'\theta_0)^{1/2}$ converges with a rate $n^{3/4}$, faster than the $n^{1/4}$-consistency rate of $\hat{\theta}_n$. This shows that $(\theta_0'\theta_0)^{1/2}$ is difficult to estimate and slow down the MLE estimator. Indeed, the MLE $(\hat{\theta}_n'\hat{\theta}_n)^{1/2}$ converges with the slower rate $n^{1/4}$.

In comparison with the MLE, the MSE of β_0, which is comparable to $\theta_0/(\theta_0'\theta_0)^{1/2}$, converges with the $\sqrt{n}$ rate at least (see Corollary 3). However, this rate is derived under much weaker conditions than for the parametric specification of Park and Phillips.

5.3. Estimation of a Single-Index Model

In this section, we discuss briefly how to apply the MSE or SMSE estimation in the previous section to a nonstationary single-index model. In estimating a nonlinear nonstationary regression model, one critical problem is the curse of dimensionality of the nonstationary regressor, as it is in nonparametric statistics. Indeed, the p-dimensional Brownian motion $W(t)$, which is the asymptotic counterpart of the nonstationary regressor x_t in (2.1), is transient as soon

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press

as the dimension p is larger than three, and it has a slow logarithmic recurrence rate when $p = 2$. This prohibits accurate estimation of some parametric models. As a consequence, the existing studies on nonlinear modeling with integrated covariates are often limited to univariate explanatory variables; see for example, Park and Phillips (2001). A possible remedy to the curse of dimensionality is, as in nonparametrics, to consider index-based models. The parametric binary choice model of Park and Phillips (2000) can be interpreted as an index model with a known link function, as shown from their equation (3). See also Chang and Park (2003) and Hu and Phillips (2004a, 2004b).

Suppose that observations z_t are generated by the following semiparametric single-index model:[2]

$$z_t = G(\beta_0' x_t) + \varepsilon_t. \tag{5.5}$$

In model (5.5), x_t is a p-vector of an integrated process, ε_t is a stationary error, and $G(\cdot)$ is an *unknown* monotone link function that is assumed to be increasing. Again, we impose a normalization restriction $\beta_0'\beta_0 = 1$ to ensure identification. Suppose that the goal is to estimate β_0.

Assume that the link function $G(x)$ diverges when $|x|$ goes to infinity, so that the generalized inverse $G^-(y) = \inf\{x; G(x) \geq y\}$ is defined for all values of y. Now denote $y_t = sgn(z_t)$. Then, by taking the *sgn* operator in (5.5), we transform the single-index model (5.5) into a binary model $y_t = sgn(G(\beta_0' x_t) + \varepsilon_t) = sgn(\beta_0' x_t + G^-(\varepsilon_t))$. Set $G^-(\varepsilon_t) = u_t$ and assume that u_t satisfies Assumption U1 or U2. Then, according to Corollary 3, the MSE or SMSE of β_0 in the resulting binary model is $\sqrt{n}$-consistent.

6. CONCLUSION

This paper studies estimation of the regression parameter of a semiparametric binary choice model with integrated regressors x_t, $t = 1, \ldots, n$, under a spherical normalization of the parameter. We show that the MSE and SMSE can achieve an $\sqrt{n}$-consistency rate under fairly general conditions on the error terms u_t, including the usual case of i.i.d. u_t but relaxing standard exogeneity conditions. This carries over for the estimation of a single-index model, a specification that can be used to circumvent the curse of dimensionality resulting from integrated regressors. A key ingredient is the divergence of the x_t's, a fact that contrasts with the standard i.i.d. case. Indeed, the integrated x_t's are of order $\sqrt{n}$ by a FCLT, so that the error terms u_t can be asymptotically neglected in the objective functions of the MSE and SMSE. Because $\beta' x_t$ is asymptotically sensitive to variations of β of magnitude $1/\sqrt{n}$, the objective functions asymptotically detect departures from the true parameter β_0 of order $1/\sqrt{n}$ and our estimators are $\sqrt{n}$-consistent at least. Such a sketchy derivation suggests that finding consistency rates is possible for more general specification of the binary choice model, including, for instance, polynomial functions of the covari-

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press

ates in the regression. A discussion of the exact asymptotic behavior of the SMSE suggests that the SMSE can converge faster than $\sqrt{n}$ when the error term is symmetric given the covariate, a condition considered natural in the i.i.d. setup. However a formal derivation of such a result would necessitate extending the asymptotic theory of Park and Phillips (1999, 2001) for integrated processes; we leave this issue as a further research topic.

NOTES

1. The two identifying conditions can be related as follows. We have in (5.1) $y_t = sgn(\theta_0' x_t - \eta_t) = sgn\{[\theta_0'/(\theta_0'\theta_0)^{1/2}]x_t - [\eta_t/(\theta_0'\theta_0)^{1/2}]\}$. Therefore, the inverse norm $1/(\theta_0'\theta_0)^{1/2}$ of θ_0 in the binary model (5.1) can be interpreted as the unknown scaling parameter of the error term in the semiparametric specification (1.1).

2. Notice that there is large literature on single-index models for stationary time series data. To mention an example of earlier work, Robinson (1982) investigated asymptotic properties of the Gaussian quasi maximum likelihood estimator of a censored Tobit model, which is a special case of the single-index model with a known link function, assuming general Gaussian stationary errors and stationary covariates. However, he assumes independence of the error and the covariates. We thank a referee for pointing us to Robinson's work.

REFERENCES

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

Borodin, A.N. & I.A. Ibragimov (1995) *Limit Theorems for Functionals of Random Walks*. Proceedings of the Steklov Institute of Mathematics, vol. 195. American Mathematical Society.

Chang, Y. & J.Y. Park (2003) Index models with integrated time series. *Journal of Econometrics* 114, 73–106.

Chang, Y. & J.Y. Park (2004) Endogeneity in Nonlinear Regression with Integrated Time Series. Mimeo, Yale University.

de Jong, R.M. (2002) Nonlinear Estimators with Integrated Regressors but without Exogeneity. Mimeo, Ohio State University.

Einhmal, U. (1989) Extensions of results of Komlós, Major and Tusnády to the multivariate case. *Journal of Multivariate Analysis* 28, 20–68.

Guerre, E. & H.R. Moon (2002) A note on the nonstationary binary choice logit model. *Economics Letters* 76, 267–271.

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

Horowitz, J.L. (1993) Optimal rates of convergence of parameter estimators in the binary response model with weak distributional assumptions. *Econometric Theory* 9, 1–18.

Horowitz, J.L. (1998) *Semiparametric Methods in Econometrics*. Lecture Notes in Statistics 131. Springer-Verlag.

Hu, L. & P.C.B. Phillips (2004a) Dynamics of the federal funds target rate: A nonstationary discrete choice approach. *Journal of Applied Econometrics* 19(7), 851–867.

Hu, L. & P.C.B. Phillips (2004b) Nonstationary discrete choice. *Journal of Econometrics* 120(1), 103–138.

Kim, J. & D. Pollard (1990) Cube root asymptotics. *Annals of Statistics* 18, 191–219.

Manski, C. (1975) Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics* 3, 205–228.

Manski, C. (1985) Semiparametric analysis of discrete response. *Journal of Econometrics* 27, 313–333.

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press

Moon, H.R. (2004) Maximum score estimation of a nonstationary binary choice model. *Journal of Econometrics* 122, 385–403.
Park, J.Y. & P.C.B. Phillips (1999) Asymptotic for nonlinear transformation of integrated time series. *Econometric Theory* 15, 269–298.
Park, J. & P.C.B. Phillips (2000) Nonstationary binary choice. *Econometrica* 68, 1249–1280.
Park, J.Y. & P.C.B. Phillips (2001) Nonlinear regressions with integrated time series. *Econometrica* 69, 117–161.
Phillips, P.C.B. & J.Y. Park (1998) Nonstationary Density Estimation and Kernel Autoregression. Cowles Foundation Discussion Paper 1181.
Robinson, P.M. (1982) On the asymptotics properties of estimators of models containing limited dependent variables. *Econometrica* 50, 27–42.
Revuz, Y. and M. Yor (1999) *Continuous Martingales and Brownian Motion*, 3rd ed. Springer-Verlag.

APPENDIX: Technical Proofs

To avoid notational burden, we sometimes abbreviate $Q_n(\beta) = Q_n(R\alpha)$ into $Q_n(\alpha)$, where R is from (3.2). In addition to the main statements given in the body of the paper, we will prove the following two intermediate results.

LEMMA 1. *Let $\widetilde{Q}_n(\beta; h_n)$ be the objective function in (4.1). Define, under Assumption K, $k(z) = 2\inf_{x;|x|\geq z}|K(x)|$ and $\widetilde{Q}_n^c(\beta) = \widetilde{Q}_n^c(\beta; z) = (1/n)\sum_{t=1}^n 1\{z \leq |\beta' x_t|\}$ $1\{sgn(\beta_0' x_t) \neq sgn(\beta' x_t)\}$. Then, under Assumption K, $\widetilde{Q}_n(\beta; h_n) \leq 1 - k(z)\widetilde{Q}_n^c(\beta)$, for any β in $\mathbb{S}_p$, any $z \geq 0$. Under Assumptions X and K, $\widetilde{Q}_n(\beta_0; h_n) = 1 + O_p(h_n/\sqrt{n})$.*

PROPOSITION 2. *Assume that Assumptions X and K hold and fix $z \geq 0$ in Lemma 1. Let $\gamma_n = \max(r_n, h_n)$ under Assumption U1 and $\gamma_n = 1$ under Assumption U2. Define $\bar{Q}_n(\lambda) = \sup_{\alpha\in\mathbb{S}_p;\|\alpha_2\|\geq\gamma_n n^{-1/2}\lambda}(1 - k(z)\widetilde{Q}_n^c(\alpha))$, where $\|\alpha_2\|$ is the euclidean norm of α_2. Then, under Assumption U1 or U2, for any $\epsilon > 0$ one can find a C_ϵ such that, for any $\lambda > 0$, $P(\bar{Q}_n(\lambda) \leq 1 - C_\epsilon(\lambda\gamma_n/\sqrt{n})) \geq 1 - \epsilon$.*

A.1. Preliminary Results.

Define $\Psi_k = \sum_{i=0}^k \Pi_j$, $\Psi = \sum_{i=0}^\infty \Pi_j = \Pi(1)$. This gives, for all $c_n \in \mathbb{R}$, $\zeta \in \mathbb{R}^p$

$$x_{1t} + c_n \frac{\zeta' x_t}{\sqrt{n}} = \left(\beta_0 + \frac{c_n}{\sqrt{n}}\,\zeta\right)' \left[\sum_{k=0}^{t-1} \Psi_k e_{t-k} + \sum_{k=0}^{\infty} (\Psi_{k+t} - \Psi_k) e_{-k}\right]. \tag{A.1}$$

We begin with a slight modification of Theorem 4 of Akonom (1993). Let $\zeta \in \mathbb{R}^p$ be a fixed vector and $\{b_n\}_{n\geq 1}$ $\{c_n\}_{n\geq 1}$ be two sequences of positive real numbers. Define $N_n^{x_1,\zeta}(a,b) = \sum_{t=1}^n 1\{a \leq x_{1t} + c_n(\zeta' x_t/\sqrt{n}) \leq b\}$, $N_n^{x_1}(a,b) = \sum_{t=1}^n 1\{a \leq x_{1t} \leq b\}$, and $N_n^{V_1}(a,b) = \sum_{t=1}^n 1\{a \leq V_1(t) \leq b\}$.

LEMMA 2. *Assume that $n^{-1/2+\kappa}/C \leq b_n \leq Cn^{1/2-\kappa}$, $0 \leq c_n \leq Cn^{1/2-\kappa}$ for some $C > 1$ and κ in $(0, \frac{1}{2}]$. Let ζ be a fixed vector of $\mathbb{R}^p$. Then under Assumption X, $N_n^{x_1,\zeta}(a, a+b_n) = N_n^{V_1}(a, a+b_n)(1 + o_p(1))$, $N_n^{x_1}(a, a+b_n) = N_n^{V_1}(a, a+b_n)(1 + o_p(1))$, and $(1/b_n\sqrt{n})N_n^{V_1}(a, a+b_n) = [(1 + o_p(1))/b_n\sqrt{n}]\int_0^n 1\{a \leq V_1(t) \leq a + b_n\}\,dt \Rightarrow L_1(0,1)$. Moreover, if $b_n = b > 0$ for all n, $(N_n^{x_1}(a, a+b) - N_n^{V_1}(a, a+b))/\sqrt{n} =$*

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press

$O_p(n^{-(1/6)+(2/3r)+\kappa'})$, *for all* $\kappa' > 0$. *If* $f_t^{x_1,\zeta}(\cdot) = f_{tm}^{x_1,\zeta}(\cdot)$ *is the probability density function of* $t^{-1/2}(x_{1t} + c_n(\zeta' x_t/\sqrt{n}))$, *then*

$$\sup_{y\in\mathbb{R}} \left| f_t^{x_1,\zeta}(y) - \frac{1}{\sqrt{2\pi}\sigma_0^2} \exp\left(-\frac{y^2}{2\sigma_0^2}\right)\right| \to 0, \qquad \sup_{t\geq 1}\sup_{y\in\mathbb{R}} f_t^{x_1,\zeta}(y) < \infty. \tag{A.2}$$

The proof of Lemma 2 works by straightforward modifications of the arguments used in Akonom (1993) and is not reported here. The next two lemmas provide some useful consequences of Lemma 2.

LEMMA 3. *Assume that Assumption X holds and let* b_n *and* c_n *be as in Lemma 2. Let* $\zeta \in \mathbb{R}^p$ *be a fixed vector, and a a fixed real number, and* $f(\cdot)$ *be a generalized Riemann integrable bounded function from* $\mathbb{R}$ *to* $\mathbb{R}$ *with* $\int |f(x)|\,dx < \infty$. *Then*

$$\frac{1}{b_n\sqrt{n}} \sum_{t=1}^{n} f\left(\frac{x_{1t} + c_n \dfrac{\zeta' x_t}{\sqrt{n}} - a}{b_n}\right) \Rightarrow L_1(0,1) \int f(x)\,dx.$$

Proof of Lemma 3. Because $f(\cdot)$ is Rieman integrable in the generalized sense, one can, for any $\epsilon > 0$, find $I = I_\epsilon \in \mathbb{N}$ and $d = d_\epsilon > 0$ such that, using (A.2),

$$\int |f(x) - f_I(x)|\,dx \leq \frac{\epsilon}{3 \sup_t \sup_y |f_t^{x_1,\zeta}(y)|}, \tag{A.3}$$

where $f_I(x) = \sum_{i=-I}^{I} f(d_i) 1\{(d_i \leq x < d_{i+1}\}$, $d_i = d \times i$. Now (A.2) and $x = (\sqrt{t}y - a)/b_n$ yield

$$\begin{aligned}
&\frac{1}{b_n} E\left| f\left(\frac{x_{1t} + c_n \dfrac{\zeta' x_t}{\sqrt{n}} - a}{b_n}\right) - f_I\left(\frac{x_{1t} + c_n \dfrac{\zeta' x_t}{\sqrt{n}} - a}{b_n}\right)\right| \\
&\quad = \frac{1}{b_n} \int \left| f\left(\frac{\sqrt{t}y - a}{b_n}\right) - f_I\left(\frac{\sqrt{t}y - a}{b_n}\right)\right| f_t^{x_1,\zeta}(y)\,dy \\
&\quad = \frac{1}{\sqrt{t}} \int |f(x) - f_I(x)| f_t^{x_1,\zeta}\left(\frac{b_n x + a}{\sqrt{t}}\right) dx \leq \frac{\epsilon}{3\sqrt{t}}.
\end{aligned}$$

This gives

$$\begin{aligned}
&E\left| \frac{1}{b_n\sqrt{n}} \sum_{t=1}^{n} \left[f\left(\frac{x_{1t} + c_n \dfrac{\zeta' x_t}{\sqrt{n}} - a}{b_n}\right) - f_I\left(\frac{x_{1t} + c_n \dfrac{\zeta' x_t}{\sqrt{n}} - a}{b_n}\right)\right]\right| \\
&\quad \leq \frac{\epsilon}{3} \frac{1}{\sqrt{n}} \int_0^{n-1} \frac{dt}{\sqrt{t}} \leq \epsilon,
\end{aligned} \tag{A.4}$$

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press

for n large enough. Lemma 2 and (A.3) yield now, because I is finite,

$$\frac{1}{b_n\sqrt{n}}\sum_{t=1}^{n} f_I\left(\frac{x_{1t}+c_n\dfrac{\zeta' x_t}{\sqrt{n}}-a}{b_n}\right) = \sum_{i=-I}^{I}\frac{f(d_i)}{b_n\sqrt{n}}N^{x_1,\zeta}(a+b_n d_i, a+b_n d_{i+1})$$

$$= (1+o_p(1))\sum_{i=-I}^{I}\frac{f(d_i)}{b_n\sqrt{n}}N^{V_1}(a+b_n d_i, a+b_n d_{i+1})$$

$$\Rightarrow L_1(0,1)\sum_{i=-I}^{I} f(d_i)(d_{i+1}-d_i) = L_1(0,1)\int f_I(x)\,dx,$$

where the last limit holds because the last limit of Lemma 2 holds jointly with respect to the d_i's, $i=-I,\ldots,I$. Taking $\epsilon>0$ small enough in (A.3) shows that Lemma 3 is proved. ■

LEMMA 4. *Consider c_n with $n^{-1/2+\kappa}/C \le c_n \le Cn^{1/2-\kappa}$ for some $C>0$ and κ in $(0,\frac{1}{2}]$, $\lambda>0$ and z in $\mathbb{R}$, and $\delta\in\mathbb{S}_{p-1}$. Then*

$$\frac{1}{c_n\sqrt{n}}\sum_{t=1}^{n}\left[1\left\{0\le x_{1t}<-z-\frac{c_n\lambda\delta' x_{2t}}{\sqrt{n}}\right\}+1\left\{z-\frac{c_n\lambda\delta' x_{2t}}{\sqrt{n}}\le x_{1t}<0\right\}\right]$$

converges in distribution to $\lambda\int_0^1|\delta' V_2(r)|\,dL_1(0,r)$ if $c_n\to\infty$ and, if $c_n=c>0$ for n large enough, to

$$\int_0^1 [|\lambda\delta' V_2(r)+z|1\{-\lambda\delta' V_2(r)-z=0\}$$

$$+|\lambda\delta' V_2(r)-z|1\{-\lambda\delta' V_2(r)+z<0\}]\,dL_1(0,r)$$

$$=\lambda\int_0^1\left[\left|\delta' V_2(r)+\frac{z}{\lambda}\right|1\left\{-\delta' V_2(r)>\frac{z}{\lambda}\right\}\right.$$

$$\left.+\left|\delta' V_2(r)-\frac{z}{\lambda}\right|1\left\{-\lambda\delta' V_2(r)<-\frac{z}{\lambda}\right\}\right]dL_1(0,r).$$

(Recall that $\{L_1(v,r)\}_{v\in\mathbb{R},\,r\in[0,1]}$ is continuous with respect to (v,r) and, for a fixed v, increases with r; see, for example, Revuz and Yor (1999). The notation $dL_1(w,r)$, with $w=0$ in the integrals in the preceding expression, indicates a Stieljes integral with respect to the increasing function $L_1(w,\cdot)$.)

Proof of Lemma 4. Define, for any $\epsilon>0$,

$$S_n=\frac{1}{c_n\sqrt{n}}\sum_{t=1}^{n}1\left\{0\le x_{1t}<-z-\frac{c_n\lambda\delta' x_{2t}}{\sqrt{n}}\right\}$$

$$=\sum_{i=-\infty}^{+\infty}\frac{1}{c_n\sqrt{n}}\sum_{t=1}^{n}1\left\{i\epsilon<\frac{\lambda\delta' x_{2t}}{\sqrt{n}}\le(i+1)\epsilon\right\}1\left\{0\le x_{1t}<-z-\frac{c_n\lambda\delta' x_{2t}}{\sqrt{n}}\right\}$$

$$=\sum_{i=-\infty}^{+\infty}S_{in},$$

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press

which corresponds to the first item of the sum of Lemma 4, the study of the second being similar.

We first study each of the S_{in}'s. Note that

$$S_{in} \leq \frac{1}{c_n\sqrt{n}} \sum_{t=1}^{n} 1\left\{ i\epsilon < \frac{\lambda\delta' x_{2t}}{\sqrt{n}} \leq (i+1)\epsilon \right\} 1\{0 \leq x_{1t} < -z - c_n i\epsilon\} = \bar{S}_{in},$$

$$S_{in} \geq \frac{1}{c_n\sqrt{n}} \sum_{t=1}^{n} 1\left\{ i\epsilon < \frac{\lambda\delta' x_{2t}}{\sqrt{n}} \leq (i+1)\epsilon \right\} 1\{0 \leq x_{1t} < -z - c_n (i+1)\epsilon\} = \underline{S}_{in}.$$

We study $\bar{S}_{in}$. The arguments of Theorem 2 in Akonom (1993) carry over and give

$$\begin{aligned}
\bar{S}_{in} &= \frac{1}{c_n\sqrt{n}} \int_0^n 1\left\{ i\epsilon < \frac{\lambda\delta' V_2(t)}{\sqrt{n}} \leq (i+1)\epsilon \right\} 1\{0 \leq V_1(t) < -z - c_n i\epsilon\}\, dt + o_p(1) \\
&\stackrel{d}{=} \frac{\sqrt{n}}{c_n} \int_0^1 1\{i\epsilon < \lambda\delta' V_2(r) \leq (i+1)\epsilon\} 1\{0 \leq \sqrt{n} V_1(r) < -z - c_n i\epsilon\}\, dr + o_p(1) \\
&= \frac{\sqrt{n}}{c_n} \int_{-\infty}^{+\infty} \int_0^1 1\{i\epsilon < \lambda\delta' V_2(r) \leq (i+1)\epsilon\} \\
&\quad \times 1\{0 \leq \sqrt{n} v < -z - c_n i\epsilon\}\, dL_1(v, r)\, dv + o_p(1).
\end{aligned}$$

Now the change of variables $w = \sqrt{n} v$ and the continuity of the local time yield

$$\begin{aligned}
\bar{S}_{in} &\stackrel{d}{=} \frac{1}{c_n} \int_{-\infty}^{+\infty} \int_0^1 1\{i\epsilon < \lambda\delta' V_2(r) \leq (i+1)\epsilon\} \\
&\quad \times 1\{0 \leq w < -z - c_n i\epsilon\}\, dL_1\left(\frac{w}{\sqrt{n}}, r \right) dw + o_p(1) \\
&= \frac{1}{c_n} \int_{-\infty}^{+\infty} \int_0^1 1\{i\epsilon < \lambda\delta' V_2(r) \leq (i+1)\epsilon\} \\
&\quad \times 1\{0 \leq w < -z - c_n i\epsilon\}\, dL_1(0, r)\, dw + o_p(1) \\
&= \int_0^1 1\{i\epsilon < \lambda\delta' V_2(r) \leq (i+1)\epsilon\} \frac{-z - c_n i\epsilon}{c_n} 1\{0 < -z - c_n i\epsilon\}\, dL_1(0, r) + o_p(1).
\end{aligned}$$

Consider now the case $c_n \to +\infty$. A similar study for $\underline{S}_{in}$ yields, jointly for each finite subset of index i,

$$\begin{pmatrix} \bar{S}_{in} \\ \underline{S}_{in} \end{pmatrix} \Rightarrow \begin{pmatrix} \displaystyle\int_0^1 1\{i\epsilon < \lambda\delta' V_2(r) \leq (i+1)\epsilon\}(-i\epsilon) 1\{i < 0\}\, dL_1(0, r) \\ \displaystyle\int_0^1 1\{i\epsilon < \lambda\delta' V_2(r) \leq (i+1)\epsilon\}(-(i+1)\epsilon) 1\{i < -1\}\, dL_1(0, r) \end{pmatrix}.$$

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press

Therefore, for any integer number $I > 0$,

$$\sum_{i=-I}^{I} \bar{S}_{ni} \Rightarrow \int_0^1 \left(\sum_{i=-I}^{I} 1\{i\epsilon < \lambda\delta' V_2(r) \leq (i+1)\epsilon\}(-i\epsilon) 1\{i < 0\} \right) dL_1(0,1)$$

$$\overset{I\to+\infty,\epsilon\to 0}{\to} \int_0^1 |\lambda\delta' V_2(r)| 1\{\lambda\delta' V_2(r) < 0\}\, dL_1(0,r),$$

$$\sum_{i=-I}^{I} \underline{S}_{ni} \Rightarrow \int_0^1 \left(\sum_{i=-I}^{I} 1\{i\epsilon < \lambda\delta' V_2(r) \leq (i+1)\epsilon\}(-(i+1)\epsilon) 1\{i < -1\} \right) dL_1(0,1)$$

$$\overset{I\to+\infty,\epsilon\to 0}{\to} \int_0^1 |\lambda\delta' V_2(r)| 1\{\lambda\delta' V_2(r) < 0\}\, dL_1(0,r).$$

Observe now that $\max_{1\leq t\leq n} |x_{2t}|/\sqrt{n} = O_p(1)$, so that, for each $\epsilon > 0$, the sum $S_n = \sum_{i=-\infty}^{+\infty} S_{in}$ reduces to $\sum_{i=-I_\epsilon}^{I_\epsilon} S_{in}$ with a high probability for I_ϵ taken large enough. Taking ϵ small enough and the preceding limit distribution therefore give the result of the lemma when $c_n \to \infty$. The case $c_n = c$ can be dealt with similarly. ■

LEMMA 5. *Assume that Assumption X holds. Let $\lambda > 0$ and z and M be some fixed nonnegative real number. For $\eta > 0$, let $Y_\eta = \inf_{\delta\in\mathbb{S}_{p-1},\lambda\geq 1} Z(\lambda,\delta) - \eta M L_1(0,1)$ where $Z(\lambda,\delta)$ is either equal to*

$$\int_0^1 |\delta' V_2(r)| dL_1(0,r) \quad \textit{or}$$

$$\int_0^1 \left[\left| \delta' V_2(r) + \frac{z}{\lambda} \right| 1\left\{ -\delta' V_2(r) > \frac{z}{\lambda} \right\} + \left| \delta' V_2(r) - \frac{z}{\lambda} \right| 1\left\{ -\lambda\delta' V_2(r) < -\frac{z}{\lambda} \right\} \right] dL_1(0,r).$$

Then, for any $\epsilon > 0$, one can find an $\eta = \eta_\epsilon > 0$ small enough and a $C_\epsilon > 0$, independent of λ, such that $P(Y_\eta \geq C_\epsilon) \geq 1 - \epsilon$, for any $\lambda \geq 1$.

Proof of Lemma 5. Because $\eta > 0$ can be taken small enough, it is sufficient to show that

$$P(\Omega) = 0, \tag{A.5}$$

where $\Omega = \{\omega : \inf_{\delta\in\mathbb{S}_{p-1},\lambda\geq 1} Z(\lambda,\delta) = 0\}$. We shall show that $P(\Omega) > 0$ is impossible because it would contradict the law of large numbers. Recall first that $L_1(0,\cdot)$ is a positive measure carried by the infinite set $\{r \in [0,1]; V_1(r) = 0\}$ (see Revuz and Yor, 1999). Moreover, integrals with respect to $L_1(0,\cdot)$ are Stieljes integral almost surely. Recall also that $\{V_1(r), V_2(r)\}_{r\in[0,1]}$ is a full-rank process under Assumption X.

Note that $(\lambda,\delta) \mapsto Z(\lambda,\delta)$ is a continuous function that has a limit when $\lambda \to \infty$. As a consequence, there is a $\delta^* = \delta^*(\omega) \in \mathbb{S}_{p-1}$, a $\lambda^* = \lambda^*(\omega) \in [1,\infty]$ such that $\inf_{\delta\in\mathbb{S}_{p-1},\lambda\geq 1} Z(\lambda,\delta) = Z(\lambda^*,\delta^*)$. Consider now an $\omega \in \Omega$ that is such that $Z(\lambda^*,\delta^*) = 0$. Then, by the definition of $Z(\cdot,\cdot)$, $|\delta' V_2(r)| = 0$ or $|\delta' V_2(r) + (z/\lambda)| 1\{-\delta' V_2(r) >$

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press

$(z/\lambda)\} + |\delta' V_2(r) - (z/\lambda)| 1\{-\lambda\delta' V_2(r) < -(z/\lambda)\} = 0$, $L_1(0,\cdot)$ almost everywhere. In any case, one can find an infinite random increasing sequence in $\{r \in [0,1]; V_1(r) = 0\}$ $\rho_k = \rho_k(\omega)$ with $\delta^* V_2(\rho_k)$ being a constant. This gives for all $K \geq 1$

$$\frac{1}{K}\sum_{k=1}^{K}(\rho_k - \rho_{k-1})^{-1}\delta^{*\prime}(V_2(\rho_k) - V_2(\rho_{k-1}))$$

$$\times (V_2(\rho_k) - V_2(\rho_{k-1}))'\delta^* = 0$$

$$= \delta^{*\prime}\left[\lim_{K\to+\infty}\frac{1}{K}\sum_{k=1}^{K}(\rho_k - \rho_{k-1})^{-1}(V_2(\rho_k)\right.$$

$$\left. - V_2(\rho_{k-1}))(V_2(\rho_k) - V_2(\rho_{k-1}))'\right]\delta^*,$$

for any $\omega \in \Omega$. Observe now that the $(\rho_k - \rho_{k-1})^{-1/2}(V_2(\rho_k) - V_2(\rho_{k-1}))$ are i.i.d. $\mathcal{N}(0,\Sigma_2)$, where Σ_2 is the covariance matrix of V_2 and is therefore of full rank. The strong law of large numbers therefore gives that $\delta^{*\prime}(\omega)\Sigma_2\delta^*(\omega) = 0$ with $\delta^*(\omega) \in S_{p-1}$ for almost all $\omega \in \Omega$, which is impossible unless (A.5) is true. ■

7.2. Proof of Main Results.

Proof of Proposition 1. Define $R_n(\beta) = R_n(\alpha) = Q_n(\beta;h_n) - \tilde{Q}_n(\beta;h_n)$ and observe that because $|K(\cdot)| \leq 1$, $\sup_{\beta\in\mathbb{S}_p}|R_n(\beta)| \leq (1/n)\sum_{t=1}^{n}|sgn(x_{1t} + u_t) - sgn(x_{1t})|$, with $x_{1t} = \beta_0' x_t$. Note that

$$|sgn(a) - sgn(b)| \leq 2 \times 1\{|a| \leq |b-a|\}, \quad \text{so that } \sup_{\beta\in\mathbb{S}_p}|R_n(\beta)| \leq \frac{2}{n}\sum_{t=1}^{n}1\{|x_{1t}| \leq u_t\}. \tag{A.6}$$

Consider first the case where Assumption U1 holds. Then, for $M > 0$ large enough, $\tilde{r}_n = Mr_n$, we have with a probability that can be made as close as wanted to 1 $\sqrt{n}\sup_{\beta\in\mathbb{S}_p}|R_n(\beta)| \leq (2/\sqrt{n})\sum_{t=1}^{n}1\{|x_{1t}| \leq \tilde{r}_n\} = O_p(r_n)$, by Lemma 3, taking $f(x) = 1\{|x| \leq 1\}$, $b_n = c_n = r_n$, $\zeta = 0$, and $a = 0$.

Consider now the case where Assumption U2 holds. Define

$$a_t = E[|sgn(x_{1t} + u_t) - sgn(x_{1t})|x_t, \ldots, x_1], \qquad b_t = |sgn(x_{1t} + u_t) - sgn(x_{1t})| - a_t,$$

which is such that $a_t \leq 2E[1\{|x_{1t}| \leq |u_t|\}|x_t] = 2P(|x_{1t}| \leq |u_t||x_{1t}) = 2(1 - F(|x_{1t}|;x_{1t}))$, so that $a_t = a(x_{1t})$ with $\int|a(x)|dx < \infty$ under Assumption U2. We have by (A.6) $\sqrt{n}\sup_{\beta\in\mathbb{S}_p}|R_n(\beta)| \leq S_{1n} + S_{2n}$, where $S_{1n} = (1/\sqrt{n})\sum_{t=1}^{n}a_t$ and $S_{2n} = (1/\sqrt{n})\sum_{t=1}^{n}b_t$. Notice that $S_{1n} = O_p(1)$ by Lemma 3 and S_{2n} is a martingale with quadratic variation

$$\frac{4}{n}\sum_{t=1}^{n}V[1\{|x_{1t}| \leq |u_t|\}|x_{1t}] \leq \frac{4}{n}\sum_{t=1}^{n}E[1\{|x_{1t}| \leq |u_t|\}|x_{1t}] = n^{-1/2}S_{1n} = O_p\left(\frac{1}{\sqrt{n}}\right).$$

Applying the same Dambis, Dubins–Schwartz martingale construction as in Park and Phillips (2000, Lem. 3), yields that $S_{2n} = O_p(n^{-1/4})$. ■

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press

Proof of Lemma 1. Let us begin with the first assertion of the lemma. Notice that $sgn(K(\beta' x_t/h_n)) = sgn(\beta' x_t)$ because the median of the cumulative distribution function $(K(\cdot)+1)/2$ is 0 under Assumption K. Observe also that, recalling $x_{1t} = \beta_0' x_t$, $1\{sgn(x_{1t}) = sgn(\beta' x_t)\} + 1\{sgn(x_{1t}) \neq sgn(\beta' x_t)\} = 1$. This gives

$$\begin{aligned}
\tilde{Q}_n(\beta; h_n) &= \frac{1}{n}\sum_{t=1}^{n} sgn(x_{1t})sgn(\beta' x_t)\left|K\left(\frac{\beta' x_t}{h_n}\right)\right| \\
&= \frac{1}{n}\sum_{t=1}^{n}\left[1\{sgn(x_{1t}) = sgn(\beta' x_t)\} - 1\{sgn(x_{1t}) \neq sgn(\beta' x_t)\}\right]\left|K\left(\frac{\beta' x_t}{h_n}\right)\right| \\
&= \frac{1}{n}\sum_{t=1}^{n}\left|K\left(\frac{\beta' x_t}{h_n}\right)\right| - \frac{2}{n}\sum_{t=1}^{n} 1\{sgn(x_{1t}) = sgn(\beta' x_t)\}\left|K\left(\frac{\beta' x_t}{h_n}\right)\right| \\
&\leq 1 - \frac{k(z)}{n}\sum_{t=1}^{n} 1\{z \leq |\beta' x_t|\}1\{sgn(x_{1t}) = sgn(\beta' x_t)\},
\end{aligned}$$

because $|K(\cdot)| \leq 1$ and by definition of $k(z)$. For the second assertion of Lemma 1, note that

$$\frac{\sqrt{n}}{h_n}(\tilde{Q}_n(\beta_0; h_n) - 1) = \frac{1}{h_n\sqrt{n}}\sum_{t=1}^{n}\left(\left|K\left(\frac{x_{1t}}{h_n}\right)\right| - 1\right) \Rightarrow L_1(0,1)\int(|K(x)| - 1)\,dx,$$

under Assumption K and by Lemma 3 (taking $f(\cdot) = |K(\cdot)| - 1$, $b_n = c_n = h_n$, $a = 0$, and $\zeta = 0$). ■

Proof of Proposition 2. Define $\tilde{R}_n(\lambda) = \sqrt{n}\gamma_n \inf_{\alpha \in S_p; \|\alpha_2\| \geq \gamma_n n^{-1/2}\lambda} \tilde{Q}_n^c(\alpha)$ and observe that the proposition is proved if we can find $C_\epsilon > 0$ with

$$P(\tilde{R}_n(\lambda) \geq C_\epsilon \lambda) \geq 1 - \epsilon. \tag{A.7}$$

Observe that, if $\delta = \alpha_2/\|\alpha_2\|$, we have for any $\alpha_2 = \|\alpha_2\|\delta$ with $\|\alpha_2\| \geq \gamma_n n^{-1/2}\lambda > 0$,

$$\begin{aligned}
&\sqrt{n}\gamma_n \tilde{Q}_n^c(\alpha) \\
&= \frac{1}{\gamma_n\sqrt{n}}\sum_{t=1}^{n} 1\{|\alpha_1 x_{1t} + \alpha_2' x_{2t}| \geq z\} \\
&\quad \times [1\{x_{1t} \geq 0, \alpha_1 x_{1t} + \alpha_2' x_{2t} < 0\} + 1\{x_{1t} < 0, \alpha_1 x_{1t} + \alpha_2' x_{2t} \geq 0\}] \\
&= \frac{1}{\gamma_n\sqrt{n}}\sum_{t=1}^{n}[1\{x_{1t} \geq 0, \alpha_1 x_{1t} + \alpha_2' x_{2t} < -z\} + 1\{x_{1t} < 0, \alpha_1 x_{1t} + \alpha_2' x_{2t} \geq z\}]
\end{aligned}$$

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press

$$= \frac{1}{\gamma_n\sqrt{n}}\sum_{t=1}^{n}\left[1\left\{x_{1t}\geq 0, \frac{\alpha_2' x_{2t}}{\|\alpha_2\|\sqrt{n}} < \frac{-z-\alpha_1 x_{1t}}{\|\alpha_2\|\sqrt{n}}\right\}\right.$$

$$\left. + 1\left\{x_{1t}<0, \frac{\alpha_2' x_{2t}}{\|\alpha_2\|\sqrt{n}} \geq \frac{+z-\alpha_1 x_{1t}}{\|\alpha_2\|\sqrt{n}}\right\}\right]$$

$$\geq \frac{1}{\gamma_n\sqrt{n}}\sum_{t=1}^{n}\left[1\left\{x_{1t}\geq 0, \frac{\delta' x_{2t}}{\sqrt{n}} < \frac{-z-x_{1t}}{\gamma_n\lambda}\right\} + 1\left\{x_{1t}<0, \frac{\delta' x_{2t}}{\sqrt{n}} \geq \frac{z-x_{1t}}{\gamma_n\lambda}\right\}\right]$$

$$= \frac{1}{\gamma_n\sqrt{n}}\sum_{t=1}^{n}\left[1\left\{0\leq x_{1t} < -z - \frac{\gamma_n\lambda\delta' x_{2t}}{\sqrt{n}}\right\} + 1\left\{z - \frac{\gamma_n\lambda\delta' x_{2t}}{\sqrt{n}} \leq x_{1t} < 0\right\}\right].$$

Define $\tilde{q}_n^c(\delta) = (1/\gamma_n\sqrt{n})\sum_{t=1}^{n}[1\{0 \leq x_{1t} < -z - (\gamma_n\lambda\delta' x_{2t}/\sqrt{n})\} + 1\{z - (\gamma_n\lambda\delta' x_{2t}/\sqrt{n}) \leq x_{1t} < 0\}]$, which is then such that $\tilde{R}_n(\lambda) \geq \inf_{\delta\in S_{p-1}}\tilde{q}_n^c(\delta)$. Consider now an η-covering of S_{p-1}, $B(\delta_j;\eta) = B(\delta_j)$, $j = 1,\ldots,J$ with balls centered at δ_j and radius η, where $\eta > 0$ will be chosen small enough subsequently. We have

$$\inf_{\delta\in S_{p-1}}\tilde{q}_n^c(\delta) \geq \min_{j=1,\ldots,J}\tilde{q}_n^c(\delta_j) - \max_{j=1,\ldots,J}\sup_{\delta\in B(\delta_j)}|\tilde{q}_n^c(\delta) - \tilde{q}_n^c(\delta_j)|. \tag{A.8}$$

Consider first the first item of (A.8) and let $Z(\lambda,\delta)$ be as in Lemma 5. Because the convergence of Lemma 4 holds jointly with respect to the δ_j's, we have for $\lambda \geq 1$,

$$\min_{j=1,\ldots,J}\tilde{q}_n^c(\delta_j) \Rightarrow \lambda \min_{j=1,\ldots,J} Z(\lambda,\delta_j). \tag{A.9}$$

Let us now study the second term of (A.8). The definition of $\tilde{q}_n^c(\cdot)$ shows that, for any $\delta \in B(\delta_j)$,

$$|\tilde{q}_n^c(\delta) - \tilde{q}_n^c(\delta_j)| \leq \frac{1}{\gamma_n\sqrt{n}}\sum_{t=1}^{n}1\left\{\left|x_{1t} + \frac{\gamma_n\lambda\delta_j' x_{2t}}{\sqrt{n}}\right| \leq \frac{\lambda\gamma_n|(\delta-\delta_j)' x_{2t}|}{\sqrt{n}}\right\}$$

$$\leq \frac{1}{\gamma_n\sqrt{n}}\sum_{t=1}^{n}1\left\{\left|x_{1t} + \frac{\lambda\gamma_n\delta_j' x_{2t}}{\sqrt{n}}\right| \leq \gamma_n\lambda\eta\max_{1\leq t\leq n}\frac{\|x_{2t}\|}{\sqrt{n}}\right\}.$$

Choose M such that $P(\max_{1\leq t\leq n}(\|x_{2t}\|/\sqrt{n}) \geq M)$ is small for n large enough. Then applying Lemma 3 with $f(x) = 1(|x| \leq \lambda M)$, $b_n = c_n = \gamma_n\zeta = 0$, and $a = 0$ yields

$$\max_{j=1,\ldots,J}\frac{1}{\gamma_n\sqrt{n}}\sum_{t=1}^{n}1\left\{\left|x_{1t} + \frac{\lambda\gamma_n\delta_j' x_{2t}}{\sqrt{n}}\right| \leq \gamma_n\lambda\eta M\right\} \Rightarrow 2M\lambda\eta L_1(0,1). \tag{A.10}$$

Expressions (A.9) and (A.10) give (A.7) by Lemma 5, choosing η small enough. ■

Proof of Theorem 2. We limit ourselves to the proof of the second assertion of the theorem assuming $h_n = 1$, the other cases being exactly similar. Fix $z \geq 0$ in Lemma 1 with $k(z) > 0$. Let $R_n(\alpha) = Q_n(\alpha;h_n) - \tilde{Q}_n(\alpha;h_n)$. Consider an arbitrary $\epsilon > 0$ and the associated $C_{\epsilon/2}$ of Proposition 2. Then Proposition 1 and Lemma 1 show that one can take a $\lambda_\epsilon > 0$ large enough such that

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press

$$\limsup_{n\to\infty} P\left(\sup_{\|\alpha\|=1} |R_n(\alpha)| \geq \frac{1}{2\sqrt{n}} C_{\epsilon/2}\lambda_\epsilon, |\tilde{Q}_n(\alpha_0;h_n)-1| \geq \frac{1}{2\sqrt{n}} C_{\epsilon/2}\lambda_\epsilon \right) \leq \frac{\epsilon}{2}.$$

Therefore, if

$$E_n = \left\{ \sup_{\alpha\in\mathbb{S}_p} |R_n(\alpha)| < \frac{C_{\epsilon/2}\lambda_\epsilon}{2\sqrt{n}} |\tilde{Q}_n(\alpha_0;h_n)-1| \right.$$

$$\left. < \frac{1}{2\sqrt{n}} C_{\epsilon/2}\lambda_\epsilon \sup_{\alpha\in\mathbb{S}_p;\|\alpha_2\|\geq n^{-1/2}\lambda_\epsilon} \tilde{Q}_n(\alpha;h_n) < 1 - \frac{C_{\epsilon/2}\lambda_\epsilon}{\sqrt{n}} \right\},$$

we have, by Lemma 1 and Propositions 1 and 2, $\liminf_{n\to\infty} P(E_n) \geq 1-\epsilon$.

The definition of E_n shows that, on E_n and because $Q_n(\alpha_0;h_n) \leq \sup_{\alpha\in S_p;\|\alpha_2\|\leq n^{-1/2}\lambda_\epsilon} Q_n(\alpha;h_n)$,

$$\sup_{\alpha\in\mathbb{S}_p;\|\alpha_2\|\leq n^{-1/2}\lambda_\epsilon} Q_n(\alpha;h_n) \geq \tilde{Q}_n(\alpha_0;h_n) + R_n(\alpha_0)$$

$$> 1 - \frac{C_{\epsilon/2}\lambda_\epsilon}{2\sqrt{n}} - \frac{C_{\epsilon/2}\lambda_\epsilon}{2\sqrt{n}} = 1 - \frac{C_{\epsilon/2}\lambda_\epsilon}{2\sqrt{n}},$$

$$\sup_{\alpha\in\mathbb{S}_p;\|\alpha_2\|\geq n^{-1/2}\lambda_\epsilon} Q_n(\alpha;h_n) \leq \sup_{\alpha\in\mathbb{S}_p;\|\alpha_2\|\geq n^{-1/2}\lambda_\epsilon} \tilde{Q}_n(\alpha:h_n) + \sup_{\alpha\in\mathbb{S}_p} |R_n(\alpha)|$$

$$< 1 - \frac{C_{\epsilon/2}\lambda_\epsilon}{\sqrt{n}} + \frac{C_{\epsilon/2}\lambda_\epsilon}{2\sqrt{n}} = 1 - \frac{C_{\epsilon/2}\lambda_\epsilon}{2\sqrt{n}},$$

so that

$$E_n \subset \left\{ \sup_{\alpha\in\mathbb{S}_p;\|\alpha_2\|\geq n^{-1/2}\lambda_\epsilon} Q_n(\alpha;h_n) < \sup_{\alpha\in\mathbb{S}_p;\|\alpha_2\|\leq n^{-1/2}\lambda_\epsilon} Q_n(\alpha;h_n) \right\} \subset \{\|\hat{\alpha}_{2n}\| \leq n^{-1/2}\lambda_\epsilon\},$$

by the definition (3.1) of the SMSE. Therefore $\sqrt{n}\hat{\alpha}_{2n} = O_p(1)$, because $\liminf_{n\to\infty} P(E_n)$ can be made arbitrarily large. To obtain the second assertion of the theorem, note that, because $\hat{\alpha}_n \in \mathbb{S}_p$, $|\hat{\alpha}_{1n}| - 1 = \sqrt{1-\|\hat{\alpha}_{2n}\|^2} - 1 = O_p(1/n)$. Moreover, Theorem 1 implies that $P(|\hat{\alpha}_{1n}| = \hat{\alpha}_{1n}) \to 1$ because $\hat{\alpha}_{1n}$ is a consistent estimator of 1. ■

https://doi.org/10.1017/S0266466606060336 Published online by Cambridge University Press