A Smoothed Maximum Score Estimator for the Binary Choice Panel Data Model with Individual Fixed Effects and Application to Labour Force Participation

by

Erwin Charlier[1]
Department of Econometrics and CentER
Tilburg University
P. O. Box 90153
5000 LE, Tilburg
The Netherlands

September 1994

Abstract

In a binary choice panel data model with individual effects and two time periods, Manski proposed the maximum score estimator, based on a discontinuous objective function, and proved its consistency under weak distributional assumptions. However, the rate of convergence of this estimator is low ($N^{1/3}$) and its limit distribution cannot be used for making inference. This paper overcomes this problem by applying the idea of Horowitz to smooth Manski's objective function. The paper extends the resulting smoothed maximum score estimator to the case of more than two time periods and to unbalanced panels (assuming away selectivity effects). Under weak assumptions the estimator is consistent and asymptotically normal with a rate of convergence that is at least $N^{2/5}$ and can be made arbitrarily close to $N^{1/2}$, depending on the strength of the smoothness assumptions imposed. Statistical inferences can be made. The estimator is applied to an equation for labour force participation of married Dutch females on the basis of annual observations from 1984 through 1988. A simulated annealing type of algorithm is used to maximize the objective function because it can have many local maxima and attention is paid to the choice of the smoothness parameter. Finally, some model specification tests are performed.

Keywords: panel data, binary choice model, semiparametric estimation, smoothing, selectivity bias, unbalanced panel.

[1] I thank Bertrand Melenberg and Arthur van Soest for many helpful comments and discussions and a referee for useful comments. All remaining errors are mine. Furthermore, I am grateful to the Netherlands Central Bureau of Statistics (CBS) for providing the data. The views expressed in this paper do not necessarily reflect the views of the CBS.

1. Introduction

In a binary choice panel data model with individual effects and two time periods, Manski (1987) proposed the maximum score estimator, based on a discontinuous objective function, and proved its consistency under weak distributional assumptions. However, the rate of convergence of this estimator is low ($N^{1/3}$) and its limit distribution cannot be used for making inference. This paper overcomes this problem by applying the idea of Horowitz (1992) to smooth Manski's objective function. Moreover, it generalizes Manski (1987) to panels with more than two time periods and to unbalanced panels.

This paper considers a binary choice panel data model with individual effects:

$$\begin{cases} y_{it}^{*} = \beta' x_{it} + \alpha_i + u_{it}, \quad i=1,..,N,\ t=1,..,T. \\ y_{it} = 1(y_{it}^{*} \geq 0) \end{cases} \tag{1.1}$$

in which $ß \in \mathbb{R}^k$ and 1(A) is the indicator function that is 1 if A is true and 0 otherwise. One observes $(y_{it}, x_{it}')'$, i=1,..,N for some (possibly all) $t \in \{1,2,..,T\}$. The index i represents the individuals or households and index t represents time. An example of such a model is a labour force participation model of married females. The dependent variable is whether a female participates or not and the explanatory variables include household characteristics and labour supply of the male.

In general, the model assumes independence across individuals and imposes rather strong assumptions with respect to the distributions of α_i and $u_i=(u_{i1},..,u_{iT})$, conditional on $x=(x_{i1},..,x_{iT})$. When, for example, α_i and u_i are assumed to be independently normally distributed and the u_{it} are i.i.d. over t, we have the Heckman and Willis (1976) model. A drawback of this approach is that the composite error terms $v_{it}=\alpha_i+u_{it}$ are equally correlated over time. A normal distribution with a general structure for the covariance of the u_{it} is assumed in Avery, Hansen and Hotz (1983). A drawback of both models is that the α_i are not allowed to depend on $(x_{i1},..,x_{iT})$. This problem has been overcome by Chamberlain (1984), who assumes normality of α_i and u_i, with unrestricted covariance matrix and allows the α_i to be correlated with $(x_{i1},..,x_{it})$. A GMM estimation procedure can be used to estimate ß.

In contrast, in a fixed effects model, the incidental parameters problem arises (Neyman and Scott (1948)). One feasible approach to deal with a fixed effects model is to assume the u_{it} to follow an i.i.d. standard logistic distribution and then use conditional maximum likelihood to estimate β. Assuming i.i.d. normal errors cannot be used to estimate β consistently, see Maddala (1987).

A drawback of all the random effects parametric models is the assumption of normal distributions for α_i and/or u_i. In general, this may yield inconsistent estimators of β if the true distributions of α_i and/or u_i are nonnormal. In a fixed effects setting the distributional assumptions are also rather restrictive. To solve the problem for a cross-section binary choice model (without the α_i), several estimators for β have been proposed that are consistent under weaker assumptions. Examples are the maximum score estimator of Manski (1985) and the smoothed maximum score estimator of Horowitz (1992). A drawback of the former is that the rate of convergence is low ($N^{1/3}$) and its limit distribution is some complicated non-normal distribution that is hard to use for inference (see Kim and Pollard (1990)). This problem has been overcome by the smoothed maximum score estimator, which is obtained by smoothing the maximum score objective function, such that the asymptotic behaviour can be analyzed using standard Taylor series approximations.

If one is willing to make strong distributional assumptions in a binary choice panel data model, one of the previous mentioned parametric approaches can be used to estimate β. However, consistency is lost if the distributional assumptions are not valid. The semiparametric literature is limited for the binary choice panel data model with individual effects. An example of such an estimator, for the case T=2, is the maximum score estimator proposed by Manski (1987). The resulting estimator for β is consistent under weak assumptions but the limit distribution shares the problems of the estimator of Manski (1985) for a cross-section. This paper aims to construct a consistent asymptotically normal estimator for β in model (1.1) with individual effects, based on relatively weak assumptions. The estimator will be derived by combining the ideas of Horowitz (1992) and Manski (1987) and the estimator will be extended to the case of more periods ($T\geq2$) and for unbalanced panels (without selectivity). The assumptions indicate that the estimator is consistent both in a fixed effects model and a random effects model, because the distribution of α_i conditional on $x=(x_{i1},..,x_{iT})$ is not restricted. Also, serial correlation between the error terms as well as forms of heteroskedasticity are allowed for. The resulting smoothed maximum score estimator is calculated for an empirical application concerning labour force participation of married Dutch females.

The remainder of this paper is organized as follows: section 2 defines the smoothed maximum score estimator for β in model (1.1) and derives its asymptotic properties. Section 3 discusses the empirical application. The results are obtained by using a global search algorithm to find the global optimum of some non-concave objective function, as proposed by Corana et al. (1987). Section 4 deals with specification testing. Concluding remarks are presented in section 5. The assumptions used to prove consistency are presented in the main text: they indicate when things may go wrong. The additional assumptions required for deriving the asymptotic limit distribution are presented in the appendix together with proofs of theorems and lemmas.

2. Smoothed Maximum Score for Panel Data

This section extends the smoothed maximum score estimation method as proposed for cross-section data by Horowitz (1992) to the case of panel data. The only assumptions concerning u_{it}, t=1,..,T, is that they are time stationary conditional on $(x_{i1},..,x_{It})$ and α_i and that the support of the distribution function of u_{it} is $\mathbb{R}$. As is common in binary choice models, one normalization has to be made for identification in model (1.1). Because no parametric distributional assumptions are made, this cannot be established by normalizing a parameter in the distribution function of the α_i or u_{it}. The normalization thus has to concern β. Following Horowitz (1992), the paper normalizes to one in absolute value an element in β that is nonzero and that is related to an absolute continuous element in $w_{its} \equiv x_{it} - x_{is}$. Arrange the components of $w_{its} = (w_{its,1},..,w_{its,k})$ such that $w_{its,1}$ satisfies this condition, then the normalization is $|b_1|=1$.

Define

$$G^{*}_{NT}(b) = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s<t} c_{its}\,\mathrm{sign}(b' w_{its})(y_{it} - y_{is}) \tag{2.1}$$

where $c_{its} = r_{it} r_{is}$, with $r_{it}=1$ if $\{y_{it},x_{it}\}$ is observed, and zero otherwise (a missing observation); hence $c_{its}=1$ if both $\{y_{it},x_{it}\}$ and $\{y_{is},x_{is}\}$ are observed and zero otherwise, and sign(z)=1 if $z \geq 0$ and -1 otherwise. From the definition of c_{its} it follows that individuals who are not observed or who are observed in only one time period do not contribute to the objective function and hence N can be interpreted as the number of individuals for whom at least two of the (y_{it},x_{it}), t=1,..,T are observed.[2] For T=2 and all $c_{its}=1$, maximization of $G^{*}_{NT}(b)$ w.r.t. b (and normalizing $\|b\|=1$) yields the maximum score estimator of Manski (1987). Let $Y=\{(y_{it},y_{is},x_{it},x_{is}) \mid y_{it} \neq y_{is}\}$. Maximizing $G^{*}_{NT}(b)$ boils down to choosing b such that the sign of $b'w_{its}$ equals the sign of $y_{it} - y_{is}$ for as many observations in Y as possible. Under the same distributional assumptions as mentioned in the beginning of this section, the resulting estimator is consistent.

The problems with the limit distribution of the estimator obtained by maximizing $G^{*}_{NT}(b)$ are caused by the sign function, which is a step function. The idea of Horowitz (1992) is to smooth the objective function. Note that maximizing $G^{*}_{NT}(b)$ boils down to maximizing

---

[2] Since this paper imposes independence between the c_{ts}'s and the other variables, there is no harm in defining N this way.

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s<t} c_{its}\frac{1}{2}\left(\mathrm{sign}(b'w_{its})+1\right)(y_{it}-y_{is}) \;=\; \frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s<t} c_{its}\,1(b'w_{its}\geq 0)\,(y_{it}-y_{is})\ . \qquad \textbf{(2.2)}$$

This objective function can be smoothed by replacing the indicator function 1(.) by some smooth function $K^N(.)$ that converges to the indicator function as $N\to\infty$. Rewriting $y_{it}-y_{is}$ as $1(y_{it}\neq y_{is})[2*1(y_{it}=1,y_{is}=0)-1]$, following Horowitz (1992), let

$$G_{NT}(b;\sigma_N) \;=\; \frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s<t} c_{its}\,1(y_{it}\neq y_{is})\left[2*1(y_{it}=1,y_{is}=0)-1\right]K\left(\frac{b'w_{its}}{\sigma_N}\right) \qquad \textbf{(2.3)}$$

where $\sigma_N\to 0$ $(N\to\infty)$ and K(.) is a continuous function of the real line into itself satisfying:

K1. $|K(v)|<M$ for some finite M and all v in $\mathbb{R}$;

K2. $\lim_{v\to-\infty} K(v)=0$ and $\lim_{v\to\infty} K(v)=1$.

K(v) could thus be a distribution function but it also might take on values larger than one or lower than zero and it need not necessarily be increasing. Two examples satisfying K1 and K2 are $K_2(v)=\Phi(v)$ and

$$K_4(v) \;=\; \begin{cases} 0 & \text{if } v < -5, \\ \frac{1}{2}+\frac{105}{64}\left[\frac{v}{5}-\frac{5}{3}(\frac{v}{5})^3+\frac{7}{5}(\frac{v}{5})^5-\frac{3}{7}(\frac{v}{5})^7\right] & \text{if } -5\leq v\leq 5, \\ 1 & \text{if } v > 5. \end{cases} \qquad \textbf{(2.4)}$$

cf. Horowitz (1992).

The derivative of $K_h(v)$ (h=2,4) with respect to v is an h^{th} order kernel. It is easily seen that if z equals zero with probability zero, then $K(z/\sigma_N)\to 1(z\geq 0)$ almost surely as $N\to\infty$ (and thus $\sigma_N\to 0$) and use this to prove that $G_{NT}(b;\sigma_N)\to G^*_{NT}(b)$ almost surely uniformly in b as N tends to infinity. Use this property, together with assumptions that are similar to Horowitz (1992), and some additional assumptions concerning exclusion of any form of selectivity bias (caused by attrition, initial nonresponse, wave nonresponse or item nonresponse, see Verbeek and Nijman (1992)), to prove consistency of the smoothed maximum score estimator in model (1.1). The continuity and differentiability of $G_{NT}(b;\sigma_N)$ makes it feasible to derive the asymptotic distribution through the usual Taylor series approximations.

Let $x=(x_1,..,x_T)$ and let F denote the population distribution of $\{(y^*_t,x_t,u_t;\ t=1,..,T),\alpha\}$. Let $F_{u|x,\alpha}$ denote the distribution of u conditional on (x,α) and let $F_{w_{ts}}$ denote the distribution of w_{ts} (i subscripts are suppressed). To prove consistency of the estimator resulting from maximization of

$G_{NT}(b;\sigma_N)$ over the set $|b_1|=1$ and $(b_2,..,b_k)$ in a compact set $\tilde{B}$, use the following **assumptions** (assumptions (i)−(iii) are the analogons of Manski (1987), assumption (iv) is from Horowitz (1992) and assumption (v) is extra):

(i) a) $F_{u_t|x,\alpha} = F_{u_s|x,\alpha}$ for all (x,α) and $s,t\leq T$;

b) The support of $F_{u_t|x,\alpha}$ is $\mathbb{R}$ for all (x,α) and all t;

(ii) a) For all t,s the support of $F_{w_{ts}}$ is not contained in any proper linear subspace of $\mathbb{R}^k$;

b) For all t,s there exists at least one j in {1,2,..,k} such that $ß_j\neq 0$ and such that, for almost every value of $\tilde{w}_{ts}=(w_{ts,1},..,w_{ts,j-1}, w_{ts,j+1},..,w_{ts,k})$ the scalar random variable $w_{ts,j}$ has everywhere positive Lebesgue density conditional on $\tilde{w}_{ts}$ and $y_t\neq y_s$. Notice that j=1 has already been used;

(iii) A random sample is drawn from F;

(iv) $|ß_1|=1$ and $\tilde{\beta}=(ß_2,..,ß_k)'$ is contained in a compact subset $\tilde{B}$ of $\mathbb{R}^{k-1}$;

(v) c_{ts} is independent of $(y_1,x_1,..,y_T,x_T)$ and $P(c_{ts}>0)>0$ for some t,s.

Assumption (i) a) says that the distribution of the error term in (1.1) is time stationary conditional on (x,α). Assumptions (i) b) and (ii) a) are regularity conditions needed for identification. For assumption (ii) b) to hold, w_{ts} should contain an absolute continuous element with non-zero coefficient. Assumptions (iii) and (iv) need no explanation. Assumption (v) allows for an unbalanced or rotating panel but requires the absence of selectivity bias. (v) implies that N, the number of observations for which at least two time periods are available, tends to infinity if the random sample grows in size. To prove consistency, it is sufficient that c_{ts} is independent of (y_t,x_t,y_s,x_s), but the slightly stronger assumption (v) that c_{ts} is independent of $(y_1,x_1,..,y_T,x_T)$ is needed to derive the limit distribution. The assumptions place no restrictions on the distribution of α conditional on x, and assumption (i) implies that no restrictions are imposed on the serial dependence between u_t and u_s $(s\neq t)$, while the form of heteroskedasticity is restricted only through (i) b). It includes heterogeneity of the form $Var(u_t|\alpha,x)=\exp(\alpha+\tau'x)$, t=1,..,T, whereas it excludes $Var(u_t|\alpha,x)=\exp(\alpha+\tau'x_t)$, t=1,..,T, so the dependence must be through x and not just through x_t.

The following corollary indicates that the present panel data problem has a median regression interpretation (cf. Manski (1987, p. 360)), which is the basis for the construction of the estimator.

Corollary 1:

Let assumption (i) hold. Then for all t,s $Median(y_t-y_s|w_{ts},y_t\neq y_s)=sign(ß'w_{ts})$ (i subscripts are suppressed). ■

This conditional median restriction can be viewed as an alternative way to write the model (conditional on $y_t \neq y_s$) as:

$$\begin{cases} z^*_{its} = \beta' w_{its} + u_{its} & \text{and} \\ y_{it} - y_{is} = \text{sign}(z^*_{its}) & \end{cases} \tag{2.5}$$

and Median($u_{its} \mid w_{its}, y_{it} \neq y_{is}$)=0, for all i, t and s.

The following theorem shows that the smoothed maximum score estimator for panel data is consistent under assumptions (i)-(v).

Theorem 1 (Consistency):
Let assumptions (i)−(v) hold. Define $\tilde{b}=(b_2,..,b_k)'$ and $\tilde{w}_{ts}=(w_{ts,2},..,w_{ts,k})'$. Let b_N be a solution to

$$\underset{b:|b_1|=1,\ \tilde{b}\in\tilde{\mathbf{B}}}{\text{maximize}}\ G_{NT}(b;\sigma_N) \tag{2.6}$$

Then $\lim_{N\to\infty} b_N$ = ß almost surely. ■

Before stating the theorem that deals with the asymptotic distribution of the smoothed maximum score estimator, this section will provide some definitions. Let z_{ts}=ß'w_{ts}. Then, because of the normalization in β, there is a one-to-one relation between $(z,\tilde{w}_{ts})$ and w_{ts} for each fixed β. By assumption (ii), the distribution of z_{ts} conditional on $\tilde{w}_{ts}$ and $y_t \neq y_s$ has everywhere positive density with respect to Lebesgue measure for almost every $\tilde{w}_{ts}$. Let $p(z_{ts} \mid \tilde{w}_{ts}, y_t \neq y_s)$ denote this density. For each positive integer i define
$p^{(i)}(z_{ts} \mid \tilde{w}_{ts}, y_t \neq y_s) = \partial^i p(z_{ts} \mid \tilde{w}_{ts}, y_t \neq y_s)/\partial z^i_{ts}$ whenever the derivative exists and let
$p^{(0)}(z_{ts} \mid \tilde{w}_{ts}, y_t \neq y_s) = p(z_{ts} \mid \tilde{w}_{ts}, y_t \neq y_s)$.
Let $P(\tilde{w}_{ts} \mid y_t \neq y_s)$ denote the cumulative distribution function of $\tilde{w}_{ts}$ conditional on $y_t \neq y_s$,
and let $F_u(-z_{ts} \mid z_{ts}, \tilde{w}_{ts}, y_t \neq y_s)$ denote the cumulative distribution of $u=u_{ts}$ conditional on z_{ts}, $\tilde{w}_{ts}$ and $y_t \neq y_s$, evaluated at $-z_{ts}$ and where u (=u_{ts}) is the error term in model **(2.5)**.
For each positive integer i, define $F^{(i)}_u(-z_{ts} \mid z_{ts}, \tilde{w}_{ts}, y_t \neq y_s) = \partial^i F_u(-z_{ts} \mid z_{ts}, \tilde{w}_{ts}, y_t \neq y_s)/\partial z^i_{ts}$ whenever the derivative exists.
Let

$$A = \sum_{t=2}^{T}\sum_{s<t} -2\int \xi^h K'(\xi)\,d\xi \sum_{i=1}^{h} \frac{1}{i!(h-i)!} \tag{2.7}$$
$$E\left\{F_u^{(i)}(0|0,\tilde{w}_{ts},y_t\neq y_s)p^{(h-i)}(0|\tilde{w}_{ts},y_t\neq y_s)\tilde{w}_{ts}|y_t\neq y_s\right\}P(c_{ts}=1)P(y_t\neq y_s)$$

$$D_1 = \sum_{t=2}^{T}\sum_{s<t}\int\left[K'(\xi_{ts})\right]^2 d\xi_{ts} E\left\{\tilde{w}_{ts}\tilde{w}_{ts}' p(0|\tilde{w}_{ts},y_t\neq y_s)\middle|y_t\neq y_s\right\}P(c_{ts}=1)P(y_t\neq y_s) \tag{2.8}$$

and

$$Q = 2\sum_{t=2}^{T}\sum_{s<t} E\left\{\tilde{w}_{ts}\tilde{w}_{ts}' F_u^{(1)}(0|0,\tilde{w}_{ts},y_t\neq y_s)p(0|\tilde{w}_{ts},y_t\neq y_s)\,\middle|\, y_t\neq y_s\right\}P(c_{ts}=1)P(y_t\neq y_s)\ . \tag{2.9}$$

In addition, let assumptions (vi) to (xi) (see appendix) hold for some $h\geq 2$. This requires the use of a smoothing function $K(v)$ such that the derivative of $K(v)$ is an h^{th} order Kernel ($h\geq 2$, examples are $K_2(v)$ and $K_4(v)$ introduced above) and some additional assumptions on the density of (suppressing the i subscript) $\beta' w_{ts}$ and the distribution of u_{ts}, both conditional on ($\tilde{w}_{ts}$, $y_t\neq y_s$), (see appendix assumptions (vii), (viii) and (ix)). The following theorem shows the main result concerning the asymptotic distribution of the smoothed maximum score estimator.

Theorem 2 (Asymptotic Distribution):

Let assumptions (i)−(xi) hold for some $h\geq 2$ (assumptions (vi)−(xi) are in the appendix) and let $\{b_N\}$ be a sequence of solutions to the maximization of problem **(2.6)**. The fastest rate of convergence in distribution is obtained by the following: Let $\sigma_N=(\lambda/N)^{1/(2h+1)}$ with $0<\lambda<\infty$; let Ω be any nonstochastic, positive semidefinite matrix such that $A'Q^{-1}\Omega Q^{-1}A\neq 0$; let E_A denote the expectation with respect to the asymptotic distribution of $N^{h/(2h+1)}(\tilde{b}_N-\tilde{\beta})$, and $MSE=E_A(\tilde{b}_N-\tilde{\beta})'\Omega(b_N-\tilde{\beta})$. MSE is minimized by setting

$$\lambda=\lambda^*=[\mathrm{trace}(Q^{-1}\Omega Q^{-1}D_1)]/(2hA'Q^{-1}\Omega Q^{-1}A),$$

in which case

$$N^{h/(2h+1)}(\tilde{b}_N-\tilde{\beta}) \to^d MVN(-(\lambda^*)^{h/(2h+1)}Q^{-1}A,\ (\lambda^*)^{-1/(2h+1)}Q^{-1}D_1Q^{-1}). \qquad \blacksquare$$

Note that the rate of convergence is lower than $N^{1/2}$ and depends on h. By choosing h large enough, the rate of convergence can be made arbitrarily close to $N^{1/2}$. As before, a larger h requires the use of a higher-order kernel and stronger requirements with respect to $p(z_{ts}|\tilde{w}_{ts},y_t\neq y_s)$ and $F_u(-z_{ts}|z_{ts},\tilde{w}_{ts},y_t\neq y_s)$, see assumptions (vii), (viii) and (ix) in the appendix. For h=1 the rate of convergence is $N^{1/3}$ and $N^{1/3}(\tilde{b}_N-\tilde{\beta})$ has an unknown limit distribution, and is therefore not useful for making inferences (see Horowitz (1992), p. 514); hence for h=1 the smoothed maximum score estimator for panel data has no apparent advantages over Manski's estimator. For $h\geq 2$, the estimator has an asymptotic bias. The structure of the asymptotic covariance matrix is similar to that of an extremum estimator or to that of a pseudo maximum likelihood estimator. The theorem stated here follows from theorems 1 and 2 in the appendix. The interested reader can find detailed information concerning lower rates of convergence (theorem 2) there.

Finally, if theorem 2 is to be used to make inferences, consistent estimators for the matrices involved in the asymptotic distribution of the smoothed maximum score estimator have to be constructed. The following theorem shows how to construct consistent estimators for A, D_1 and Q, where the expressions for $T_{NT}(b_N,\sigma_N)$ and $Q_{NT}(b_N,\sigma_N)$ are the (familiar) first-order derivatives and the second-order derivatives of the objective function $G_{NT}(b,\sigma_N)$ with respect to $\tilde{b}$, respectively.

Theorem 3:

Let b_N be a consistent smoothed maximum score estimator based on $\sigma_N=O(N^{-1/(2h+1)})$. For $b\in\{-1,1\}x\tilde{B}$ and i=1,..,N, define

$$a_{its}(b,\sigma) \ = \ c_{its}1(y_{it}\neq y_{is})\left[2*1(y_{it}=1,y_{is}=0)-1\right]K'\left(\frac{b'w_{its}}{\sigma}\right)\frac{\tilde{w}_{its}}{\sigma} \qquad \textbf{(2.10)}$$

Let $\sigma_N^*=O(N^{-\delta/(2h+1)})$, where $0<\delta<1$. Then

(a) $\hat{A}_N=(\sigma_N^*)^{-h}T_{NT}(b_N;\sigma_N^*)$ converges in probability to A;

(b) the matrix

$$\hat{D}_{1N} \ = \ \frac{\sigma_N}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s<t} a_{its}(b_N;\sigma_N)\, a_{its}(b_N;\sigma_N)' \qquad \textbf{(2.11)}$$

converges in probability to D_1;

(c) $Q_{NT}(b_N;\sigma_N)$ converges in probability to Q. ■

Note that $T_{NT}(b_N;\sigma_N)=0$ by the first-order condition of the optimization problem (2.6). Because σ_N^* is of lower order than σ_N, $T_{NT}(b_N;\sigma_N^*)$ is not identically zero.

3. Empirical Example

We examine what kind of problems arise when applying the smoothed maximum score estimator, by applying the estimation procedure to an empirical model explaining labour force participation of married Dutch females in age between 18 and 65. Participation is defined as having a job or looking for a job. The α_i (individual specific effects) are introduced to deal with characteristics that are not observed and thus are not included in x_{it}. Estimates are based upon the October waves of 1984 through 1988 of the Socio−Economic Panel (SEP), drawn by the Netherlands Central Bureau of Statistics. Hence T=5. The endogenous variable (IEF) is one if the female participates, and zero if she does not. Descriptions of the endogenous and explanatory variables are given in table 1.[3]

Table 1: overview of variables

variable	description
IEF	dummy variable indicating participation of the female (IEF=1) or no participation (IEF=0)
T	time (in years after 1900)
OI	after tax other family income, excluding female's earnings and earnings of children living with the family (Dutch Guilders per week), including husband's earnings and benefits and excluding the female's benefits
HM	the number of hours per week that the male is working
NCH	number of children younger than 18 years old, living with the family
DCH6	dummy, indicating whether the family contains one or more children with an age less than 6 years. DCH6=1 if this is the case, DCH6=0 otherwise
IEM	dummy, IEM=1 if the husband is working and IEM=0 if the husband is not working
AGE2	age squared

Instead of using OI itself, the model uses the natural logarithm of (OI+1) as an explanatory variable. This variable will be denoted by LOI from now on. The variables NCH and DCH6 represent the household characteristics; IEM and HM represent the actual labour supply of the male. The female's labour force participation decision is thus made conditional on the male's actual labour supply and income. The variable T corrects for time effects, as does the variable

[3] AGE and HM are integer values.

AGE2. The variable AGE is left out because estimation is based on differences between two time periods and the difference in AGE is perfectly correlated with the difference in T. This implies that the estimated coefficient on T should be interpreted as a combination of a time effect and an age effect. The dataset used in estimation was constructed by linking the five SEP waves and selecting the married females that are present in at least two waves and for whom information on the variables of interest (see table 1) is available.[4] This yields a dataset consisting of N=3174 married Dutch females. Sample statistics are presented in table 2.

Table 2: sample statistics (11675 observations)

Variable	Mean	Standard Dev.	Minimum	Maximum
IEF	0.4277	0.4948	0	1
LOI	6.1931	0.9829	0	9.2606
HM	35.4571	17.7123	0	97
NCH	1.1522	1.1156	0	7
DCH6	0.2940	0.4556	0	1
IEM	0.8394	0.3672	0	1
AGE2	1596.02	893.6183	324	4096

In the period 1984-1988, on average 43% of the married Dutch females were participating, whereas 84% of their males had a job. Over time, labour force participation of the females increased gradually, whereas the average of IEM did not change much. The averages of NCH and DCH6 did not change that much over time although they tend to decrease slightly.

Furthermore, from the objective function it is obvious that the only observations that contain information on β are the ones for which changes in the participation have taken place, i.e. females who shifted from participating to non-participating or vice versa. This yields 2563 combinations of (y_{it},y_{is}), i=1,..,N, s,t=1,2,..,T, such that $y_{it} \neq y_{is}$. For the two subsamples $(y_{it},y_{is})=(1,0)$ and $(y_{it},y_{is})=(0,1)$, sample statistics on these differences are given in table 3.

[4] The only problem that occurred here was that for some observations OI and/or HM were/was missing (item nonresponse). These observations were left out. The initial panel contained 4268 individuals and 13629 observations; after leaving out the observations with item nonresponse, the panel shrunk to 12583 observations.

Table 3: sample statistics for differences used in estimation (2563 observations)

Variable*	Total 2563 observations	
	IEF=1 (1325 obs)	IEF=−1 (1238 obs)
T	2.129 (1.023)	2.054 (0.981)
LOI	0.031 (0.909)	0.084 (1.121)
HM	−1.004 (12.369)	−0.271 (13.685)
NCH	0.018 (0.515)	0.286 (0.740)
DCH6	−0.097 (0.362)	0.238 (0.493)
IEM	0.002 (0.262)	0 (0.273)
AGE2	150.165 (80.420)	146.439 (84.325)

[*] Note that the variables refer to differences between levels in different time periods

It must be concluded that the only effect (ignoring the standard errors) that occurs is that DCH6 has a negative effect on the willingness to participate (due to a negative effect on z^*_{ts}). For the other explanatory variables the effects are unclear.

The only exogenous variable that satisfies assumption (ii) b) is LOI and we expect it to have a non-zero effect on the willingness to participate (y^*_{it}). Therefore, the coefficient related to LOI will be normalized to one in absolute value. Before conducting smoothed maximum score, a standard probit was performed first, treating the 2563 combinations as a cross-section. The estimates will be used as a comparison to the ones resulting from smoothed maximum score. Note that, even with normally distributed error terms u_{it} in (1.1) and in the absence of individual effects, the transformed model (2.5) does not satisfy the assumptions of the probit model. However, assuming that the u_{its} in model (2.5) are i.i.d. $N(0,\sigma_u^2)$, probit fits in model (2.5). If the distributional assumptions are not valid, the probit estimator, as well as the standard errors, may be inconsistent. The probit estimator will be used to compare the estimation results with those of smoothed maximum score on the basis of the same data. The probit results both for normalization σ=1 and normalization b_{LOI}=−1 are presented in table 4. The estimator in the second column is denoted b_{probit}.

Table 4: results from probit estimation (standard errors in parentheses), dependent variable IEF

Variable	Normalization σ=1	Normalization b_{LOI}=−1
T	0.412 (0.055)	13.773 (11.150)
LOI	−0.030 (0.026)	−1 n.a.
HM	−0.011 (0.004)	-0.378 (0.315)
NCH	−0.176 (0.053)	-5.866 (5.029)
DCH6	−1.182 (0.079)	-39.482 (32.184)
IEM	0.446 (0.180)	14.912 (11.847)
AGE2	−0.005 (0.001)	-0.163 (0.132)
σ	1. n.a.	33.405 (26.972)

With the normalization σ=1, all the coefficients are significant except for b_{LOI}. The coefficients have the expected sign (except maybe for IEM). The fact that LOI does not enter the model significantly is unfortunate because its coefficient is (going to be) normalized at (minus) one. It indicates that it might be wise to carry out the optimization of the smoothed maximum score function over both $b_{LOI}=-1$ and $b_{LOI}=1$. To test whether the assumption of normality is justified by the data, a specification test was performed. For the moment, normalize σ to one and let f and Φ denote the density of the standard normal and its distribution function, respectively. We performed a LM test on $H_0{:}\gamma_1\gamma_2=0$ in the family of probability distributions $P(u_{ts}\leq t \mid w_{ts})=\Phi(t+\gamma_1 t^2+\gamma_2 t^3)$, generalizing the standard normal. This class was proposed by Ruud (1984), and Newey (1985) showed that the test statistic can easily be computed using the R^2 of an OLS regression of a vector of ones on the scores and the moments

$$m_i(b_{probit}, w_{its}) = \frac{f(b'_{probit}w_{its})\left[1(y_{it}=1,y_{is}=0) - b'_{probit}w_{its}\right]}{\Phi(b'_{probit}w_{its})\left[1-\Phi(b'_{probit}w_{its})\right]}\left[(b'_{probit}w_{its})^2, (b'_{probit}w_{its})^3\right] . \qquad \textbf{(3.1)}$$

Under the null, the distribution of the test statistic is χ^2_2. The value of the test statistic was 45.6 which leads to a rejection of the hypothesis of normally distributed errors at a significance level of 5% and it implies that we have to be a bit careful when interpreting the probit estimates.

To perform smoothed maximum score, two problems have to be solved: 1) σ_N has to be chosen and 2) a non-concave function has to be maximized. A few arbitrary choices could be made for σ_N (keeping in mind that it has to be of some order, as stated in theorem 1) and then maximize $G_{NT}(b;\sigma_N)$ w.r.t. b. This, however, does not seem to be tractable because since one does not know what σ_N should be, one would have to conduct a global optimization algorithm quite often, which is time consuming. To provide some indication of how to choose σ_N, we carried out (non-smoothed) maximum score to get a consistent estimator b_{MS} for b. b_{MS} is then used to determine σ_N as follows: transform the observations on w_{its} linearly in such a way that the sample covariance matrix of the transformed w_{its} equals the identity matrix. Transform b_{MS} in the reverse way, so that $b'w_{its}$ remains the same for all i, t and s. The smoothed maximum score objective function is drawn as a function of one of the elements in b, keeping the other values at their value in b_{MS}. This is repeated for all free parameters in b_{MS} and for various choices of σ_N. σ_N is determined as that value for which all these figures are smooth (i.e. not too erratic and not too flattened out). With the choice for σ_N, $G_{NT}(b;\sigma_N)$ can then be optimized. To save time we tried to use only a local search algorithm starting from b_{MS} (steepest descent). It appeared that the solution obtained from local search was not as good as the one returned by the global optimization algorithm. The following strategy therefore holds:

(i) calculate b_{MS} using a global optimization algorithm;

(ii) transform the data such that the empirical variance-covariance matrix equals the identity, (reversely) transform b_{MS}, choose the function K(.) and determine σ_N as described previously;

(iii) use a global optimization algorithm on the transformed dataset with the transformed estimates b_{MS} as the starting solution;

(iv) transform back the final solution.

When optimizing $G^*_{NT}(b)$ (maximum score, step (i)) over the set $|b_{LOI}|=1$, one is confronted with the problem of maximizing an objective function that has no properties that would simplify locating the global maximum (e.g. concavity); hence, one must use a global search maximization algorithm. The algorithm used is the one proposed by Corana et al. (1987). Goffe et al. (1994) show that it performs well compared to several local maximization algorithms. The algorithm runs as follows: for each free parameter an initial parameter search interval must be provided. For a given starting point (possibly randomly drawn from the search intervals) and an initial ’temperature’, T_0, compute the value of the objective function. Alter the coordinate of the first free parameter by randomly choosing an element in the parameter search interval. If the value of the objective function in this candidate point is higher, this point is accepted. If it is lower it is accepted with a probability depending on the difference in the objective function value and the temperature. The procedure is repeated for the second free parameter in the last accepted point. Repeat this until all free parameters have come in turn. The whole procedure is repeated N_S times. After that the search intervals are adjusted. A search interval is increased if many of the candidate points in this direction were accepted. The interval is decreased if few points were accepted, and the interval remains unchanged if approximately 50 percent of the candidate points in this direction are accepted. All this is repeated N_T times, after which temperature is reduced by a factor $r_T<1$ so that decreases in objective function values are less frequently accepted. Call the previous procedures a round. The last accepted point in the last round is compared with the optimal solution found so far and also with the last accepted points in the previous N_ε rounds. If the absolute value of the difference between all these points is lower than ε, the algorithm has converged. If the stopping criterion is not met, the algorithm continues with the next round. To apply the algorithm, one must choose several parameters; the choices used are mentioned in the tables. The parameters c and v have not been mentioned previously: these involve the modification of the search intervals. For the exact expressions, see Corana et al. (1987). The domain and T_0 are problem specific and choosing v equal to half the length of the initial parameter search interval performs quite well. T_0 should be chosen large relative to the range of the objective function in the domain. The optimization has to be conducted both for $b_{LOI}=1$ and $b_{LOI}=-1$. For K(.), K_4(.) is used, so h=4.

The estimation results for the maximum score estimator after normalizing $b_{LOI}=-1$ are in table 5.

Table 5: maximum score estimates[5]

Variable	Parameter Estimate
T	4.288
LOI	−1
HM	−0.228
NCH	0.552
DCH6	−14.219
IEM	9.463
AGE2	−0.046

Value objective function : 991

* Note that the variables refer to differences between levels in different time periods

The value of the objective function when $b_{LOI}=1$ was 957. For both normalizations the optimization algorithm took approximately five hours on a vax/vms mainframe. For comparison, the value of the objective function for the probit estimates as reported in the second column of table 4 is 895. This implies that using b_{MS} instead of b_{probit} leads to an increase in matching sign(y_t-y_s) with sign($\beta' w_{ts}$) from 1729 to 1777. The difference between b_{MS} and b_{probit} seems substantial when normalizing $b_{LOI}=-1$. However, if both estimators are normalized to have norm one, it appears that the estimates for T, DCH6 and AGE2 are nearly the same, whereas the estimates for the other parameters differ substantially both in sign and magnitude.

---

[5] Choices for parameters in the Corana et al. (1987) algorithm (for notation see the main text):

Domain	: [−50, 50]x{−1}x[−5, 5]x[−25,25]x[−75, 75]x[−50, 50]x[−5, 5]
c	: [2, 2, 2, 2, 2, 2] (free parameters only)
v	: [50, 5, 25, 75, 50, 5] (free parameters only)
r_T	: 0.95
T_0	: 10000
ε	: 0.000001
N_ε	: 4
N_s	: 30
N_t	: 20

To apply smoothed maximum score we have to fix the smoothness parameter. Using the previously proposed determination process for σ_N, it is fixed at 0.5. Again simulated annealing is used to locate the global optimum (step (iii)). Transforming back the optimal solution to the original data and normalizing $b_{LOI}=-1$, resulted in the estimates as reported in table 6. Call the bias corrected estimates b_{SMS}.

Table 6: smoothed maximum score estimates

Ω=I

δ=0.7

K(.)=K_4(.)

Variable	Bias corrected estimate	Bias	Standard errors
T	4.880*	0.141	0.435
LOI	−1	−	−
HM	−0.127*	−0.005	0.023
NCH	2.914*	0.133	0.563
DCH6	−15.895*	−0.811	2.127
IEM	4.762*	0.134	0.823
AGE2	−0.054*	0.001	0.006

Mean Square Error is 6.42 and choices for parameters in the Corana et al. (1987) algorithm are the same as in the previous table.

* significant at 5%

The asymptotic bias and asymptotic standard errors are calculated using the expressions in theorem 2. For Ω the identity matrix was used and the choice for δ did not change the results dramatically. The results in the table are reported for δ=0.7. I conclude that the bias is low in comparison to the standard errors and that the standard errors are low in comparison to the parameter estimates so that all the parameters are significant. Small standard errors were also encountered in Horowitz (1993), where smoothed maximum score is applied in a cross-section context.

The results shown in table 6 should be interpreted as the effect of changes in certain explanatory variables on the participation decision. The estimates imply that, *ceterus paribus*, time has a positive effect when AGE2 is low and a negative effect when AGE2 is high, that the hours that the male is working have a negative effect, that the number of children living with the family has a positive effect, that the dummy indicating whether the family contains children under the age of six years has a negative effect, that the dummy indicating whether the male participates has a positive effect and that age squared has a negative effect on the willingness to participate. The coefficient related to time consists both of a true time effect and an age effect because including age in x_{it} would lead to the same difference as the difference in time. Hence no distinction can be made between both effects. An increase in the number of working hours of an already working male increases HM and LOI and hence leads to a decrease in the willingness to participate for the female. An increase in the number of working hours for a previously unemployed male leads to negative effects on the willingness to participate through LOI and HM, but to a (relatively large) positive effect through IEM. The total effect, hence, depends on the number of hours that the male works. If the number of working hours is low, it will have a positive effect on the willingness to participate, but the effect turns negative if the amount is high. The birth of a child has a positive effect on the willingness to participate if the family already had a child under age six (such an effect seems a bit strange). On the other hand, if the family had no child under age six, the effect is severely negative.

Comparing the probit and smoothed maximum score estimates was done after normalizing the parameter estimates to norm one and the results are presented in table 7. This is done to correct for possible differences in b_{LOI} (which were normalized at -1 for both estimators). The estimates for T, HM, DCH6, IEM and AGE2 are similar for both estimators. In the probit estimates, the coefficient related to LOI is less than half the estimates in smoothed maximum score. The estimates for NCH vary both in magnitude and in sign. The standard errors for the probit estimates decreased tremendously as compared to the estimates with normalization b_{LOI}=-1 (see table 4). Except for LOI, all the coefficients are significant after normalizing $\|b\|=1$. All the parameters are significant in the smoothed maximum score estimates. It can be concluded that for most coefficients the smoothed maximum score estimates are similar to the estimates based on ordinary probit. Differences in magnitude appear for LOI and a difference in sign appears for NCH. This implies that the probit and the smoothed maximum score estimates are similar for most of the parameters, although the probit specification was rejected on the basis of a conditional moment test on the normality assumption.

Table 7: probit and bias corrected smoothed maximum score estimates with normalisation $\|b\|=1$

	probit	Smoothed MS*
T	0.307* (0.044)	0.278* (0.035)
LOI	-0.022 (0.018)	-0.057* (0.007)
HM	-0.008* (0.003)	-0.007* (0.001)
NCH	-0.131* (0.039)	0.166* (0.022)
DCH6	-0.881* (0.044)	-0.905* (0.016)
IEM	0.333* (0.121)	0.271* (0.034)
AGE2	-0.004* (0.0006)	-0.003* (0.0004)

* significant at 5%

4. Specification testing

Finally this paper will test the specification of the model on which the smoothed maximum score estimator is based. Although the model assumptions are weak, the implicit assumptions of a constant β over time and/or linearity of the effect of $\beta' x_{it}$ on y^*_{it} could be wrong. Such a test can be based on the following relationship that is implied by assumptions (i)-(v):

$$\operatorname{sign}\left(P(y_t - y_s = 1 \mid \beta' w_{ts},\ y_t \neq y_s) - \frac{1}{2}\right) = \operatorname{sign}(\beta' w_{ts}) \tag{4.1}$$

This relationship holds for all t and s, $1 \leq s < t \leq T$.

The idea is to construct uniform confidence bands for $P(y_t - y_s = 1 \mid \beta' w_{ts}, y_t \neq y_s)$, for each separate pair (s,t) using a nonparametric regression of $1(y_t - y_s = 1)$ on $b_{SMS}' w_{ts}$ for those observations for which

$y_t \neq y_s$. This was suggested by Manski as reported in Horowitz (1993, footnote 11). A requirement for the nonparametric method to apply is that $P(y_t - y_s = 1 \mid \beta' w_{ts}, y_t \neq y_s)$ is a continuous function of $\beta' w_{ts}$. The uniform confidence bands are constructed using a (slightly adapted) proposition by Horowitz (1993). Heuristically, the argument is that the (bias corrected) semiparametric estimator b_{SMS} has a larger rate of convergence than does the nonparametric kernel regression and hence β may be replaced by b_{SMS} without affecting the limiting distribution. For each (s,t), $1 \leq s < t \leq T$, use the subsample of observations for which $c_{its}=1$ and $y_{it} \neq y_{is}$. Let $\hat{F}_n(\beta' w_{ts})$ denote the nonparametric estimate for $P(y_t - y_s = 1 \mid \beta' w_{ts}, y_t \neq y_s)$. Instead of $y_t - y_s$, consider $1(y_t - y_s = 1)$. $\hat{F}_n(\beta' w_{ts})$ is essentially a weighted average of observations $1(y_t - y_s = 1)$ for which $b'_{SMS} w_{ts}$ is close to (the chosen value) of $\beta' w_{ts}$. The weights are determined by the choice of the kernel, the smoothness parameter and the distance between $b'_{SMS} w_{ts}$ and $\beta' w_{ts}$. Note that the number of observations used (n) may depend on (s,t). Let the kernel (K) be a probability density that is symmetric around zero, has bounded support, and with first derivative of bounded variation. Take the bandwidth $\omega_n = dn^{-\tau}$, $1/5 < \tau < 1/3$, $d>0$. Let f(.) denote the probability density function of $\beta' w_{ts}$. Let $\hat{f}_n$ denote the kernel estimate of f based on $b'_{SMS} w_{ts}$, kernel K and bandwidth ω_n. Let S be a closed interval on the real line on which f is strictly positive. Assume that f is twice differentiable. Then, for any real z, $x \in S$,

$$P\left(\sqrt{2\tau \log(n)}\left[\sqrt{\frac{n\omega_n}{c_K}}\ \sup_{x \in S} \sqrt{\frac{\hat{f}_n(x)}{\hat{\sigma}_n^2(x)}}\,|\hat{F}_n(x) - F(x)|\ -\ d_n\right] < z\right) \rightarrow \exp(-2\exp(-z)), \quad (n \rightarrow \infty) \qquad \textbf{(4.2)}$$

where

$$\hat{\sigma}_n^2(x) = \hat{F}_n(x)\left[1 - \hat{F}_n(x)\right] \qquad \textbf{(4.3)}$$

$$d_n = \sqrt{2\tau \log(n)} + \frac{1}{\sqrt{2\tau \log(n)}}\left(\log\left[\sqrt{\frac{C_2}{2 d^2 \pi^2}}\right]\right) \qquad \textbf{(4.4)}$$

$$c_K = \int_{-\infty}^{\infty} K(u)^2 du \qquad \textbf{(4.5)}$$

$$C_2 = \frac{1}{2c_K} \int_{-\infty}^{\infty} K'(u)^2 du \qquad \textbf{(4.6)}$$

The expression for d_n as presented in Horowitz (1993) is not completely correct. The difference is the factor d^2 in the denominator of the last term of d_n. This arises from modifying theorem 3.1 of Bickel and Rosenblatt (1973) (on which theorem 4.3.1 of Härdle (1990) is based) to more flexible bandwidths of the form $dn^{-\tau}$, $d>0$, instead of $n^{-\tau}$. The idea is to rewrite the expressions with the flexible bandwidth to the ones with bandwidth $n^{-\tau}$ and then to apply theorem 3.1 of Bickel and Rosenblatt (1973).

For each pair (s,t), s<t, the bandwidth ω_n was determined using Generalized Cross Validation as discussed in Craven and Wahba (1979). This was used instead of cross-validation because it is computationally much more convenient and appears to work quite well in practice (cf. Newey, Powell and Walker (1990)). τ is chosen to be 4/15 and for given n and ω_n this determines d. The 95% uniform confidence bands for $P(y_t-y_s=1 \mid \beta' w_{ts}, y_t \neq y_s)-0.5$ are presented in figure 1.[6] It must be concluded that the hypothesis of correct specification cannot be rejected for nearly all the combinations of (s,t), s<t. For the combination of years (84,86) the lower confidence band is above zero for values of $\beta' w_{ts}$ just below zero. This also occurs for the years (84,87) and (84,85). In the latter case, things go completely wrong for values of $\beta' w_{ts}$ between 5 and 8. The latter is caused by the few observations on $b_{SMS}' w_{ts}$ in this area. The accurate estimates as suggested by the confidence bands, are due to the fact that a limited number of observations $b_{SMS}' w_{ts}$ are used in calculating $\tilde{F}$. The observations for which $1(y_t-y_s=1)$ was zero was given most weight and hence $\tilde{F}$ is close to zero and thus $\tilde{\sigma}^2$ is also close to zero. In the area with $\beta' w_{ts}$ between 5 and 8, $\tilde{\sigma}^2$ is closer to zero than $\tilde{f}$ is. This explains the very narrow confidence bands. For all the other combinations, $b_{SMS}' w_{ts}$ was distributed more or less uniformly over the intervals displayed, so this problem does not occur there. These results might indicate that something is going on for the year 1984, although this is not immediately obvious from the data. It might suggest that β is not constant over the time period of five years, being especially different for 1984. Allowing τ to vary (keeping each d as before) led to closer confidence bands for τ=1/5 and to wider confidence bands for τ=1/3. In general, τ=1/5 led to similar figures as in figure 1 (i.e. the confidence bands did not get that much closer) whereas τ=1/3 led to better figures in the sense that the problems around $\beta' w_{ts}=0$ disappeared for the years (84,85), (84,86) and (84,87).

5. Conclusions

This paper has described a smoothed maximum score estimator for the binary choice panel data model with individual fixed/random effects. The estimator was derived combining the ideas of

---

[6] The number of observations for each combination of years are respectively 175, 210, 264, 289, 171, 244, 289, 274, 374 and 273. The bandwidths used are respectively 0.45, 1.50, 2.10, 2.00, 1.05, 2.00, 1.60, 2.60, 1.70 and 1.80.

Horowitz (1992) with those of Manski (1985, 1987). The estimator has also been extended to the case of more than two periods and an unbalanced panel under the assumption that there is no selectivity or attrition bias. Under slightly more restrictive assumptions than in Manski (1987), it is found that the smoothed maximum score estimator converges more rapidly than does that of Manski, and has a tractable asymptotic distribution. Use of a sufficiently large sample makes it possible to estimate consistently the parameters of the asymptotic distribution and to make statistical inferences. Optimizing the objective function requires a global optimization algorithm because the objective function can have many local maxima. The smoothed maximum score estimator for the binary choice panel data model with individual effects is applied to labour force participation of married Dutch females in age between 18 and 65. Interpreting the smoothed maximum score estimates yields fairly good results: most coefficients have the expected sign. For example, the coefficient related to the log of other family income is negative, the parameter related to a dummy indicating whether the family has children under age six is negative and the parameter related to age squared is negative.

Comparing the probit estimates with the bias corrected smoothed maximum score estimates, it can be concluded that the estimates for T, HM, DCH6, IEM and AGE2 are similar for both estimators. In the probit estimates, the coefficient related to LOI is less than half the estimates in smoothed maximum score. The estimates for NCH vary both in magnitude and in sign. Except for LOI, all the coefficients are significant after normalizing $\|b\|=1$. All the parameters are significant in the smoothed maximum score estimates. It can be concluded that for most coefficients the smoothed maximum score estimates are similar to the estimates based on ordinary probit. Differences in magnitude appear for LOI and a difference in sign appears for NCH. This implies that the probit and the smoothed maximum score estimates are similar for most of the parameters, although the probit specification was rejected on the basis of a conditional moment test on the normality assumption.

Finally, specification tests on the model on which the smoothed maximum score estimator is based, were performed. The hypothesis of correct specification was not rejected except for some tests where 1984 was involved. This might indicate that something is going on for the year 1984, although nothing is immediately obvious from the data.

References

Avery, R., L. Hansen and V. Hotz (1983), "Multiperiod Probit Models and Orthogonality Condition Estimation," *International Economic Review*, 24 (1), 21-35.

Bickel, P. and M. Rosenblatt (1973), "On Some Global Measures of the Deviations of Density Function Estimates," *The Annals of Statistics*, 1 (6), 1071-1095.

Chamberlain, G. (1984), "Panel Data." In *Handbook of Econometrics*, eds. Z. Griliches and M. Intrilligator, vol. 2, 1248-1318, Amsterdam: North−Holland Publishing Co.

Corana, A., M. Marchesi, C. Martini and S. Ridella (1987), "Minimizing Multimodal Functions of continuous variables with the 'Simulated Annealing Algorithm," ACM Transactions on Mathematical Software, 13, 262-280.

Craven, P. and G. Wahba (1979), "Smoothing Noisy Data with Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross Validation," *Numerische Mathematik*, 31, 377-403.

Goffe, W., G. Ferrier and J. Rogers (1994), "Global Optimization of Statistical Functions with Simulated Annealing," *Journal of Econometrics*, 60, 65-101.

Gourieroux, C. and A. Monfort (1993), "Simulation Based Inference : A Survey with Special Reference to Panel Data Models", *Journal of Econometrics*, 59, 5-33.

Härdle, W. (1990), Applied Nonparametric Regression. Cambridge University Press, New York.

Heckman, J. and R. Willis (1976), "Estimation of a Stochastic Model of Reproduction: An Econometric Approach." In *Household Production and Consumption*, ed. N. Terleckyj, New York: National Bureau of Economic Research;

Horowitz, J. (1992), "A Smoothed Maximum Score Estimator for the Binary Choice Response Model," *Econometrica*, 60 (3), 505-531.

Horowitz, J. (1993), "Semiparametric Estimation of a Work−trip Mode Choice Model," *Journal of Econometrics*, 58, 49-70.

Kim, J. and D. Pollard (1990), "Cube Root Asymptotics," *The Annals of Statistics*, 18 (1), 191-219.

Maddala, G. (1987), "Limited Dependent Variable Models Using Panel Data," *Journal of Human Resources*, 22 (3), 307-338.

Manski, C. F. (1985), "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics*, 27, 313-334.

Manski, C.F. (1987), "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55 (2), 357-362.

Newey, W. (1985), "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica*, 53 (5), 1047-1070.

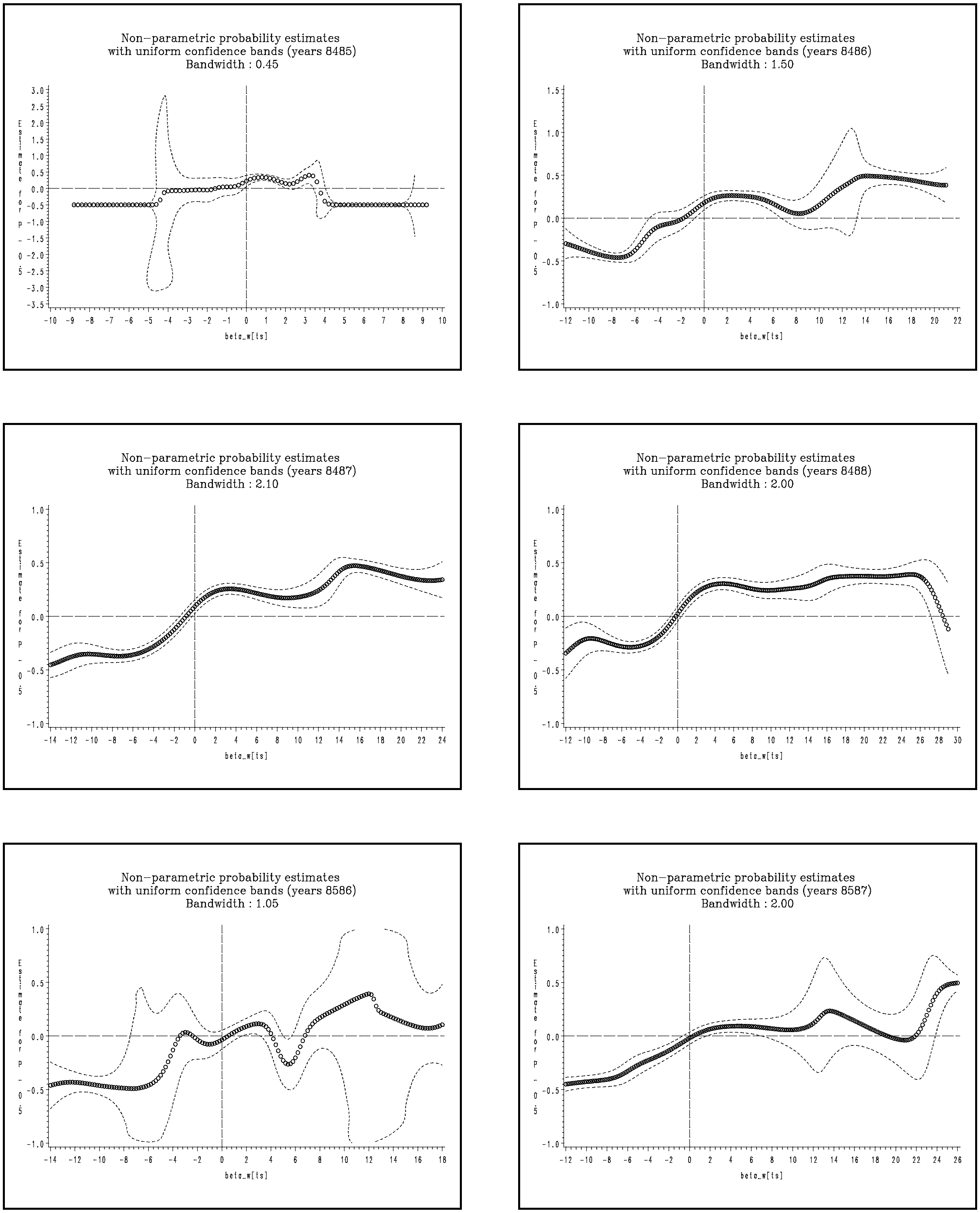
Newey, W., J. Powell and J Walker (1990), "Semiparametric Estimation of Selection Models: Some Empirical Results," *American Economic Review Papers and Proceedings*, 80 (2), 324-328.

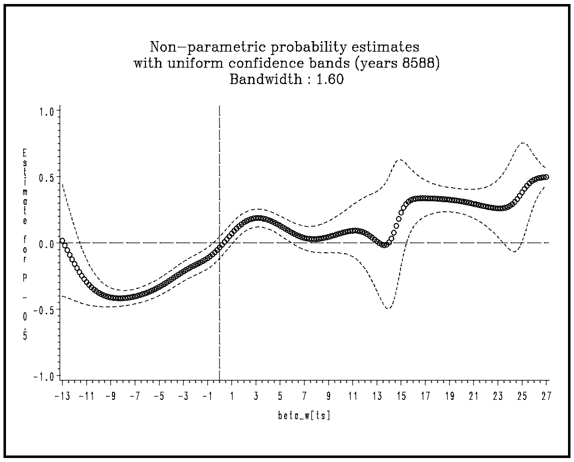
Neyman, J., and E. L. Scott (1948), "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16,1-32.

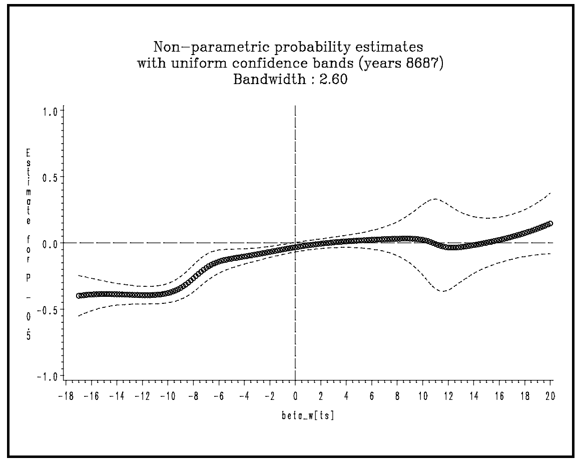
Ruud, P. (1984), "Tests of Specification in Econometrics," *Econometric Reviews*, 3, 211-242.

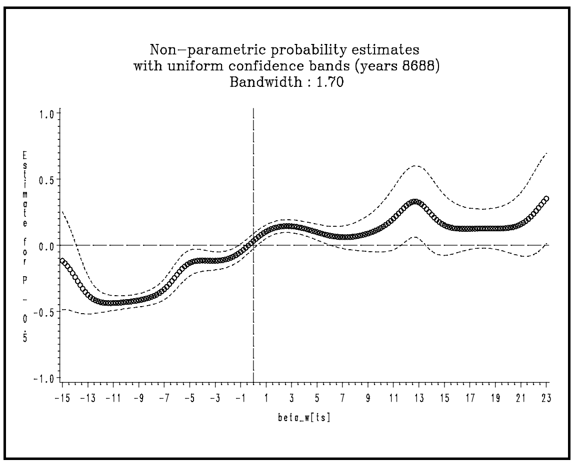
Verbeek, M. and T. Nijman (1992), "Incomplete Panels and Selection Bias." In *The Econometrics of Panel Data*, eds. L. Mátyás and P. Sevestre, Kluwer Academic Publishers, The Netherlands.

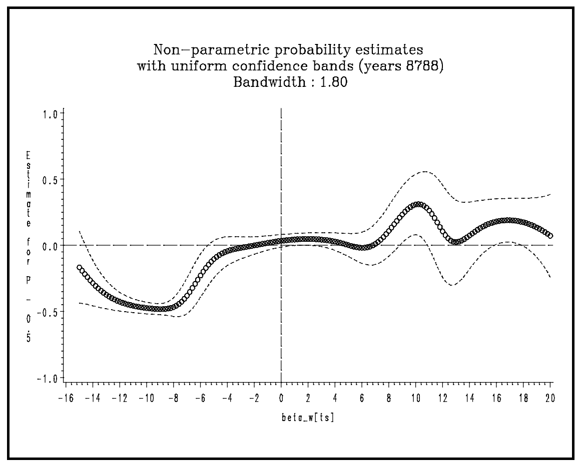
Figure 1: specification testing, 95% uniform confidence bands




Non−parametric probability estimates
with uniform confidence bands (years 8588)
Bandwidth : 1.60
beta_w[ts]


Non−parametric probability estimates
with uniform confidence bands (years 8687)
Bandwidth : 2.60
beta_w[ts]


Non−parametric probability estimates
with uniform confidence bands (years 8688)
Bandwidth : 1.70
beta_w[ts]


Non−parametric probability estimates
with uniform confidence bands (years 8788)
Bandwidth : 1.80
beta_w[ts]

APPENDIX

In this appendix proofs of the theorems stated in the text are given. These theorems in turn are proven using several lemmas. The proofs of these lemmas are also reported. The lemmas and theorems are similar to those in Horowitz (1992). The numbering of the lemmas corresponds with the numbering in Horowitz (1992) and the numbering of the theorems corresponds with the numbering in the main text. Lemmas 1 to 4 are used to prove theorem 1 (strong consistency of the smoothed maximum score estimator). This theorem, together with lemmas 5 to 9 are used to prove theorem 2 (asymptotic distribution of the smoothed maximum score estimator) and theorem 3 (consistent estimators for the matrices involved in the asymptotic distribution).

In all the lemmas and theorems one should keep in mind that the results of Horowitz (1992) are extended to panel data models with individual effects, with more than two time periods and with missing observations. Extending the results in the direction of the inclusion of individual effects and more than two time periods relies heavily on Manski (1985 and 1987) whereas the extension in the direction of unbalanced panels is possible by assuming away selectivity.

Define the expectation of $G^{*}_{NT}(b)$ by (i subscripts are suppressed)

$$G_T(b) = E\left\{ \sum_{t=2}^{T} \sum_{s<t} c_{ts} \mathrm{sign}(b^{/} w_{ts})(y_t - y_s) \right\} \tag{4.7}$$

where the expectation is taken over c_{ts}, w_{ts}, y_t and y_s.[6]

Lemma 1:

Let $b \in \{-1,1\} x \mathbb{R}^{k-1}$. Under assumptions (i), (ii) and (v), $G_T(b) \leq G_T(ß)$ with equality holding only if b=ß.

Proof:

From assumption (i) and (ii) it follows, similar to Manski (1987, lemma 3), that for all $t,s \leq T$ and all $b \in \{-1,1\} x \mathbb{R}^{k-1}$

$$E\{\mathrm{sign}(b^{/} w_{ts})(y_t - y_s)\} \leq E\{\mathrm{sign}(\beta^{/} w_{ts})(y_t - y_s)\} \tag{4.8}$$

with equality only if $b=\beta$.

This result together with assumption (v) implies that if there exist t and s, $2 \leq t \leq T$, $s<t$, such that

[6] These expressions are closely related to the definitions of H(b) and $H_N(b)$ in Manski (1987, p. 361).

$E\{c_{ts}\}>0$, then

$$G_T(b) = \sum_{t=2}^{T}\sum_{s<t} E\left\{c_{ts}\text{sign}(b'w_{ts})(y_t-y_s)\right\} \le \sum_{t=2}^{T}\sum_{s<t} E\left\{c_{ts}\text{sign}(\beta'w_{ts})(y_t-y_s)\right\} = G_T(\beta) \quad \textbf{(4.9)}$$

with equality only if b=ß. Q.E.D.

Lemma 2:

Under assumptions (iii) and (v), $G^*_{NT}(b) \to G_T(b)$ almost surely uniformly over $b \in \mathbb{R}^k$.

Proof:

Let $\sup_b f(b)$ denote the supremum of f(b) over all b. Then

$$\begin{aligned}
&\sup_b \left|G^*_{NT}(b) - G_T(b)\right| \\
&\le \sum_{t=2}^{T}\sum_{s<t} \sup_b \left|\frac{1}{N}\sum_{i=1}^{N} c_{its}\text{sign}(b'w_{its})(y_{it}-y_{is}) - E\{c_{ts}\}E\left\{\text{sign}(b'w_{ts})(y_t-y_s)\right\}\right| \\
&\le \sum_{t=2}^{T}\sum_{s<t} \Bigg[\sup_b \left|\frac{1}{N}\sum_{i=1}^{N} (c_{its}-E\{c_{ts}\})\text{sign}(b'w_{its})(y_{it}-y_{is})\right| \qquad \textbf{(4.10)} \\
&\qquad + \sup_b \left|E\{c_{ts}\}\left(E\left\{\text{sign}(b'w_{ts})(y_t-y_s)\right\} - \frac{1}{N}\sum_{i=1}^{N}\text{sign}(b'w_{its})(y_{it}-y_{is})\right)\right| \Bigg]
\end{aligned}$$

Because $|E\{c_{ts}\}|=P(c_{ts}=1)\le 1$, the second term in the summations converges to zero uniformly over b using Manski (1985, lemma 4) for each t and s, which requires assumption (iii). The first term is smaller than or equal to $|N^{-1}\Sigma^N_{i=1}(c_{its}-E\{c_{ts}\})| \sup_b |\text{sign}(b'w_{its})(y_{it}-y_{is})| \le |N^{-1}\Sigma^N_{i=1}(c_{its}-E\{c_{ts}\})|$, which converges to zero almost surely uniformly in b by the strong law of large numbers. Q.E.D.

Lemma 3:

Under assumptions (i), (ii) and (v), $G_T(b)$ is continuous at all b such that $b_1\neq 0$.

Proof:

Using (v), the result can be derived analogously to Manski (1985, lemma 5).

Lemma 4:

Under assumptions (ii) and (iii), $|G_{NT}(b;\sigma_N)-G^*_{NT}(b)| \to 0$ almost surely uniformly over $b\in B^*$ where $B^*=\{-1,1\}x\mathbb{R}^{k-1}$.

Proof:

$$|G_{NT}(b;\sigma_N) - G^*_{NT}(b)| \;\le\; \frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s<t}\left|1(b'w_{its}\ge 0) - K\left(\frac{b'w_{its}}{\sigma_N}\right)\right| \;=\; \sum_{t=2}^{T}\sum_{s<t}\frac{1}{N}\sum_{i=1}^{N}\left|1(b'w_{its}\ge 0) - K\left(\frac{b'w_{its}}{\sigma_N}\right)\right| \qquad \textbf{(4.11)}$$

Horowitz (1992, lemma 4) immediately implies that $|G_{NT}(b;\sigma_N) - G^*_{NT}(b)| \rightarrow 0$ $(N\rightarrow\infty)$ almost surely, uniformly over $b\in B^*$. Q.E.D.

Assumptions (i)−(v) and the results of lemmas 1−4 imply strong consistency of the smoothed maximum score estimator.

Proof of theorem 1:

The proof of theorem 1 is analogously to Horowitz (1992, theorem 1).

To obtain the limit distribution of the smoothed maximum score estimator for the panel data model a few additional definitions and assumptions are needed. The definitions of A, D_1 and Q are stated in the main text. Similar to Horowitz (1992, p. 509, 511) define the matrices

$$T_{NT}(b;\sigma_N) \;=\; \frac{\partial G_{NT}(b;\sigma_N)}{\partial\tilde{b}} \;=\; \frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s<t} c_{its}\,1(y_{it}\neq y_{is})\left[2*1(y_{it}=1,y_{is}=0) - 1\right]K'\left(\frac{b'w_{its}}{\sigma_N}\right)\frac{\tilde{w}_{its}}{\sigma_N} \qquad \textbf{(4.12)}$$

$$Q_{NT}(b;\sigma_N) \;=\; \frac{\partial^2 G_{NT}(b;\sigma_N)}{\partial\tilde{b}\,\partial\tilde{b}'} \;=\; \frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s<t} c_{its}\,1(y_{it}\neq y_{is})\left[2*1(y_{it}=1,y_{is}=0) - 1\right]K''\left(\frac{b'w_{its}}{\sigma_N}\right)\frac{\tilde{w}_{its}}{\sigma_N}\,\frac{\tilde{w}_{its}}{\sigma_N} \qquad \textbf{(4.13)}$$

$$\begin{aligned} D_2 \;=\; &\sum_{S} 2\,P(c_{ts}=1,c_{kl}=1)\,P(y_t\neq y_s,y_k\neq y_l) \\ &\int\left\{P(u_{ts}\ge 0,u_{kl}\ge 0\,|\,b_3) + P(u_{ts}<0,u_{kl}<0\,|\,b_3) - P(u_{ts}\ge 0,u_{kl}<0\,|\,b_3) - P(u_{ts}<0,u_{kl}\ge 0\,|\,b_3)\right\} \\ &\int\!\!\int K'(\xi_{ts})K'(\xi_{kl})\,d\xi_{ts}\,d\xi_{kl}\,\tilde{w}_{ts}\tilde{w}_{kl}\,p(0,0\,|\,b_{1'})\,dP(\tilde{w}_{ts},\tilde{w}_{kl}\,|\,y_t\neq y_s,y_k\neq y_l) \end{aligned} \qquad \textbf{(4.14)}$$

where, in case of D_2, $S=\{\{(t,s),(k,l)\}\,|\,s<t,\; l<k,\; t\neq k \text{ or } s\neq l\}$, $b_{1'}=\{\tilde{w}_{ts},\tilde{w}_{kl},y_t\neq y_s,y_k\neq y_l\}$ and $b_3=\{0,\tilde{w}_{ts},0,\tilde{w}_{kl},y_t\neq y_s,y_k\neq y_l\}$.

Apart from the terms related to c_{ts} and $y_t\neq y_s$ these expressions are similar to the ones in Horowitz (1992), with one exception: D_1 corresponds to Horowitz's D whereas D_2 is extra. The expression D_2 is a consequence of the correlation between different terms in the summation in $T_{NT}(b;\sigma_N)$ which are absent in a cross−section context.

4 Restating assumption (8) and (9) of Horowitz (1992) in terms of $F_u(-z_{ts} \mid z_{ts},\tilde{w}_{ts},y_t \neq y_s)$ and $p(z_{ts} \mid \tilde{w}_{ts},y_t \neq y_s)$ will enable us to obtain the limit distribution of the smoothed maximum score estimator for panel data as in Horowitz (1992).

Additional Assumptions (vi)−(xi):

(vi) a) The components of $\tilde{w}_{ts}$ and of the matrices $\tilde{w}_{ts}\tilde{w}_{kl}'$, $s<t$, $l<k$, and $\tilde{w}_{ts}\tilde{w}_{ts}'\tilde{w}_{kl}\tilde{w}_{kl}'$, $s<t$, $l<k$, have finite first absolute moments conditional on $(y_t \neq y_s, y_k \neq y_l)$;

b) $(\log N)/(N\sigma_N^4) \to 0$ as $N \to \infty$;

(vii) a) K is twice differentiable everywhere, $|K'(.)|$ and $|K''(.)|$ are bounded, and each of the following integrals over $(-\infty,\infty)$ is finite: $\int[K'(v)]^4 dv$, $\int[K''(v)]^2 dv$ and $\int |v^2 K''(v)| dv$;

b) for some integer $h \geq 2$ and each integer j $(0 \leq j \leq h)$, $\int |v^j K'(v)| dv < \infty$ and

$$\int_{-\infty}^{\infty} v^j K'(v)dv = \begin{cases} 0 \text{ if } j<h \\ d \text{ (nonzero) if } j=h \end{cases} \tag{5}$$

c) For any integer j between 0 and h, any $\mu>0$, and any sequence $\{\sigma_N\}$ converging to 0,

$$\begin{aligned} &\lim_{N\to\infty} \sigma_N^{j-h} \int_{|\sigma_N v| > \mu} |v^j K'(v)| dv = 0 \\ &\lim_{N\to\infty} \sigma_N^{-1} \int_{|\sigma_N v| > \mu} |K''(v)| dv = 0 \end{aligned} \tag{6}$$

(viii) For each integer j such that $1 \leq j \leq h-1$, all z_{ts} in a neighbourhood of 0, almost every $(\tilde{w}_{ts},y_t \neq y_s)$ and some $M<\infty$, $p^{(j)}(z_{ts} \mid \tilde{w}_{ts},y_t \neq y_s)$ exists and is a continuous function of z satisfying $|p^{(j)}(z_{ts} \mid w_{ts},y_t \neq y_s)| < M$. In addition, $|p(z_{ts} \mid \tilde{w}_{ts},y_t \neq y_s)| < M$ for all z and almost every $(\tilde{w}_{ts},y_t \neq y_s)$ and $|p(z_{ts},z_{kl} \mid \tilde{w}_{ts},\tilde{w}_{kl},y_t \neq y_s,y_k \neq y_l)| < M$ for all (z_{ts},z_{kl}) and almost every $(\tilde{w}_{ts},\tilde{w}_{kl},y_t \neq y_s,y_k \neq y_l)$.

(ix) For each integer j such that $1 \leq j \leq h$, all z in a neighbourhood of 0, almost every $(\tilde{w}_{ts},y_t \neq y_s)$ and some $M<\infty$, $F_u^{(j)}(-z_{ts} \mid z_{ts},\tilde{w}_{ts},y_t \neq y_s)$ exists and is a continuous function of z_{ts} satisfying $|F_u^{(j)}(-z_{ts} \mid z_{ts},\tilde{w}_{ts},y_t \neq y_s)| < M$;

(x) $ß^{\sim}$ is an interior point of $B^{\sim}$;

(xi) The matrix Q is negative definite.

Compared to Horowitz (1992) assumption (vii) b) has been extended to include j=0 which has to do with the covariance terms in $Var[T_{NT}(b;\sigma_N)]$. In assumption (viii) we have the additional requirement that $|\,p(z_{ts},z_{kl}\,|\,\tilde{w}_{ts},\tilde{w}_{kl},y_t\neq y_s,y_k\neq y_l)\,|<M$ for all (z_{ts},z_{kl}) and almost every $(\tilde{w}_{ts},\tilde{w}_{kl},y_t\neq y_s,y_k\neq y_l)$. This has to do with the same issue.

Lemma 5:

Let assumptions (i)−(iii) and (v)−(ix) hold. Then

a) $E\{\sigma_N^{-h}T_{NT}(\beta;\sigma_N)\} \rightarrow A$ $\quad (N\rightarrow\infty)$

b) $Var\{(N\sigma_N)^{1/2}T_{NT}(\beta;\sigma_N)\} \rightarrow D_1$ $\quad (N\rightarrow\infty)$

Proof:

Under assumption (v) we have

$$E\left\{\sigma_N^{-h}T_{NT}(\beta;\sigma_N)\right\} = \sigma_N^{-h}\sum_{t=2}^{T}\sum_{s<t}P(c_{ts}=1)\,P(y_t\neq y_s)\,E\left\{\left[2*1(y_t=1,y_s=0)-1\right]K'\left(\frac{\beta' w_{ts}}{\sigma_N}\right)\frac{\tilde{w}_{ts}}{\sigma_N}\,\middle|\,y_t\neq y_s\right\}$$

Analogously to Horowitz (1992, lemma 5) it can be shown that

$$\sigma_N^{-h}\;E\left\{\left[2*1(y_t=1,y_s=0)-1\right]K'\left(\frac{\beta' w_{ts}}{\sigma_N}\right)\frac{\tilde{w}_{ts}}{\sigma_N}\,\middle|\,y_t\neq y_s\right\} \qquad \textbf{(11)}$$

$$\rightarrow\; -2\int\xi^h K'(\xi)\,d\xi\sum_{i=1}^{h}\frac{1}{i!(h-i)!}\,E\left\{F_u^{(i)}(0\,|\,0,\tilde{w}_{ts},y_t\neq y_s)\,p^{(h-i)}(0\,|\,\tilde{w}_{ts},y_t\neq y_s)\,\tilde{w}_{ts}\,\middle|\,y_t\neq y_s\right\} \qquad (N\rightarrow\infty)$$

To prove part b), define $\quad t_{NT}(\beta;\sigma_N) = \sum_{t=2}^{T}\sum_{s<t}c_{ts}1(y_t\neq y_s)\left[2*1(y_t=1,y_s=0)-1\right]K'\left(\frac{\beta' w_{ts}}{\sigma_N}\right)\frac{\tilde{w}_{ts}}{\sigma_N}$,

then

$$\mathrm{Var}\left[\sqrt{N\sigma_N}\,T_{NT}(\beta;\sigma_N)\right]$$
$$= \sigma_N E\left\{t_{NT}(\beta;\sigma_N)t_{NT}(\beta;\sigma_N)^{/}\right\} + o(1)$$
$$= \sigma_N\left\{\sum_{t=2}^{T}\sum_{s<t} E\{a_{ts}a_{ts}^{/}\} + \sum_{S} 2E\{a_{ts}a_{kl}^{/}\}\right\} + o(1),$$
$$\text{where } a_{ts} = c_{ts}1(y_t \neq y_s)\left[2*1(y_t=1,y_s=0)-1\right]K^{/}\left(\frac{\beta^{/}w_{ts}}{\sigma_N}\right)\frac{\tilde{w}_{ts}}{\sigma_N}$$

We will start concentrating on $E\{a_{ts}a_{kl}^{'}\}$. We have

$$E\{a_{ts}a_{kl}^{/}\}$$
$$= E\left\{c_{ts}c_{kl}1(y_t\neq y_s)1(y_k\neq y_l)\left[2*1(y_t=1,y_s=0)-1\right]\left[2*1(y_k=1,y_l=0)-1\right]K^{/}\left(\frac{\beta^{/}w_{ts}}{\sigma_N}\right)K^{/}\left(\frac{\beta^{/}w_{kl}}{\sigma_N}\right)\frac{\tilde{w}_{ts}}{\sigma_N}\frac{\tilde{w}_{kl}^{/}}{\sigma_N}\right\}$$
$$= P(c_{ts}=1,c_{kl}=1)P(y_t\neq y_s,y_k\neq y_l)E\left\{\left[2*1(y_t=1,y_s=0)-1\right]\left[2*1(y_k=1,y_l=0)-1\right]K^{/}\left(\frac{\beta^{/}w_{ts}}{\sigma_N}\right)K^{/}\left(\frac{\beta^{/}w_{kl}}{\sigma_N}\right)\right.$$
$$\left.\frac{\tilde{w}_{ts}}{\sigma_N}\frac{\tilde{w}_{kl}^{/}}{\sigma_N}\right| y_t\neq y_s,y_k\neq y_l\Bigg\} \tag{12}$$

where the last step follows from assumption (v).

Define z_{ts}=ß'w_{ts}, z_{kl}=ß'w_{kl}, $\xi_{ts}=z_{ts}/\sigma_N$, $\xi_{kl}=z_{kl}/\sigma_N$,

$b_1=\{z_{ts},\tilde{w}_{ts},z_{kl},\tilde{w}_{kl},y_t\neq y_s,y_k\neq y_l\}$, $b_{1^{/}}=\{\tilde{w}_{ts},\tilde{w}_{kl},y_t\neq y_s,y_k\neq y_l\}$

$b_2=\{\sigma_N\xi_{ts},\tilde{w}_{ts},\sigma_N\xi_{kl},\tilde{w}_{kl},y_t\neq y_s,y_k\neq y_l\}$, and

$b_3=\{0,\tilde{w}_{ts},0,\tilde{w}_{kl},y_t\neq y_s,y_k\neq y_l\}$,

then

$$E\{a_{ts}a_{kl}^{/}\}$$
$$= P(c_{ts}=1,c_{kl}=1)P(y_t\neq y_s,y_k\neq y_l)\iiint E\left\{\left[2*1(y_t=1,y_s=0)-1\right]\left[2*1(y_k=1,y_l=0)-1\right]\middle|b_1\right\}$$
$$K^{/}\left(\frac{z_{ts}}{\sigma_N}\right)K^{/}\left(\frac{z_{kl}}{\sigma_N}\right)\frac{\tilde{w}_{ts}}{\sigma_N}\frac{\tilde{w}_{kl}^{/}}{\sigma_N}p(z_{ts},z_{kl}\,|\,b_{1^{/}})\,dz_{ts}\,dz_{kl}\,dP(\tilde{w}_{ts},\tilde{w}_{kl}\,|\,y_t\neq y_s,y_k\neq y_l)$$

$$= P(c_{ts}=1,c_{kl}=1)P(y_t\neq y_s,y_k\neq y_l)\iiint\left\{P(z_{ts}^{*}\geq 0,z_{kl}^{*}\geq 0\,|\,b_2)+P(z_{ts}^{*}<0,z_{kl}^{*}<0\,|\,b_2)-P(z_{ts}^{*}\geq 0,z_{kl}^{*}<0\,|\,b_2)\right.$$
$$\left.-P(z_{ts}^{*}<0,z_{kl}^{*}\geq 0\,|\,b_2)\right\}K^{/}(\xi_{ts})K^{/}(\xi_{kl})\tilde{w}_{ts}\tilde{w}_{kl}^{/}p(\sigma_N\xi_{ts},\sigma_N\xi_{kl}\,|\,b_{1^{/}})\,d\xi_{ts}\,d\xi_{kl}\,dP(\tilde{w}_{ts},\tilde{w}_{kl}\,|\,y_t\neq y_s,y_k\neq y_l)$$

$$= P(c_{ts}=1,c_{kl}=1)P(y_t \neq y_s, y_k \neq y_l)\int\int\int\{P(u_{ts} \geq -\sigma_N\xi_{ts}, u_{kl} \geq -\sigma_N\xi_{kl} \mid b_2) + P(u_{ts} < -\sigma_N\xi_{ts}, u_{kl} < -\sigma_N\xi_{kl} \mid b_2)$$
$$- P(u_{ts} \geq -\sigma_N\xi_{ts}, u_{kl} < -\sigma_N\xi_{kl} \mid b_2) - P(u_{ts} < -\sigma_N\xi_{ts}, u_{kl} \geq -\sigma_N\xi_{kl} \mid b_2)\} K'(\xi_{ts})K'(\xi_{kl})\tilde{w}_{ts}\tilde{w}_{kl}'$$
$$p(\sigma_N\xi_{ts},\sigma_N\xi_{kl} \mid b_{1'})\, d\xi_{ts}\, d\xi_{kl}\, dP(\tilde{w}_{ts},\tilde{w}_{kl} \mid y_t \neq y_s, y_k \neq y_l)$$

$$\rightarrow P(c_{ts}=1,c_{kl}=1)P(y_t \neq y_s, y_k \neq y_l)\int\{P(u_{ts} \geq 0, u_{kl} \geq 0 \mid b_3) + P(u_{ts} < 0, u_{kl} < 0 \mid b_3) - P(u_{ts} \geq 0, u_{kl} < 0 \mid b_3)$$
$$- P(u_{ts} < 0, u_{kl} \geq 0 \mid b_3)\}\int\int K'(\xi_{ts})K'(\xi_{kl})\, d\xi_{ts}\, d\xi_{kl}\, \tilde{w}_{ts}\tilde{w}_{kl}'\, p(0,0 \mid b_{1'})\, dP(\tilde{w}_{ts},\tilde{w}_{kl} \mid y_t \neq y_s, y_k \neq y_l) \quad (N \rightarrow \infty)$$

(13)

The last step follows from applying the Lebesgue dominated convergence theorem, using assumptions (vii) and (viii).

It now follows immediately that $\Sigma_S 2E\{a_{ts}a_{kl}'\} \rightarrow D_2$ $(N \rightarrow \infty)$, where the summation is over all elements in S.

Completely analogously it follows that

$$\sigma_N \sum_{t=2}^{T} \sum_{s<t} E\{a_{ts}a_{ts}'\}$$
$$\rightarrow \sum_{t=2}^{T} \sum_{s<t} P(c_{ts}=1)P(y_t \neq y_s)\int[K'(\xi_{ts})]^2 d\xi_{ts}\, E\{\tilde{w}_{ts}\tilde{w}_{ts}'\, p(0 \mid \tilde{w}_{ts}, y_t \neq y_s) \mid y_t \neq y_s\} \quad (N \rightarrow \infty)$$

(14)

Summarizing:

(1) $\sigma_N \Sigma_{t=2}^{T} \Sigma_{s<t} E\{a_{ts}a_{ts}'\} \rightarrow D_1$ $(N \rightarrow \infty)$;

(2) $\Sigma_{t=2}^{T} \Sigma_{s<t} E\{a_{ts}a_{ts}'\}$ does not converge as $N \rightarrow \infty$;

(3) $\sigma_N 2\Sigma_S E\{a_{ts}a_{kl}'\} \rightarrow 0$ $(N \rightarrow \infty)$;

(4) $2\Sigma_S E\{a_{ts}a_{kl}'\} \rightarrow D_2$ $(N \rightarrow \infty)$.

Lemma 5 b) follows from (1) and (3) above. Q.E.D.

Lemma 6:

Let assumptions (i)−(iii) and (v)−(ix) hold.

(a) If $N\sigma_N^{2h+1} \rightarrow \infty$ as $N \rightarrow \infty$, $\sigma_N^{-h} T_{NT}(\beta,\sigma_N)$ converges in probability to A;

(b) If $N\sigma_N^{2h+1}$ has a finite limit λ as $N \rightarrow \infty$, $(N\sigma_N)^{½} T_{NT}(\beta,\sigma_N)$ converges in distribution to $MVN(\lambda^{½}A, D_1)$.

Proof:

The proof of (a) is similar to the proof in Horowitz (1992, lemma 6), which requires (i)−(iii) and (v)−(ix). To prove (b) define

$$t_{Nits} \;=\; \sum_{t=2}^{T}\sum_{s<t} c_{its}1(y_{it}\neq y_{is})\big[2*1(y_{it}=1,y_{is}=0)-1\big]\frac{\tilde{w}_{its}}{\sigma_N}K'\left(\frac{z_{its}}{\sigma_N}\right) \tag{15}$$

Applying the results of lemma 5 and using t_{Nits} instead of the t_{Nn} in the proof of lemma 6 in Horowitz (1992), result (b) follows. Q.E.D.

Lemma 7:

Let assumptions (i)−(iii) and (vi)−(ix) hold. Assume that $\|\tilde{w}_{ts}\|\leq a$ for all t, $2\leq t\leq T$ and $s<t$ for some $a>0$. Let $\eta>0$ be such that $F_u^{(1)}(-z_{ts}\,|\,z_{ts},\tilde{w}_{ts},y_t\neq y_s)$, $F_u^{(2)}(-z_{ts}\,|\,z_{ts},\tilde{w}_{ts},y_t\neq y_s)$ and $p^{(1)}(z\,|\,\tilde{w}_{ts},y_t\neq y_s)$ exist for all t,s, and are bounded for almost every $(\tilde{w}_{ts},y_t\neq y_s)$ if $|z_{ts}|\leq\eta$. For $\Theta\in\mathbb{R}^{k-1}$, define $T^*_{NT}(\Theta)$ by

$$T^*_{NT}(\Theta) \;=\; \frac{1}{N\sigma_N^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{s<t} c_{its}1(y_{it}\neq y_{is})\big[2*1(y_{it}=1,y_{is}=0)-1\big]\tilde{w}_{its}K'\left(\frac{z_{its}}{\sigma_N}+\Theta'\tilde{w}_{its}\right)$$

Define the sets Θ_N (N=1,2,..) by $\{\Theta\,|\,\Theta\in\mathbb{R}^{k-1},\sigma_N\|\Theta\|\leq\eta/2a\}$. Then

$$\operatorname*{plim}_{N\to\infty}\;\sup_{\Theta\in\Theta_N}\left\|T^*_{NT}(\Theta)-E\{T^*_{NT}(\Theta)\}\right\| \;=\; 0 \tag{16}$$

In addition, there are finite numbers α_1 and α_2 such that, for all $\Theta\in\Theta_N$

$$\left\|E\{T^*_{NT}(\Theta)\}-Q\Theta\right\| \;\leq o(1)+\alpha_1\sigma_N\|\Theta\|+\alpha_2\sigma_N\|\Theta\|^2 \tag{17}$$

uniformly over $\Theta\in\Theta_N$.

Proof:

Define $G_{Ni}(\Theta)$ by

$$\begin{aligned} G_{Ni}(\Theta) \;=\; &\sum_{t=2}^{T}\sum_{s<t} c_{its}1(y_{it}\neq y_{is})\big[2*1(y_{it}=1,y_{is}=0)-1\big]K'\left(\frac{z_{its}}{\sigma_N}+\Theta'\tilde{w}_{its}\right)\tilde{w}_{its} \\ &- E\left\{\sum_{t=2}^{T}\sum_{s<t} c_{its}1(y_{it}\neq y_{is})\big[2*1(y_{it}=1,y_{is}=0)-1\big]K'\left(\frac{z_{its}}{\sigma_N}+\Theta'\tilde{w}_{its}\right)\tilde{w}_{its}\right\} \end{aligned} \tag{18}$$

Given any $\delta>0$, divide each set Θ_N into nonoverlapping subsets Θ_{Nj} such that the distance between any two points in the same subset does not exceed $\delta\sigma_N^2$ and the number Γ_N of subsets does not exceed $C\sigma_N^{-3(q-1)}$, then (A17) in Horowitz (1992) remains valid with g_{Nn} replaced by G_{Ni}. Using that $E\{G_{Ni}(\Theta)\}=0$ and the independence of $G_{Ni}(\Theta)$ over i, Hoeffding's inequality is still applicable (see Horowitz (1992, proof of lemma 7), though c_2 now depends on T). Assumptions (vii) a) and (vi) imply that the right hand side of (A17) of Horowitz (1992) in terms of G_{Ni} instead of g_{Nn} converges to zero as N tends to infinity and, consequently,

$$\operatorname*{plim}_{N\to\infty} \sup_{\Theta\in\Theta_N} \left\| T^*_{NT}(\Theta) - E\left\{T^*_{NT}(\Theta)\right\}\right\| = 0$$

Furthermore,

$E\{T^*_{NT}(\Theta)\}=\Sigma^T_{t=2}\Sigma_{s<t}P(c_{ts}=1)P(y_t\neq y_s)K_{N1}+K_{N2}+\Sigma^T_{t=2}\Sigma_{s<t}P(c_{ts}=1)P(y_t\neq y_s)J_{N2}+\Sigma^T_{t=2}\Sigma_{s<t}P(c_{ts}=1)P(y_t\neq y_s)I_{N2}$ (K_{N1}, J_{N2} and I_{N2} depend on t and s, but these subscripts will be dropped), where

$$K_{N1} = -2 \int\limits_{|\xi_{ts}-\Theta'\tilde{w}_{ts}|\le\frac{\eta}{\sigma_N}} F_u^{(1)}(0|0,\tilde{w}_{ts},y_t\neq y_s)p(0|\tilde{w}_{ts},y_t\neq y_s)\tilde{w}_{ts}\xi_{ts}K'(\xi_{ts})\,d\xi_{ts}\,dP(\tilde{w}_{ts}|y_t\neq y_s) \tag{19}$$

$$\begin{aligned} K_{N2} = &\sum_{t=2}^{T}\sum_{s<t} 2P(c_{ts}=1)P(y_t\neq y_s) \\ &\int\limits_{|\xi_{ts}-\Theta'\tilde{w}_{ts}|\le\frac{\eta}{\sigma_N}} F_u^{(1)}(0|0,\tilde{w}_{ts},y_t\neq y_s)p(0|\tilde{w}_{ts},y_t\neq y_s)\Theta'\tilde{w}_{ts}\tilde{w}_{ts}K'(\xi_{ts})\,d\xi_{ts}\,dP(\tilde{w}_{ts}|y_t\neq y_s)\end{aligned} \tag{20}$$

$$\begin{aligned} J_{N2} = &\frac{-1}{\sigma_N^2}\int\limits_{|z_{ts}|\le\eta}\left[2F_u^{(1)}(0|0,\tilde{w}_{ts},y_t\neq y_s)p^{(1)}(\xi_{2,ts}|\tilde{w}_{ts},y_t\neq y_s) + F_u^{(2)}(-\xi_{1,ts}|\xi_{1,ts},\tilde{w}_{ts},y_t\neq y_s)p(z_{ts}|\tilde{w}_{ts},y_t\neq y_s)\right] \\ &\tilde{w}_{ts}z_{ts}^2K'\left(\frac{z_{ts}}{\sigma_N}+\Theta'\tilde{w}_{ts}\right)dz_{ts}\,dP(\tilde{w}_{ts}|y_t\neq y_s)\end{aligned} \tag{21}$$

where $\xi_{1,ts}$ and $\xi_{2,ts}$ are between 0 and z_{ts}, and,

$$I_{N2} = \frac{1}{\sigma_N^2}\int\limits_{|z_{ts}|>\eta}\left[1-2F_u(-z_{ts}|z_{ts},\tilde{w}_{ts},y_t\neq y_s)\right]K'\left(\frac{z_{ts}}{\sigma_N}+\Theta'\tilde{w}_{ts}\right)\tilde{w}_{ts}p(z_{ts}|\tilde{w}_{ts},y_t\neq y_s)\,dz_{ts}\,dP(\tilde{w}_{ts}|y_t\neq y_s) \tag{22}$$

By assumption (vii) c) we have that

$\lim_{N\to\infty}\sup_{\Theta\in\Theta_N}\|I_{N2}\|=0$ (cf. Horowitz (1992, lemma 7, (A19))),

$\lim_{N\to\infty} \sup_{\Theta\in\Theta_N} |K_{N1}| = 0$ for all t and s (cf. Horowitz (1992, lemma 7, (A22)),

$$\lim_{N\to\infty} \left\| \sup_{\Theta\in\Theta_N} 2 \int_{|\xi_{ts}-\Theta'\tilde{w}_{ts}|\le\frac{\eta}{\sigma_N}} F_u^{(1)}(0|0,\tilde{w}_{ts},y_t\neq y_s)p(0|\tilde{w}_{ts})\Theta'\tilde{w}_{ts}\tilde{w}_{ts}K'(\xi_{ts})\,d\xi_{ts}\,dP(\tilde{w}_{ts}|y_t\neq y_s) - \Theta'Q_{ts} \right\| = 0$$

for all t and s (cf. (A24) of Horowitz (1992)), which implies that $\lim_{N\to\infty} \left\| \sup_{\Theta\in\Theta_N} \left(K_{N2} - \Theta'Q\right) \right\| = 0$.

Finally, using that $\|J_{N2}\| \le o(1)+\alpha_{1ts}\sigma_N\|\Theta\| + \alpha_{2ts}\sigma_N\|\Theta\|^2$ for some finite α_{1ts} and α_{2ts} (compare (A25) in Horowitz (1992, lemma 7)) it follows that

$$\left\|E\left\{T^*_{NT}(\Theta)\right\} - Q\Theta\right\| \le o(1) + \alpha_1\sigma_N\|\Theta\| + \alpha_2\sigma_N\|\Theta\|^2$$

where $\alpha_k=\Sigma^T_{t=2}\Sigma_{s<t}\alpha_{kts}$. Q.E.D.

Lemma 8:

Let assumptions (i)−(xi) hold, and define $\Theta_N=(\tilde{b}_N-\tilde{\beta})/\sigma_N$, where b_N is a smoothed maximum score estimator. Then $\text{plim}_{N\to\infty}\,\Theta_N=0$.

Proof:

The proof is analogous to the proof in Horowitz (1992, lemma 8), which requires (i)−(xi). The adapted lemma 7 is required in the proof.

Lemma 9:

Let assumptions (i)−(iii) and (v)−(x) hold. Let $\{ß_N\}=\{ß_{N1},\tilde{\beta}_N\}$ be any sequence in $B=\{-1,1\}x\tilde{B}$ such that $(ß_N-ß)/\sigma_N\to 0$ as $N\to\infty$. Then

$$\text{plim}_{N\to\infty}\; Q_{NT}(\beta_N;\sigma_N) = Q$$

Proof:

Assume that $ß_{N1}=ß_1$, since this is true for all sufficiently large N. Define $\Theta_N=(\tilde{\beta}_N-\tilde{\beta})/\sigma_N$. Let a_N be a sequence such that $a_N\to\infty$ and $a_N\Theta_N\to 0$ as $N\to\infty$. Define $W_N=\{\tilde{w}_{ts},\ t=1,..,T,\ s<t \mid \|\tilde{w}_{ts}\|\le a_N\}$.
Then it suffices to show that $E\{Q_{NT}(ß_N,\sigma_N)\mid W_N\}\to Q$ and $Var\{Q_{NT}(ß_N,\sigma_N)\mid W_N\}\to 0$ (see Horowitz (1992, proof of lemma 9)). Let $P_N(\tilde{w}_{ts})$ denote the distribution of $\tilde{w}_{ts}$, conditional on W_N and $y_t\neq y_s$,

and let $p_N(\tilde{w}_{ts},\tilde{w}_{kl})$ denote the distribution of $(\tilde{w}_{ts},\tilde{w}_{kl})$ conditional on W_N, $y_t \neq y_s$ and $y_k \neq y_l$. Then, using Taylor series approximations for both $F_u(.\,|\,.)$ and $p(.\,|\,.)$ around zero,

$$E\{Q_{NT}(\beta_N;\sigma_N)\,|\,W_N\} \;=\; \sum_{t=2}^{T}\sum_{s<t} P(c_{ts}=1)\left\{I_{N1} + P(y_t \neq y_s\,|\,W_N)\left[I_{N2} + I_{N3}\right]\right\} \tag{23}$$

where

$$I_{N1} \;=\; P(y_t \neq y_s\,|\,W_N)\frac{-2}{\sigma_N^2}\int_{|z_{ts}|\le\eta} \tilde{w}_{ts}\tilde{w}_{ts}^{\prime}F_u^{(1)}(0\,|\,0,\tilde{w}_{ts},y_t \neq y_s)p(0\,|\,\tilde{w}_{ts},y_t \neq y_s)\,z_{ts}\,K^{\prime\prime}\left(\frac{z_{ts}}{\sigma_N} + \Theta_N^{\prime}\tilde{w}_{ts}\right)dz_{ts}\,dP_N(\tilde{w}_{ts}) \tag{24}$$

$$\begin{aligned} I_{N2} \;=\; \frac{-1}{\sigma_N^2}\int_{|z_{ts}|\le\eta} & \tilde{w}_{ts}\tilde{w}_{ts}^{\prime}\left[2F_u^{(1)}(0\,|\,0,\tilde{w}_{ts},y_t \neq y_s)p^{(1)}(\xi_{2,ts}\,|\,\tilde{w}_{ts},y_t \neq y_s) + F_u^{(2)}(-\xi_{1,ts}\,|\,\xi_{1,ts},\tilde{w}_{ts},y_t \neq y_s)\right] \\ & z_{ts}^2 K^{\prime\prime}\left(\frac{z_{ts}}{\sigma_N} + \Theta_N^{\prime}\tilde{w}_{ts}\right)dz_{ts}\,dP_N(\tilde{w}_{ts}) \end{aligned} \tag{25}$$

$$I_{N3} \;=\; \frac{1}{\sigma_N^2}\int_{|z_{ts}|\ge\eta}\left[1 - 2F_u(-z_{ts}\,|\,z_{ts},\tilde{w}_{ts},y_t \neq y_s)\right]\tilde{w}_{ts}\tilde{w}_{ts}^{\prime}K^{\prime\prime}\left(\frac{z_{ts}}{\sigma_N} + \Theta_N^{\prime}\tilde{w}_{ts}\right)p(z_{ts}\,|\,\tilde{w}_{ts},y_t \neq y_s)\,dz_{ts}\,dP_N(\tilde{w}_{ts}) \tag{26}$$

with $\xi_{1,ts}$ and $\xi_{2,ts}$ between 0 and z_{ts}.

Similar to Horowitz (1992, lemma 9), which requires (i)−(iii) and (v)−(x), it can now be shown that $I_{N1} \to P(y_t \neq y_s)Q_{ts}$, $|\,I_{N2}\,| \to 0$ and $|\,I_{N3}\,| \to 0$ as $N \to \infty$. This immediately implies that

$$E\{Q_{NT}(\beta_n;\sigma_N)\,|\,W_N\} \;\to\; \sum_{t=2}^{T}\sum_{s<t} P(c_{ts}=1)P(y_t \neq y_s)Q_{ts} \;=\; Q \qquad (N \to \infty) \tag{27}$$

Furthermore,

$$\begin{aligned} & Var\left[Q_{NT}(\beta_N;\sigma_N)\,|\,W_N\right] \;=\; \frac{1}{N}E\{vec[q_{NT}]vec[q_{NT}]^{\prime}\,|\,W_N\} \;+\; O\left(\frac{1}{N}\right), \\ & \text{where } q_{NT} \;=\; \sum_{t=2}^{T}\sum_{s<t} c_{ts}1(y_t \neq y_s)\left[2*1(y_t=1,y_s=0)-1\right]K^{\prime\prime}\left(\frac{z_{ts}}{\sigma_N} + \Theta_N^{\prime}\tilde{w}_{ts}\right)\frac{\tilde{w}_{ts}}{\sigma_N}\frac{\tilde{w}_{ts}^{\prime}}{\sigma_N} \end{aligned} \tag{28}$$

Now

$$\frac{1}{N}E\left\{\mathrm{vec}[q_{NT}]\mathrm{vec}[q_{NT}]'\,|\,W_N\right\} = \sum_{t=2}^{T}\sum_{s<t}P(c_{ts}=1)L_1 + \sum_{S}P(c_{ts}=1,c_{kl}=1)L_2\ , \qquad \textbf{(29)}$$

with

$$L_1 = \frac{P(y_t\neq y_s\,|\,W_N)}{N\sigma_N^4}\int\!\!\int \mathrm{vec}\left[\tilde{w}_{ts}\tilde{w}_{ts}'\right]\mathrm{vec}\left[\tilde{w}_{ts}\tilde{w}_{ts}'\right]'\left[K''\left(\frac{z_{ts}}{\sigma_N}+\Theta_n'\tilde{w}_{ts}\right)\right]^2 p(z_{ts}\,|\,\tilde{w}_{ts},y_t\neq y_s)\,dz_{ts}\,dP_N(\tilde{w}_{ts}) \qquad \textbf{(30)}$$

$$\leq \frac{M}{N\sigma_N^3}\int\!\!\int \mathrm{vec}\left[\tilde{w}_{ts}\tilde{w}_{ts}'\right]\mathrm{vec}\left[\tilde{w}_{ts}\tilde{w}_{ts}'\right]'[K''(\xi_{ts})]^2\,d\xi_{ts}\,dP_N(\tilde{w}_{ts})$$

for some finite M, where $\xi_{ts}=z_{ts}/\sigma_N+\Theta_N'\tilde{w}_{ts}$, and

$$L_2 = \frac{P(y_t\neq y_s,y_k\neq y_l\,|\,W_N)}{N\sigma_N^4}\int\Big\{P(u_{ts}\geq -z_{ts},u_{kl}\leq -z_{kl}\,|\,b_2)+P(u_{ts}<-z_{ts},u_{kl}<-z_{kl}\,|\,b_2)-P(u_{ts}\geq -z_{ts},u_{kl}<-z_{kl}\,|\,b_2)$$

$$-P(u_{ts}<-z_{ts},u_{kl}\geq z_{kl}\,|\,b_2\Big\}\,\mathrm{vec}\left[\tilde{w}_{ts}\tilde{w}_{ts}'\right]\mathrm{vec}\left[\tilde{w}_{kl}\tilde{w}_{kl}'\right]'K''\left(\frac{z_{ts}}{\sigma_N}+\Theta_N'\tilde{w}_{ts}\right)K''\left(\frac{z_{kl}}{\sigma_N}+\Theta_N'\tilde{w}_{kl}\right)p(z_{ts},z_{kl}\,|\,b_1')$$

$$dz_{ts}\,dz_{kl}\,dP_N(\tilde{w}_{ts},\tilde{w}_{kl})$$

$$\leq \frac{M}{N\sigma_N^2}\int \mathrm{vec}\left[\tilde{w}_{ts}\tilde{w}_{ts}'\right]\mathrm{vec}\left[\tilde{w}_{kl}\tilde{w}_{kl}'\right]'K''(\xi_{ts})\,K''(\xi_{kl})\,d\xi_{ts}\,d\xi_{kl}\,dP_N(\tilde{w}_{ts},\tilde{w}_{kl})$$

(31)

where the last step follows from assumption (viii). Notice that L_1 is similar to (A32) of Horowitz(1992) whereas L_2 is a consequence of the correlation between different terms in the summation in $Q_{NT}(b;\sigma_N)$. Due to boundedness of both integrals in L_1 and L_2 (from assumption (vi) and (vii)) both L_1 and L_2 tend to zero as $N\rightarrow\infty$, under assumption (vi). It now follows that $\mathrm{Var}\{Q_{NT}(ß_N,\sigma_N)\,|\,W_N\}\rightarrow 0$. This completes the proof. Q.E.D.

Theorem 2 (Asymptotic Distribution):

Let assumptions (i)−(xi) hold for some $h\geq 2$, and let $\{b_N\}$ be a sequence of solutions to the maximization of problem **(3)**.

(a) If $N\sigma_N^{2h+1}\rightarrow\infty$ as $N\rightarrow\infty$, then $\sigma_N^{-h}(\tilde{b}_N-\tilde{\beta})\rightarrow^p -Q^{-1}A$;

(b) If $N\sigma_N^{2h+1}$ has a finite limit λ as $N\rightarrow\infty$, then

$$\sqrt{N\sigma_N}(\tilde{b}_N - \tilde{\beta}) \to^d \text{MVN}(-\sqrt{\lambda}Q^{-1}A, Q^{-1}D_1Q^{-1})$$

(c) Let $\sigma_N=(\lambda/N)^{1/(2h+1)}$ with $0<\lambda<\infty$; Ω be any nonstochastic, positive semidefinite matrix such that $A'Q^{-1}\Omega Q^{-1}A\neq 0$; let E_A denote the expectation with respect to the asymptotic distribution of $N^{h/(2h+1)}(\tilde{b}_N-\tilde{\beta})$, and $\text{MSE}=E_A(\tilde{b}_N-\tilde{\beta})'\Omega(\tilde{b}_N-\tilde{\beta})$. MSE is minimized by setting

$$\lambda=\lambda^*=[\text{trace}(Q^{-1}\Omega Q^{-1}D_1)]/(2hA'Q^{-1}\Omega Q^{-1}A),$$

in which case

$$N^{h/(2h+1)}(\tilde{b}_N-\tilde{\beta}) \to^d \text{MVN}(-(\lambda^*)^{h/(2h+1)}Q^{-1}A,\ (\lambda^*)^{-1/(2h+1)}Q^{-1}D_1Q^{-1}).$$

Proof of theorem 2:

Similar to Horowitz (1992, theorem 2), using theorem 1' and lemmas 6, 8 and 9, which requires (i)−(xi).

Note that the matrix D_2 does not show up here. This is caused by the fact that the covariances between different terms in the summation in $T_{NT}(b;\sigma_N)$ are of order σ_N (which tends to zero as N tends to infinity) whereas the other terms are of order 1. These other terms are represented by the matrix D_1.

Proof of theorem 3:

The proof of part (a) is exactly the same as in Horowitz (1992, theorem 3), which requires (i)−(xi).

Proof of part (b):

Let $\Theta_N=(\tilde{b}_N-\tilde{\beta})/\sigma_N$ and let $\xi_{ts}=z_{ts}/\sigma_N-\Theta_N'\tilde{w}_{ts}$, then

$$E\{\hat{D}_{1N}(b_N;\sigma_N)\}$$

$$= \frac{1}{\sigma_N}\sum_{t=2}^{T}\sum_{s<t} P(c_{ts}=1)P(y_t\neq y_s)\int\int \tilde{w}_{ts}\tilde{w}_{ts}'\left[K'\left(\frac{z_{ts}}{\sigma_N}+\Theta_N'\tilde{w}_{ts}\right)\right]^2 p(z_{ts}\,|\,\tilde{w}_{ts},y_t\neq y_s)\,dz_{ts}\,dP(\tilde{w}_{ts}\,|\,y_t\neq y_s)$$

$$= \sum_{t=2}^{T}\sum_{s<t} P(c_{ts}=1)P(y_t\neq y_s)\int\int \tilde{w}_{ts}\tilde{w}_{ts}'[K'(\xi_{ts})]^2 p(\sigma_N(\xi_{ts}-\Theta_N'\tilde{w}_{ts})\,|\,\tilde{w}_{ts},y_t\neq y_s)\,d\xi_{ts}\,dP(\tilde{w}_{ts}\,|\,y_t\neq y_s) \quad \textbf{(32)}$$

$$\to D_1 \quad (N\to\infty)$$

where the last step follows from assumptions (vi) and (viii) and Lebesgue's Dominated Convergence theorem.

Furthermore,

$$\mathrm{VAR}\left[\hat{D}_{1N}\right] = \frac{\sigma_N^2}{N}\mathrm{VAR}\left[\sum_{t=2}^{T}\sum_{s<t} a_{ts}a_{ts}'\right]$$

$$= \frac{\sigma_N^2}{N}E\left\{\sum_{t=2}^{T}\sum_{s<t} \mathrm{vec}[a_{ts}a_{ts}']\mathrm{vec}[a_{ts}a_{ts}']' + 2\sum_{S}\mathrm{vec}[a_{ts}a_{ts}']\mathrm{vec}[a_{kl}a_{kl}']'\right\} + o(1)$$

$$= \sum_{t=2}^{T}\sum_{s<t} P(c_{ts}=1)P(y_t \neq y_s)I_1 + 2\sum_{S} P(c_{ts}=1,c_{kl}=1)P(y_t \neq y_s, y_k \neq y_l)I_2 + o(1) \tag{33}$$

with

$$\begin{aligned} I_1 &= \frac{1}{N\sigma_N^2}\int\int \mathrm{vec}[\tilde{w}_{ts}\tilde{w}_{ts}']\mathrm{vec}[\tilde{w}_{ts}\tilde{w}_{ts}']'\left[K'\left(\frac{z_{ts}}{\sigma_N} + \Theta_N'\tilde{w}_{ts}\right)\right]^4 p(z_{ts}\,|\,\tilde{w}_{ts}, y_t \neq y_s)\,dz_{ts}\,dP(\tilde{w}_{ts}\,|\,y_t \neq y_s) \\ &= \frac{1}{N\sigma_N}\int\int \mathrm{vec}[\tilde{w}_{ts}\tilde{w}_{ts}']\mathrm{vec}[\tilde{w}_{ts}\tilde{w}_{ts}']'\left[K'(\xi_{ts})\right]^4 p(\sigma_N[\xi_{ts} - \Theta_N'\tilde{w}_{ts}]\,|\,\tilde{w}_{ts}, y_t \neq y_s)\,d\xi_{ts}\,dP(\tilde{w}_{ts}\,|\,y_t \neq y_s) \end{aligned} \tag{34}$$

where $\xi_{ts}=z_{ts}/\sigma_N-\Theta_N'\tilde{w}_{ts}$, and

$$\begin{aligned} I_2 &= \frac{1}{N\sigma_N^2}\int\int\int \mathrm{vec}[\tilde{w}_{ts}\tilde{w}_{ts}']\mathrm{vec}[\tilde{w}_{kl}\tilde{w}_{kl}']'\left[K'\left(\frac{z_{ts}}{\sigma_N} + \Theta_N'\tilde{w}_{ts}\right)\right]^2\left[K'\left(\frac{z_{kl}}{\sigma_N} + \Theta_N'\tilde{w}_{kl}\right)\right]^2 \\ &\qquad p(z_{ts}, z_{kl}\,|\,\tilde{w}_{ts}, y_t \neq y_s, \tilde{w}_{kl}, y_k \neq y_l)\,dz_{ts}\,dz_{kl}\,dP(\tilde{w}_{ts}, \tilde{w}_{kl}\,|\,y_t \neq y_s, y_k \neq y_l) \\ &= \frac{1}{N}\int\int\int \mathrm{vec}[\tilde{w}_{ts}\tilde{w}_{ts}']\mathrm{vec}[\tilde{w}_{kl}\tilde{w}_{kl}']'\left[K'(\xi_{ts})\right]^2\left[K'(\xi_{kl})\right]^2 \\ &\qquad p(\sigma_N[\xi_{ts} - \Theta_N'\tilde{w}_{ts}], \sigma_N[\xi_{kl} - \Theta_N'\tilde{w}_{kl}]\,|\,\tilde{w}_{ts}, y_t \neq y_s, \tilde{w}_{kl}, y_k \neq y_l)\,d\xi_{ts}\,d\xi_{kl}\,dP(\tilde{w}_{ts}, \tilde{w}_{kl}\,|\,y_t \neq y_s, y_k \neq y_l) \end{aligned} \tag{35}$$

where $\xi_{ts}=z_{ts}/\sigma_N-\Theta_N'\tilde{w}_{ts}$ and $\xi_{kl}=z_{kl}/\sigma_N-\Theta_N'\tilde{w}_{kl}$. Both I_1 and I_2 converge to 0 when N tends to infinity because both integrals are bounded as $N\rightarrow\infty$ (by assumption (vii)) and because $N\rightarrow\infty$ and $N\sigma_N\rightarrow\infty$ $(N\rightarrow\infty)$. This implies that

$$\mathrm{VAR}\left[\hat{D}_{1N}(b_N,\sigma_N)\right] \rightarrow 0 \quad (N\rightarrow\infty) \tag{36}$$

Part (c) follows immediately from lemma 9. Q.E.D.