Tilburg University

A smoothed maximum score estimator for the binary choice panel data model with an application to labour force participation

Charlier, G.W.P.; Melenberg, B.; van Soest, A.H.O.

Published in:
Statistica Neerlandica

Publication date:
1995

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):

Charlier, G. W. P., Melenberg, B., & van Soest, A. H. O. (1995). A smoothed maximum score estimator for the binary choice panel data model with an application to labour force participation. *Statistica Neerlandica*, 49(3), 324-342.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

A smoothed maximum score estimator for the binary choice panel data model with an application to labour force participation

E. Charlier, B. Melenberg and A. H. O. van Soest¹

*Department of Econometrics and CentER, Tilburg University,
P.O. Box 90153, 5000 LE, Tilburg, The Netherlands*

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. The rate of convergence is low ($N^{1/3}$) and its limit distribution cannot easily be used for statistical inference. In this paper we apply the idea of Horowitz to smooth Manski's objective function. The resulting smoothed maximum score estimator is consistent and asymptotically normal with a rate of convergence that can be made arbitrarily close to $N^{1/2}$, depending on the strength of the smoothness assumptions imposed. The estimator can be applied to panels with more than two time periods and to unbalanced panels. We apply the estimator to analyze labour force participation of married Dutch females.

Key Words & Phrases: semiparametric estimation, binary choice model, unbalanced panel.

1 Introduction

The model under consideration is a binary choice panel data model with individual effects:

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

in which $\beta \in \mathbb{R}^k$ and $1(A)$ is the indicator function which is 1 if A is true and 0 otherwise. One observes $(y_{it}, x'_{it})'$, $i = 1, \dots, N$, for some (possibly all) $t \in \{1, 2, \dots, T\}$. 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 equals one if the female participates and

¹ We thank the co-editor Tom Wansbeek, an anonymous referee and Marno Verbeek for useful comments. All remaining errors are ours. Research of the third author is possible due to a fellowship of the Netherlands Royal Academy of Arts and Sciences (KNAW). We are 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.

zero otherwise. The explanatory variables are for example household characteristics and the husband's labour supply.

Usually, in applications of this model, independence across individuals is assumed and strong assumptions are imposed with respect to (w.r.t.) the distributions of α_i and $u_i = (u_{i1}, \dots, u_{iT})$, conditional on $(x_{i1}, \dots, x_{iT})$. 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 covariance structure 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 $(x_{i1}, \dots, x_{iT})$. These models will be referred to as random effects models. In contrast, we define fixed effects models to be models where the α_i may be correlated with $(x_{i1}, \dots, x_{iT})$, without specifying the conditional distribution of α_i given $(x_{i1}, \dots, x_{iT})$. One fruitful approach for dealing with this fixed effects model is to assume a logistic distribution for u_{it} and then to use conditional maximum likelihood to estimate β , see CHAMBERLAIN (1980). This approach does not work if the u_{it} are normally distributed, see MADDALA (1987). In between random and fixed effects models is the CHAMBERLAIN (1984) model in which the α_i and u_i are conditionally normal with the conditional mean of α_i affected by $(x_{i1}, \dots, x_{iT})$, and with unrestricted covariance matrices for u_i and α_i . A Generalized Method of Moments (GMM) estimation procedure can be used to estimate β . In this paper we will focus on models in which the α_i may be correlated with $(x_{i1}, \dots, x_{iT})$ by either specifying the relationship or leaving it unrestricted.

A drawback of the models mentioned above is the distributional assumption on the 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 misspecified. To solve the analogous 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 its 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 estimator of HOROWITZ (1992), that is obtained by smoothing the maximum score objective function, such that the asymptotic behaviour can be analyzed using standard Taylor series expansions.

The literature dealing with the panel data binary choice model with individual effects that may be correlated with the explanatory variables and weak distributional assumptions for u_i is very limited. The only example, 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). The aim of this paper is to construct a consistent asymptotically normal estimator for β in model (1) with individual fixed effects, based on relatively weak assumptions. This estimator will be derived by combining the ideas

of HOROWITZ (1992) and MANSKI (1987). We extend the estimator to the case of more periods ($T \geq 2$) and allow for unbalanced panels (without selectivity or attrition). The resulting smoothed maximum score estimator is applied to a model of labour force participation of married Dutch females and the results are compared with various estimates based on stronger assumptions. In the empirical application we will focus on models in which the α_i may be correlated with $(x_{i1}, \dots, x_{iT})$ by either specifying the conditional distribution of α_i given $(x_{i1}, \dots, x_{iT})$ or leaving it unrestricted.

The remainder of this paper is organized as follows: in Section 2, the smoothed maximum score estimator for β in model (1) is defined and its asymptotic properties are derived. In Section 3 we discuss 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 (i)–(v) used to prove consistency are in the main text. The additional assumptions (vi)–(xi) to derive the asymptotic distribution as well as the proofs of theorems and lemmas are available (upon request from the first author) in CHARLIER (1994). The assumptions (vi)–(xi) are similar to assumptions 5–11 in HOROWITZ (1992).

2 Smoothed maximum score for panel data

In this section the smoothed maximum score estimation method as proposed for cross-section data by HOROWITZ (1992) is extended to the case of panel data. The starting point is model (1). The only assumptions concerning u_{it} , $t = 1, \dots, T$, are that they are time stationary conditional on $(x_{i1}, \dots, x_{iT})$ and α_i , and that the support of u_{it} is $\mathbb{R}$, see assumption (i) below. As is common in binary choice models, one normalization has to be imposed for identification. Because the distributions of the α_i and u_{it} are not characterized by finite dimensional parameters, the normalization has to concern β . First define w_{it} by $w_{its} \equiv x_{it} - x_{is}$. We assume that one component of w_{its} is absolutely continuous. Without loss of generality, we assume that this is the first one. Then the normalization imposed is $|\beta_1| = 1$. Notice that this differencing of the x s eliminates the individual effects (see below). We now introduce the objective function that is used for obtaining the estimator. Define

$$G_{NT}^*(b) = \frac{1}{N} \sum_{i=1}^N \sum_{t=2}^T \sum_{s < t} c_{its} \text{sign}(b'w_{its})(y_{it} - y_{is})$$

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. $\text{Sign}(z) = 1$ 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. Hence N can be interpreted as the number of individuals for whom at least two of the (y_{it}, x_{it}) , $t = 1, \dots, T$ are observed. For $T = 2$ and $c_{i21} = 1$ for all $i = 1, \dots, N$, maximization of $G_{NT}^*(b)$ w.r.t. b (and normalizing $\|b\| = 1$) yields the maximum score estimator of MANSKI (1987). Let

$Y = \{(y_{it}, y_{is}, x_{it}, x_{is}) | y_{it} \neq y_{is}\}$. 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 weak assumptions ((i)–(v) below), the resulting estimator can be proven to be consistent.

However, the limit distribution of this estimator is non-normal and unattractive. The reason is that $G_{NT}^*(b)$ contains the sign function, which is a step function. The idea of HOROWITZ (1992) is to smooth the objective function. This idea can also be used here. Note that maximizing $G_{NT}^*(b)$ boils down to maximizing

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N \sum_{t=2}^T \sum_{s<t} c_{its} \frac{1}{2} (\text{sign}(b'w_{its}) + 1) (y_{it} - y_{is}) \\ &= \frac{1}{N} \sum_{i=1}^N \sum_{t=2}^T \sum_{s<t} c_{its} 1(b'w_{its} \geq 0) (y_{it} - y_{is}) \end{aligned}$$

This objective function can be smoothed by replacing the indicator function $1(v)$ in the right-hand side by some smooth function $K^N(v)$ that converges to the indicator function as $N \rightarrow \infty$. Let

$$G_{NT}(b; \sigma_N) = \frac{1}{N} \sum_{i=1}^N \sum_{t=2}^T \sum_{s<t} c_{its} K\left(\frac{b'w_{its}}{\sigma_N}\right) (y_{it} - y_{is})$$

where $\sigma_N \rightarrow 0$ ($N \rightarrow \infty$) and $K(v)$ 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. $\lim_{v \rightarrow -\infty} K(v) = 0$ and $\lim_{v \rightarrow \infty} K(v) = 1$.

Thus $K^N(v) = K(v/\sigma_N)$, $v \in \mathbb{R}$. Notice that $K(v)$ could be a distribution function but it also might take on values larger than one or lower than zero and it need not be increasing. Two examples satisfying K1 and K2 are $K_2(v) = \Phi(v)$, where Φ denotes the standard normal distribution function, and (cf. HOROWITZ 1992)

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

The derivative $K'_h(v)$ of $K_h(v)$ ($h = 2, 4$) w.r.t. v is an h th order kernel, so $\int v^j K'_h(v) dv$ is zero for $j = 1, \dots, h-1$ and nonzero for $j = h$. Similar to $K_2(v)$ and $K_4(v)$, $K_h(v)$ can be constructed for any other h , such that $K'_h(v)$ is a kernel of order h . This is important since the rate of convergence is influenced by h (see theorem 2 below). In the empirical application we will restrict attention to $h = 4$.

It is easily seen that if z equals zero with probability zero, then $K(z/\sigma_N) \rightarrow 1$ ($z \geq 0$) almost surely as $N \rightarrow \infty$ (and thus $\sigma_N \rightarrow 0$). This is used to prove that $|G_{NT}(b; \sigma_N) - G_{NT}^*(b)| \rightarrow 0$ almost surely uniformly in b as $N \rightarrow \infty$. This property together with regularity assumptions as in HOROWITZ (1992) and additional assump-

tions excluding any form of selectivity bias (caused by attrition, initial nonresponse, wave nonresponse or item nonresponse, see VERBEEK and NIJMAN, 1992), is used to prove consistency of the smoothed maximum score estimator. The continuity and differentiability of $G_{NT}(b; \sigma_N)$ makes it feasible to derive the asymptotic distribution using a Taylor series expansion.

To be precise, let $x = (x_1, \dots, x_T)$ (we suppress the i subscripts from now on) and let F denote the population distribution of $\{(y_t^*, x_t, u_t; t = 1, \dots, T), \alpha\}$, cf. (1). 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} . To prove consistency of the estimator resulting from maximizing $G_{NT}(b; \sigma_N)$ over the set $\{b = (b_1, \dots, b_k): |b_1| = 1 \text{ and } (b_2, \dots, b_k) \in \tilde{B} \subset \mathbb{R}^{k-1}\}$, the following assumptions are needed (assumptions (i)–(iii) are the analogues 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 $s, t = 1, \dots, T$;
 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 the elements in w_{ts} can be rearranged such that $\beta_1 \neq 0$ and for almost every value of $\tilde{w}_{ts} = (w_{ts,2}, \dots, w_{ts,k})$, $w_{ts,1}$ has everywhere positive Lebesgue density conditional on $\tilde{w}_{ts}$ and $y_t \neq y_s$.
- (iii) A random sample is drawn from F ;
- (iv) $|\beta_1| = 1$ and $\tilde{\beta} = (\beta_2, \dots, \beta_k)'$ is contained in a compact subset $\tilde{B}$ of $\mathbb{R}^{k-1}$;
- (v) c_{ts} is independent of $(y_1, x_1, \dots, y_T, x_T)$ and $P(c_{ts} > 0) > 0$ for some t, s .

Assumption (i) a) says that the distribution of the error term in (1) is time stationary conditional on (x, α) . Assumptions (i) b) and (ii) a) are regularity conditions needed for identification. Assumption (ii) b) implies that w_{ts} should contain an absolute continuous element with nonzero coefficient. Assumptions (iii) and (iv) are self-explanatory. 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 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 $(y_1, x_1, \dots, y_T, x_T)$ is needed in deriving the limit distribution. The assumptions do not place restrictions on the distribution of α conditional on x or on the serial dependence between u_t and u_s ($s \neq t$). (i) a) allows for some time-independent form of heteroskedasticity, like $V(u_t|\alpha, x) = \exp(\alpha + \tau'x)$, $t = 1, \dots, T$, whereas it excludes for example, $\text{Var}(u_t|\alpha, x) = \exp(\alpha + \lambda'x_t)$, $t = 1, \dots, T$.

The following corollary of MANSKI (1987) 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 $(y_t - y_s | w_{ts}, y_t \neq y_s) = \text{sign}(\beta' w_{ts})$.

This conditional median restriction can be used to rewrite the model (conditional on $y_t \neq y_s$) as:

$$\begin{cases} z_{ts}^* = \beta' w_{ts} + u_{ts} & \text{and} \\ y_t - y_s = \text{sign}(z_{ts}^*) \end{cases} \quad (2)$$

and Median $(u_{ts} | w_{ts}, y_t \neq y_s) = 0$, for all i, t and s .

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

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

$$\underset{b: |b_1|=1, \tilde{b} \in \tilde{B}}{\text{maximize}} G_{NT}(b; \sigma_N) \quad (3)$$

Then $b_N \rightarrow \beta$ almost surely (a.s.).

To derive the limit distribution of the smoothed maximum score estimator we need some additional definitions and assumptions (vi)–(xi), similar to those in HOROWITZ (1992). If h increases, the assumptions (vii), (viii) and (ix) impose stronger smoothness conditions on the kernel and the distributions of u_{ts} and $\beta' w_{ts}$.

The following theorem shows the main result concerning the asymptotic distribution of the smoothed maximum score estimator (expressions for A , D and Q are in the appendix, available upon request from the first author).

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 problem (3). Then, by theorem 1, $b_{N,1} \rightarrow \beta_1$ a.s. The optimal rate of convergence in distribution for the remaining parameters $\tilde{b}_N$ is obtained for $\sigma_n = (\lambda/N)^{1/(2h+1)}$ with $0 < \lambda < \infty$ (fixed). Then

$$N^{h/(2h+1)}(\tilde{b}_N - \tilde{\beta}) \xrightarrow{d} N_{k-1}(-(\lambda)^{h/(2h+1)} Q^{-1} A, (\lambda)^{-1/(2h+1)} Q^{-1} D Q^{-1}).$$

Let Ω be any nonstochastic, positive semidefinite matrix such that $A' Q^{-1} \Omega Q^{-1} A \neq 0$ and define $MSE = \lim_{N \rightarrow \infty} E\{N^{2h/(2h+1)}(\tilde{b}_N - \tilde{\beta})' \Omega (\tilde{b}_N - \tilde{\beta})\}$. Then, for a given Ω , the MSE minimizing value for λ is $\lambda = \lambda^* = [\text{trace}(Q^{-1} \Omega Q^{-1} D)] / (2h A' Q^{-1} \Omega Q^{-1} A)$.

Note that the rate of convergence is less than $N^{1/2}$ and depends on h . By choosing h large enough the rate of convergence can be made arbitrarily close to $N^{1/2}$. 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, therefore not applicable for statistical inference (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 that can be estimated consistently (cf. HOROWITZ, 1992). The

structure of the asymptotic covariance matrix is similar to that of an extremum estimator.

Finally, if we want to use theorem 2 for inference, 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 and Q . The expressions for $T_{NT}(b_N, \sigma_N)$ and $Q_{NT}(b_N, \sigma_N)$ denote the (familiar) first and second order derivations of the objective function $G_{NT}(b, \sigma_N)$ w.r.t. $\tilde{b}$, evaluated at (b_N, σ_N) .

THEOREM 3. *Let b_N be a consistent smoothed maximum score estimator based on $\sigma_N = O(N^{-1/(2h+1)})$. For $b \in \{-1, 1\} \times \tilde{B}$ and $i = 1, \dots, N$, define*

$$a_{its}(b, \sigma) = c_{its} K' \left(\frac{b' w_{its}}{\sigma} \right) (y_{it} - y_{is}) \frac{\tilde{w}_{its}}{\sigma}$$

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

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

converges in probability to D ;

- (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 (3). Because σ_N^* is of lower order than σ_N , $T_{NT}(b_N; \sigma_n^*)$ is not identically zero.

3 Empirical application

To investigate the performance of the smoothed maximum score estimator in practice, we apply it in an empirical example and we compare the results to those of several estimators discussed in Section 1. We explain labour force participation of married Dutch females between 18 and 65 years of age, using model (1). 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 till 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 endogenous and explanatory variables are given in Table 1. The variables NCH and $DCH6$ represent family composition, IEM and HM represent the husband's actual labour supply, and EDF is the female's education level. We thus explain the female's labour force participation decision conditional on the husband's actual labour supply and income. The variable T corrects for time effects as do AGE and $AGE2$. Note that EDF is time

Table 1. Overview of variables and sample statistics (11,675 observations)

Variable	Description	Mean	Standard Dev.	Minimum	Maximum
<i>IEF</i>	Dummy variable indicating participation of the female ($IEF = 1$) or no participation ($IEF = 0$)	0.4277	0.4948	0	1
<i>LOI</i>	Logarithm of 1 + after tax other family income (Dutch Guilders per week), excluding female's earnings and benefits and earnings of children living with the family, including husband's earnings and benefits	6.1931	0.9829	0	9.2606
<i>HM</i>	the number of hours per week that the husband works	35.4571	17.7123	0	97
<i>NCH</i>	Number of children younger than 18 years old, living with the family	1.1522	1.1156	0	7
<i>DCH6</i>	Dummy, indicating whether the family contains children with age less than 6 years ($DCH6 = 1$) or not ($DCH6 = 0$).	0.2940	0.4556	0	1
<i>IEM</i>	dummy, $IEM = 1$ if the husband works and $IEM = 0$ otherwise	0.8394	0.3672	0	1
<i>EDF</i>	Education level of the female. EDF is increasing in the education level.	2.3212	0.9723	1	5
<i>AGE</i>	Age of the female in years	38.4499	10.8460	18	64
<i>T</i>	Time (year - 1900)				
<i>AGE2</i>	Age of the female squared				

constant and thus cannot be included for estimators based on differencing the x_{it} s like (smoothed) maximum score. T and AGE cannot be included both because the difference between them is time constant. 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 in Table 1 is available. For some observations, LOI and/or HM were/was missing (item nonresponse). These observations were left out. The initial panel contained 4268 individuals and 13,629 observations. Leaving out the observations with data on one year only or with item nonresponse, the dataset reduced to 3174 individuals and 11,675 observations. Sample statistics for the main variables are presented in Table 1. We conclude from Table 1 that in the period 1984–1988 on average 43% of the married Dutch females have been participating whereas 84% of their husbands had a job. Over time, labour force participation of married females increased gradually whereas the husband's participation rate did not change much. The average values of NCH and $DCH6$ decreased slightly.

From the objective function (3) it is obvious that the only observations that are used to estimate β by (smoothed) maximum score are the ones for which at least one change in the female's participation has taken place, i.e. from participation to non-participation or vice versa. This yields 2563 combinations of (y_{is}, y_{it}) , $i = 1, \dots, N$, $1 \leq s < t \leq T$, such that $y_{is} \neq y_{it}$. For the two subsamples $(y_{is}, y_{it}) = (0, 1)$ (changing from non-participation to participation) and $(y_{is}, y_{it}) = (1, 0)$ (changing from participation to non-participation) sample statistics on the differences $w_{its} = x_{it} - x_{is}$ are given in Table 2. From this table we see that T , IEM and $AGE2$ seem to have a positive effect on the willingness to participate, whereas LOI , HM , NCH and $DCH6$ seem to have a negative effect on the willingness to participate.

Table 2. Sample statistics for differences (Δ) used in estimation (2563 observations), standard errors in parentheses

Variable	Total 2563 observations	
	$\Delta IEF = 1$ (1325 obs)	$\Delta IEF = -1$ (1238 obs)
ΔT	2.129 (1.023)	2.054 (0.981)
ΔLOI	0.031 (0.909)	0.084 (1.121)
ΔHM	-1.004 (12.369)	-0.271 (13.685)
ΔNCH	0.08 (0.515)	0.286 (0.740)
$\Delta DCH6$	-0.097 (0.362)	0.238 (0.493)
ΔIEM	0.02 (0.262)	0.000 (0.273)
$\Delta AGE2$	150.165 (80.420)	146.439 (84.325)

The only exogenous variable that satisfies assumption (ii) b) is *LOI* and we expect it to have a nonzero 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 we discuss three other relevant models, discussed in Section 1. The estimators related to the first two models are based on the balanced subpanel which consists of 1017 individuals. The first model is the fixed effects logit model with the conditional maximum likelihood estimator, proposed by CHAMBERLAIN (1980). The u_{it} are assumed to be i.i.d. standard logistically distributed, i.e. one normalization is made in the distribution of the error term. To avoid the incidental parameter problem (see NEYMAN and SCOTT, 1948), the α_i are conditioned out using conditional maximum likelihood, conditional on $\Sigma_i y_{it}$. The parameter related to *EDF* is not identified and the parameters related to *T* and *AGE* are not simultaneously identified. The results with normalization $\|b\| = 1$ are presented in Table 3. The estimator is denoted by $b_{FElogit}$.

The second model is the specification suggested by CHAMBERLAIN (1984). He also allows the individual effects to be correlated with the α_i : $\alpha_i = a'x_i + v_i$ where $x_i = (x_{i1}, \dots, x_{iT})'$ and $v_i \sim N(0, \sigma_v^2)$ independent of x_i . The $u_i = (u_{i1}, \dots, u_{iT})'$ are assumed to be i.i.d. $N(0, \Sigma)$. Define $\mu_i = (\sigma_i^2 + \sigma_v^2)^{1/2}$, where σ_i^2 is the i th diagonal element of Σ . Then $P(y_{it} = 1|x_i) = \Phi(\mu_i^{-1}[\beta'x_{it} + a'x_i]) = \Phi(\pi_i'x_i)$ where

Table 3. Results from CHAMBERLAIN (1984) estimation, maximum score estimates, smoothed maximum score estimates, CHAMBERLAIN (1980) fixed effects logit estimates and probit estimates on the transformed data of Table 2 (standard errors between parentheses, bias between brackets), * means significant at the 5% level

Variable	Chamberlain (1984) $\mu_i = 1$	Maximum score estimates	Bias corrected smoothed maximum score estimates $\Omega = I, \delta = 0.7, K = K_4$	Chamberlain fixed effects logit, $\ b\ = 1$	Probit $\ b\ = 1$	Smoothed max. score, $\ b\ = 1$
<i>T</i>	-0.010 (0.007)	4.288	4.880* (0.435) [0.141]	0.216* (0.057)	0.307* (0.044)	0.278* (0.035)
<i>LOI</i>	0.003 (0.002)	-1	-1	0.046 (0.044)	-0.022 (0.018)	-0.057* (0.007)
<i>HM</i>	-0.05* (0.002)	-0.228	-0.127* (0.023) [-0.005]	-0.010* (0.03)	-0.008* (0.003)	-0.007* (0.001)
<i>NCH</i>	-0.144* (0.030)	0.552	2.914* (0.563) [0.133]	-0.109* (0.055)	-0.131* (0.039)	0.166* (0.022)
<i>DCH6</i>	-0.160* (0.060)	-14.219	-15.895* (2.127) [-0.811]	-0.837* (0.085)	-0.881* (0.044)	-0.905* (0.016)
<i>IEM</i>	0.165 (0.109)	9.463	4.762* (0.823) [0.134]	0.488* (0.155)	0.333* (0.121)	0.271* (0.034)
<i>AGE2</i>	-0.001* (0.0004)	-0.046	-0.054* (0.006) [0.001]	-0.002* (0.0007)	-0.004* (0.0006)	-0.003* (0.0004)
<i>EDF</i>	0.351* (0.047)					
<i>AGE</i>	0.057 (0.031)					
Estimates b_{Cham} b_{MS} b_{SMS} $b_{FElogit}$ b_{probit} denoted by						

$\pi_t = \mu_t^{-1}(a_1, \dots, a_{t-1}, \beta + a_t, a_{t+1}, \dots, a_r)$. The estimation procedure consists of two steps. In the first step cross-equation restrictions are ignored, and the π_t are estimated by probit for each time period separately. The second step is a minimum distance step. For identification, one of the μ_t , $t = 1, \dots, T$ has to be normalized to 1, and only (*LOI*, *HM*, *NCH*, *DCH6*, *IEM*) can be included in the fixed effect. This implies that the coefficients related to (*T*, *EDF*, *AGE*, *AGE2*) should be interpreted as partly related to the fixed effect. The results are presented in Table 3 (b_{Cham}). Specification tests on the normality assumption and the independence assumption between x_i and (v_i, u_i') were performed after the first step. We performed conditional moments tests as discussed in NEWBY (1985). The hypothesis of normally distributed errors was rejected for the years 1984 and 1985 at the 5% level. To test the hypothesis of homoskedasticity, we focused on the alternative of exponential heteroskedasticity: $\text{Var}(u_{it} + v_i | x_i) = \exp(\gamma' x_i)$. The hypothesis of homoskedasticity was rejected for all years except 1988 at the 5% level. This indicates that the model is misspecified and that the estimates b_{Cham} of β may be inconsistent.

Because smoothed maximum score basically stems from model (2), the third reference case could be a model conditional on $(y_{it} \neq y_{is})$. This essentially allows us to use the unbalanced panel. A very simple example of such a model is a standard probit on the differences, treating the 2563 combinations as a cross-section. The binary endogenous variable is 1 if the changes from non-participation to participation, and 0 in case of the reverse change. This model will be referred to as differenced probit. Note that, even with normally distributed error terms u_{it} in (1) and in the absence of individual effects, the transformed model (2) does not satisfy the assumptions of the differenced probit model. The differenced probit will only be used to give a first indication of correlations in the differenced data and to compare the estimation results with those of smoothed maximum score on the basis of the same transformation of the data. The differenced probit results for normalization $\|b\| = 1$ are presented in Table 3 (b_{probit}). To test normality, a specification test proposed by NEWBY (1985) was performed as before. Under the null, the asymptotic distribution of the test statistic is χ^2_2 . The value of the test statistic was 45.6 which leads to a clear rejection of the differenced probit specification at the 5% level. The fact that *LOI* does not enter the model significantly is unfortunate because to perform smoothed maximum score its coefficient is normalized at one in absolute value. It suggests that we should carry out the optimization of the objective function for both normalizations, 1 and -1 .

To perform smoothed maximum score, two practical problems have to be solved: 1) σ_N has to be chosen and 2) a non-concave function has to be maximized. To determine the smoothness parameter in nonparametric estimation, cross-validation or generalized cross validation (CRAVEN and WAHBA, 1979) can be used. However, the structure of the smoothed maximum score estimator does not allow these procedures to be applied here. We could perform a 'grid search' over σ_N and then, for each choice for σ_N , maximize $G_{NT}(b; \sigma_N)$ w.r.t. b . This, however, is very time consuming because $G_{NT}(b; \sigma_N)$ has no properties that simplify locating the global

maximum like e.g. concavity. Therefore we chose for a more practical solution: we carried out (non-smoothed) maximum score to get a consistent estimator b_{MS} for β . 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} is the identity matrix. Transform b_{MS} is the reverse way, so that $b'_{MS}w_{its}$ 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 elements 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 then determined as that value for which all these figures are reasonably smooth (i.e. not too erratic and not too flattened out). With the resulting choice for σ_N , $G_{NT}(b; \sigma_N)$ can then be optimized. To save computer time we have first used 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. Summarizing, the following strategy is applied:

- (i) calculate b_{MS} using a global optimization algorithm;
- (ii) transform the data such that the empirical variance-covariance matrix is the identity, (reversely) transform b_{MS} , choose the function K and determine σ_N as described above;
- (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.

The global search algorithm used in maximum score and smoothed maximum score is the one proposed by CORANA et al. (1987). GOFFE et al. (1994) find that it performs quite well when compared to several local maximization algorithms. This simulated annealing type of algorithm guarantees asymptotic convergence to the global optimum as the number of random drawings tends to infinity. The optimization is conducted with the coefficient related to LOI normalized at 1 and at -1 . For K we used K_4 , so $h = 4$ and the optimal rate of convergence is $N^{4/9}$. The estimation results for the maximum score estimator after normalizing the coefficient related to LOI at -1 are in Table 3 (b_{MS}).

The maximum value of the objective function $G_{NT}^*(b_{MS})$ when the coefficient related to LOI is normalized at -1 was 991 whereas it was 957 with normalization 1. For both maximizations, the optimization algorithm took approximately 5 hours on a vax/vms 8700 mainframe. For comparison, the value of the same objective function at p_{probit} is 895. This implies that using b_{MS} instead of b_{probit} leads to an increase in matching sign ($y_i - y_s$) with sign ($\beta'w_{is}$) from 1729 to 1777. The figures to determine σ_N induced us to choose $\sigma_N = 0.5$ for the smoothed maximum score estimator. Again simulated annealing is used to locate the global optimum (step (iii)). Correcting for the bias, transforming back the optimal solution to the original data and normalizing the coefficient related to LOI at -1 , resulted in the estimates reported in Table 3 (b_{SMS}). The asymptotic bias and asymptotic standard errors are calculated using theorem 2. For Ω the identity matrix was used because that choice

is convenient and no obvious better alternative is available. The choice of δ did not affect the results substantially. The results in the table are those for $\delta = 0.7$. From the table we conclude that the bias is low in comparison with the standard errors, and that the standard errors are low in comparison with the parameter estimates so that all the parameters are significant. Small standard errors were also encountered in cross-section applications in HOROWITZ (1992, 1993) and MELENBERG and VAN SOEST (1995). According to HOROWITZ (1992, 1993) this might imply that the sample is still too small to invoke the asymptotic properties.

The results in Table 3 should be interpreted as the effect of changes in the explanatory variables on the participation decision. Note that the coefficient related to time consists both of a true time effect and an age effect: we cannot distinguish between the two. On the basis of the smoothed maximum score estimates we conclude that, *ceteris paribus*, the time (or age) effect, consisting of a combined *AGE/Time* effect and *AGE2*, is positive for all married females in the sample. This probably reflects the increase in the female participation rate over time. An increase in *LOI* (*ceteris paribus*) has a negative effect on the wife's willingness to participate. This is not in line with the standard life cycle model with complete certainty, but does correspond to life cycle theory if the change in *LOI* is unanticipated and increases permanent income. The negative sign then implies that female leisure is a normal good. Hours worked by the husband has a negative effect suggesting that the husband's and the wife's leisure are substitutes. On the other hand however, the dummy for the husband's work has a positive effect on the willingness to participate. The number of children living with the family has a positive effect which seems counterintuitive. As expected, the dummy indicating whether the family contains children aged less than six years has a negative effect. In general, the effect of the birth of the first child on the husband's willingness to participate is less strong (see for example VAN SOEST, 1995). This suggests that the birth of the first child induces the wife to stop working. An increase in the number of working hours of an already working husband increases *LOI* and *HM* 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 husband leads to negative effects on the willingness to participate through *LOI* and *HM*, but to a positive effect through *IEM*. The total effect depends on the number of hours the husband is going to work and on this income increase.

All the estimates can be compared after normalizing the vector of parameter estimates to norm one, see Table 3. This is done to correct for possible differences in the coefficient related to *LOI* (normalized at -1 in case of b_{SMS}). The results of b_{SMS} with norm one are presented in column 7 of Table 3. As discussed before, the CHAMBERLAIN (1984) estimates, b_{Cham} , related to (*T*, *EDF*, *AGE*, *AGE2*) cannot be compared with the related coefficients in b_{SMS} with norm one, b_{MS} , b_{probit} and $b_{FElogit}$, and we focus on the remaining parameters. Because of the normalization, only relative effects are identified. Comparing for example the ratio of the parameters related to *DCH6* and *IEM* in b_{Cham} with the same ratio in b_{SMS} with norm one we

can conclude immediately that the b_{Cham} estimates differ substantially from the smoothed maximum score estimates. The same holds for the ratio of the parameters related to NCH and $DCH6$. The estimates for T , HM , $DCH6$ and $AGE2$ are very similar in b_{SMS} with norm one, b_{probit} and $b_{FElogit}$. In b_{MS} , the coefficient related to the parameters T , LOI , $DCH6$ and $AGE2$ are similar to those in b_{SMS} with norm one, but because standard errors for maximum score are not available it is difficult to draw conclusions from these results. The coefficient related to LOI in b_{probit} is only half as large as the estimate in b_{SMS} with norm one and it has the reverse sign in $b_{FElogit}$. The estimates for NCH are very different both in magnitude and in sign. For the fixed effects logit and differenced probit estimates, all the coefficients are significant, except for LOI . All coefficients are significant in the smoothed maximum score estimates. We conclude that for most of the coefficients the smoothed maximum score estimates are similar to the estimates based on fixed effects logit or differenced probit. However, differences in sign and magnitude appear for LOI and a significant difference in sign appears for NCH .

4 Specification testing

The final topic considered in this paper is testing the specification of the model underlying the smoothed maximum score estimator. Although the model assumptions are weak, the normalization $|\beta_1| = 1$ and implicit assumptions of a constant beta over time and/or linearity of the effect of x_{it} and y_{it}^* could be wrong. A test can be based on the following relationship:

$$\text{sign}(P(y_t - y_s = 1 | \beta'w_{ts}, y_t \neq y_s) - \frac{1}{2}) = \text{sign}(\beta'w_{ts})$$

Given a correct specification, this relationship should hold for all relevant t and s , with $s < t$. The idea is to construct uniform confidence bands for $P(y_t - y_s = 1 | \beta'w_{ts}, y_t \neq y_s)$, 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 $P(y_t - y_s = 1 | \beta'w_{ts}, y_t \neq y_s)$ 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 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 $P(y_t - y_s = 1 | \beta'w_{ts}, y_t \neq y_s)$. 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, with bounded support, and with first derivative of bounded variation. We take the bandwidth $\omega_n = dn^{-\tau}$, $1/5 < \tau < 1/3$, $d > 0$. Let f denote the probability density function of $\beta'w_{is}$. Let $\hat{f}_n$ denote the kernel estimate of f based on $b'_{SMS}w_{is}$, 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 the uniform confidence bounds are based upon the following: for any real z , $x \in S$,

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

where

$$d_n = \sqrt{(2\tau \log(n))} + \frac{1}{\sqrt{(2\tau \log(n))}} \left(\log \left[\sqrt{\left(\frac{C_2}{2d^2\pi^2}\right)} \right] \right) \\ \hat{\sigma}_n^2(x) = \hat{F}_n(x)[1 - \hat{F}_n(x)] \\ d_n = \sqrt{(2\tau \log(n))} + \frac{1}{\sqrt{(2\tau \log(n))}} \left(\log \left[\sqrt{\left(\frac{C_2}{2d^2\pi^2}\right)} \right] \right) \\ c_K = \int_{-\infty}^{\infty} K(u)^2 du \\ C_2 = \frac{1}{2c_K} \int_{-\infty}^{\infty} K'(u)^2 du$$

Note that we have slightly changed the expression for d_n as presented in HOROWITZ (1993). 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) 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 apply theorem 3.1 of BICKEL and ROSENBLATT (1973).

We used the Gaussian kernel. Although it does not have bounded support, the results will not change when the truncation point is sufficiently large. 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. NEWBY, 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 $P(y_t - y_s = 1 | \beta'w_{is}, y_t \neq y_s) - 0.5$ are presented in Figure 1. 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. If the specification is correct, then for negative values of $b'_{SMS}w_{is}$ (on the horizontal axis) the lower bound for $P(y_t - y_s = 1 | \beta'w_{is}, y_t \neq y_s) - 0.5$ (vertical axis) should be below the horizontal axis

Fig. 1. Specification testing, point estimates and 95% uniform confidence bands.

Fig. 1. (continued)

and for positive or zero values of $b'_{SMS}w_{is}$ the upper bound should be above the horizontal axis. We conclude that the hypothesis of correct specification is rejected for the combination of years (84, 86), (84, 87) and (84, 85), since the lower confidence band is above zero for values of $\beta'w_{is}$ just below zero. In the latter case things go completely wrong for values of $\beta'w_{is}$ between 5 and 8. This is probably caused by the few observations on $b'_{SMS}w_{is}$ in this area. The very accurate estimates as suggested by the confidence bands are due to the fact that a few observations $b'_{SMS}w_{is}$ are used in calculating $\tilde{F}$. For all the other combinations $b'_{SMS}w_{is}$ was distributed more or less uniformly over the intervals displayed, so this problem did not occur. A possible explanation for the problems with 1984 is that the data for 1984 might suffer from panel set up problems. Letting τ vary to change the bandwidth (keeping each d as before) resulted in narrower confidence bands for $\tau = 1/5$ and to wider confidence bands for $\tau = 1/3$. In general, $\tau = 1/5$ led to similar figures as in Figure 1 (i.e. the confidence bands did not get much closer) whereas $\tau = 1/3$ led to better figures in the sense that the problems around $\beta'w_{is} = 0$ disappeared for the years (84, 85), (84, 86) and (84, 87).

5 Conclusions

This paper describes a smoothed maximum score estimator for the binary choice panel data model with individual fixed effects. The estimator is derived combining the ideas of HOROWITZ (1992) with those of MANSKI (1985, 1987). The estimator is also 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), we show that the smoothed maximum score estimator converges more rapidly than Manski's does, and that it has a tractable asymptotic distribution. With a sufficient large sample, the parameters of the asymptotic distribution can be estimated consistently and statistical inference is possible. Optimizing the objective function requires a global optimization algorithm because the objective function can have many local maxima. The algorithm proposed by CORANA et al. (1987) suits this purpose. We find that it performs well compared with the local optimization algorithms. 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. 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 contains children in age less than 6 is negative, and the parameter related to age squared is negative.

Compared with the CHAMBERLAIN (1984) estimates, the smoothed maximum score estimates differ substantially. Specification tests on the normality and homoskedasticity assumptions in the CHAMBERLAIN (1984) model led to rejection of these assumptions. Compared with probit on the differenced data, the CHAMBERLAIN (1980) fixed effects logit estimates are similar. The normality in the differenced probit model is rejected. Comparing these estimates with the smoothed maximum score estimates we conclude that, although for most coefficients the smoothed maximum score estimates are similar to the estimates based on fixed effects logit or differenced probit, differences in sign and magnitude appear for *LOI* and a significant difference in sign appears for *NCH*, which indicates that the estimates based on differenced probit or fixed effects logit lead to the wrong conclusion as far as the effect of the variable *NCH* is concerned.

Finally, we performed specification tests of the model on which the smoothed maximum score estimator is based. The hypothesis of no misspecification is not rejected at the 5% level except for some tests where 1984 was involved. This might indicate that something is going on for the year 1984 that might have to do with panel data set up problems.

References

- EVERY, R., L. HANSEN and V. HOTZ (1983), Multiperiod probit models and orthogonality condition estimation, *International Economic Review* 24, 21–35.

- BICKEL, P. and M. ROSENBLATT (1973), On some global measures of the deviations of density function estimates, *The Annals of Statistics* 1, 1071–1095.
- CHAMBERLAIN, G. (1980), Analysis of covariance with qualitative data, *Review of Economic Studies* 47, 225–228.
- CHAMBERLAIN, G. (1984), Panel data, North-Holland Publishing Co, Amsterdam, in: Z. GRILICHES and M. INTRILIGATOR (eds.), *Handbook of Econometrics*, 1248–1318.
- CHARLIER, E. (1994), A smoothed maximum score estimator for the binary choice panel data model with individual fixed effects and application to labour force participation, CentER Discussion Paper no. 9481, Tilburg University.
- 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.
- HECKMAN, J. and R. WILLIS (1976), Estimation of a stochastic model of reproduction: an econometric approach, National Bureau of Economic Research. New York, in: N. TERLECKYJ (ed.), *Household production and consumption*.
- HOROWITZ, J. (1992), A smoothed maximum score estimator for the binary choice response model, *Econometrica* 60, 505–531.
- HOROWITZ, J. (1993), Semiparametric estimation of a work-tip model choice model, *Journal of Econometrics* 58, 49–70.
- KIM, J. and D. POLLARD (1990), Cube root asymptotics, *The Annals of Statistics* 18, 191–219.
- MACURDY, T. (1981), An empirical model of labor supply in a life-cycle setting, *Journal of Political Economy* 89, 1059–1085.
- MADDALA, G. (1987), Limited dependent variable models using panel data, *Journal of Human Resources* 22, 307–333.
- 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, 357–362.
- MELENBERG, B. and A. VAN SOEST (1995), Parametric and semiparametric estimation of ordered response models with an application to satisfaction with life, working paper, Tilburg University.
- NEWBY, W., J. POWELL and J. WALKER (1990), Semiparametric estimation of selection models: some empirical results, *American Economic Review Papers and Proceedings* 80, 324–328.
- NEYMAN, J., and E. L. SCOTT (1948), Consistent estimates based on partially consistent observations, *Econometrica* 16, 1–32.
- VAN SOEST, A. (1995), Structural models of family labour supply: a discrete choice approach, *Journal of Human Resources*, forthcoming.
- VERBEEK, M. and T. NIJMAN (1992), Incomplete panels and selection bias, in: L. MÁTYÁS and P. SEVESTRE (eds.), *The econometrics of panel data*, Kluwer Academic Publishers, The Netherlands, 262–302.

Received: May 1994. Revised: February 1995.