Parametric and Semi-Parametric Estimation of the Binary Response Model of Labour Market Participation
Author(s): Michael Gerfin
Source: *Journal of Applied Econometrics*, Vol. 11, No. 3 (May - Jun., 1996), pp. 321-339
Published by: John Wiley & Sons
Stable URL: http://www.jstor.org/stable/2285067
Accessed: 23/04/2009 06:22

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=jwiley.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.

John Wiley & Sons is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Applied Econometrics.*

http://www.jstor.org

JOURNAL OF APPLIED ECONOMETRICS, VOL. 11, 321–339 (1996)

PARAMETRIC AND SEMI-PARAMETRIC ESTIMATION OF THE BINARY RESPONSE MODEL OF LABOUR MARKET PARTICIPATION

MICHAEL GERFIN
Volkswirtschaftliches Institut, Abteilung für angewandte Wirtschaftsforschung und Wirtschaftspolitik, Universität Bern, Gesellschaftsstr. 49, 3012 Bern, Switzerland
E-Mail: GERFIN@VWI.UNIBE.CH

SUMMARY

This paper compares the familiar probit model with three semiparametric estimators of binary response models in an application to labour market participation of married women. This exercise is performed using two different cross-section data sets from Switzerland and Germany. For the Swiss data the probit specification cannot be rejected and the models yield similar results. In the German case the probit model is rejected, but the coefficient estimates do not vary substantially across the models. The predicted choice probabilities, however, differ systematically for a subset of the sample. The results of this paper indicate that more work is necessary on specification tests of semiparametric models and on simulations using these models.

1. INTRODUCTION

Binary response models are of major importance in applied microeconometrics. Examples include labour market participation, union membership, choice of transportation mode for travel to work, the choice to seek medical care, or the choice to participate in welfare programs. Apart from being interesting in their own right, binary response models are also often used to correct for selectivity bias in censoring models. The most frequently used form of binary response models is given by

$$\begin{aligned} y_n^* &= x_n\beta + u_n \\ y_n &= 1 \text{ if } y_n^* \geqslant 0 \\ y_n &= 0 \text{ otherwise} \end{aligned} \tag{1}$$

where y_n is the indicator of the nth individual's response determined by the underlying latent variable y_n^*, x_n is a $1 \times q$ vector of explanatory variables, β is a $q \times 1$ vector of parameters, u_n is a random error term, and $n = 1, \ldots, N$.

Let $F(u \mid x)$ denote the cumulative distribution function of u conditional on the event $x_n = x$. Then

$$P(y_n = 1 \mid x_n, \beta) = F(x_n\beta \mid x_n) \tag{2}$$

In most applications, it is assumed that F is either the cumulative normal (probit model) or the cumulative logistic distribution function (logit model).

CCC 0883–7252/96/030321–19
© 1996 by John Wiley & Sons, Ltd.

Received December 1993
Revised August 1995

In practice, there usually is no prior knowledge to justify this distributional assumption. Hence it is necessary to test whether it is consistent with the data. In the past years a number of specification tests for binary response models have been developed (see Blundell, 1987). However, these tests are still not standard practice in applied work although they are easily computed. This is unfortunate, because several Monte Carlo studies have shown that both the probit and the logit estimators can be severely biased when the distribution of u is heteroscedastic or asymmetric (see e.g. Manski and Thompson, 1986; Powell *et al.*, 1989; Klein and Spady, 1993; Horowitz, 1992). Ruud (1983) has shown that under specific assumptions about the distribution of x the estimates of β can be consistent up to location and scale even when the parametric model is incorrect. However, as Ruud (1983) states, these assumptions are too restrictive to be generally applicable. Furthermore, Ruud's 'robustness result' does not carry over to choice probabilities and probability changes. But exactly these are of major interest in most applied work.

In principle, the probability $P(y_n = 1 \mid x_n)$ could be estimated by nonparametric regression without making any distributional assumptions. However, this estimation will suffer from the 'curse of dimensionality' when x is multidimensional (Stone, 1980). Furthermore, in applied work we are usually interested in β and $F(x_n\beta \mid x_n)$ which are not identified by nonparametric regression. If y is determined by utility maximization β contains important behavioural information. Second, if β and $F(\cdot)$ are known it is possible to predict y at values of x outside the support of x.

These problems have motivated the development of semiparametric models of binary response that allow β to be estimated consistently without specifying the distribution of u. In this paper I consider three semiparametric estimators that have recently been proposed in the literature. The paper provides nothing new from a theoretical point of view, but aims at illustrating the performance of these estimators in an application to real data. The first estimator is based on the semi-nonparametric approach of Gallant and Nychka (1987), adapted for the binary response case by Gabler *et al.* (1993). The second estimator has been proposed by Klein and Spady (1993) and belongs to the class of single-index models. Finally, the smoothed maximum score estimator proposed by Horowitz (1992) is applied.

These semiparametric estimators differ in several respects. First, the restrictions imposed by the underlying distributional assumptions vary considerably across the models. The models also differ in their ability to predict y. For the probit and the semi-nonparametric models $F(x_n\beta \mid x_n)$ is identified for any value of x. In the single-index model and the smoothed maximum score model, $F(x_n\beta \mid x_n)$ can be estimated by nonparametric regression in the support of $x_n\beta$. However, for the smoothed maximum score model this estimate will not be very precise if x is multidimensional (Horowitz, 1993). The usefulness of the Klein–Spady and especially of the smoothed maximum score model for simulations is limited for this reason.

Semiparametric methods for binary response models have not yet been used much in applications. To my knowledge the only study that systematically compares different semiparametric estimators is Horowitz (1993). He estimated a fixed coefficient probit model, a random coefficient probit model, a single-index model, and two versions of maximum score models of the choice of the transportation mode to work. Specification tests rejected the fixed coefficient probit model and the single-index model. The random coefficient and the smoothed maximum score model could not be rejected. The coefficient estimates and predicted choice probabilities varied considerably across models. Newey *et al.* (1990) estimated a probit and a semiparametric single-index model of labour market participation. The estimates of the parameters of both models turned out to be quite close. A similar result was obtained by Melenberg and van Soest (1991), who estimated the probability to go on vacation using a probit and a semiparametric single-index model. Das (1991) estimated logit and maximum score models of the decision whether to idle a cement kiln. An informal examination of the estimation results suggested that the logit model may have been misspecified, but no formal test was carried out.

In this paper I estimate a reduced-form labour force participation model of married women employing the four models mentioned above using data from Germany and Switzerland. In Section 2 I describe the different models. Section 3 describes the data. In Section 4 I present the results, specification tests and the choice probabilities implied by the different models. Section 5 concludes.

2. THE MODELS

The first three models are estimated by Maximum Likelihood. The general likelihood function for these models can be written as

$$\ln L(\beta) \sum_{n-1}^{N} y_n \ln[F(x_n\beta \mid x_n)] + (1 - y_n)\ln[1 - F(x_n\beta \mid x_n)] \tag{3}$$

where $F(x_n\beta \mid x_n)$ is defined by equation (2). The models below differ only by the specification of $F(x_n\beta \mid x_n)$.

2.1. Probit

Assuming a normal distribution for u independent of x yields the familiar probit specification. The probit ML estimator of β, b_{pr}, is obtained by maximizing the likelihood function (3) with

$$F(x_n\beta \mid x_n) = F(x_n\beta) = \Phi(x_n b_{\text{pr}}/\sigma_u) \tag{4}$$

where Φ is the cumulative normal distribution function and σ_u the standard deviation of u. The model is identified only up to a constant scale factor. Usually, scale normalization is achieved by setting σ_u equal to 1. The probit results and specification tests will be presented based on the this normalization. In order to compare the probit estimates with the semiparametric estimates another normalization is preferable, since for the latter estimates it would be cumbersome to restrict the variance to 1. I choose to set the coefficient of one of the components of x, non-labour income, equal to -1.

2.2. The Semi-nonparametric Model (Gabler, Laisney, and Lechner, 1993)

In the semi-nonparametric model, based on an approach by Gallant and Nychka (1987), the density underlying the distribution function $F(x_n\beta)$ is given by the Hermite form

$$h^*(u) = \sum_{i,j=0}^{K} \alpha_i \alpha_j u^{i+j} \exp(-(u/\delta^2) \tag{5}$$

which approximates any smooth density that has a moment generating function with tails at most as fat as the t-distribution.

Following Gabler *et al.* scale normalization is achieved by setting δ equal to $\sqrt{2}$. Furthermore, we have to ensure that the density of u, $h(u)$, integrates to unity. This is achieved by defining $h(u)$ as $h^*(u)/S$, where

$$S = \int_{-\infty}^{\infty} h^*(u)\,\mathrm{d}u \tag{6}$$

Since this is invariant through multiplication of α by a scalar, a further normalization is necessary. Hence α_0 is set to $(2\pi)^{-1/4}$, which leads to the standard normal for $K=0$. The seminonparametric model nests the probit model and provides a test for the normality assumption.

The semi-nonparametric estimator b_{snp} of β is obtained by maximizing the likelihood function (2) with

$$F(x_n b_{\text{snp}} \mid x_n) = S^{-1} \int_{-x_n b_{\text{snp}}}^{\infty} h^*(u)\,\mathrm{d}u \tag{7}$$

conditional on a given value of K.

Finally, a location normalization is necessary. Following again Gabler *et al.* this is achieved by imposing $E(u)=0$ (see Melenberg and van Soest, 1993, for an alternative normalization). Gabler *et al.* show that this restriction leads to a branching problem with different sets of restrictions that have to be imposed on either α_1 or α_2. They propose to estimate under either assumptions separately and compare the values of the objective function obtained at the optimum (cf. Gabler *et al.* for details).

The proposed quasi-maximum likelihood estimator is asymptotically normal and allows for the familiar testing techniques. The assumption of homoscedasticity of u is not relaxed in this specification.

2.3. The Single-index Model of Klein and Spady (1993)

The single-index specification of the binary response model is given by

$$P(y_n=1 \mid x_n) = F(x_n\beta \mid x_n) = G(x_n\beta) \tag{8}$$

where G is an unknown function (not necessarily a distribution function) whose range is contained in $[0,1]$. Root-N consistent, asymptotically normal estimators of β in single-index models have been developed by Ichimura (1993), Klein and Spady (1993) and Powell *et al.* (1989). The estimator of Klein and Spady achieves the asymptotic efficiency bound of Chamberlain (1986) and Cosslett (1987) if G is a continuous distribution function and certain other regularity conditions are satisfied. Klein and Spady assume that the model satisfies the index restriction $E(y \mid x) = E(y \mid x\beta)$.

The intercept component of β is subsumed in G and is not identified. Klein and Spady propose an estimator of β, b_{ks}, that is obtained by maximizing the (quasi) log-likelihood function (2), where $F(x_n\beta \mid x_n)$ is defined as $G_N(x_n\beta)$, a nonparametric estimate of $G(x_n\beta)$. Klein and Spady present the asymptotic theory of the quasi-maximum likelihood estimator and methods for estimating asymptotic standard errors. $N^{1/2}$ times the centered estimator of β is asymptotically normal. The index restriction permits multiplicative heteroscedasticity of a general but known form and heteroscedasticity of an unknown form if it depends only on the index (cf. Klein and Spady).

As proposed by Klein and Spady, G_N is calculated from nonparametric kernel estimates of the density of xb_{ks} conditional on y. Define $P_N = N^{-1}\Sigma_{n=1}^{N} y_n$, i.e. P_N is the sample proportion of women who participate in the labour market. Then for any real v

$$G_N(v) = \frac{P_N g_N(v \mid y=1)}{P_N g_N(v \mid y=1) + (1-P_N) g_N(v \mid y=0)} \tag{9}$$

where $g_N(\cdot \mid y)$ is a kernel estimate of $g(\cdot \mid y)$, the conditional density of $x\beta$. This estimate is given by

$$g_N(v \mid y=1) = (NP_N h_N)^{-1} \sum_{n=1}^{N} y_n K[(v - x_n b_{ks})/h_N] \tag{10}$$

and

$$g_N(v \mid y=0) = (N(1-P_N) h_N)^{-1} \sum_{n=1}^{N} (1-y_n) K[(v - x_n b_{ks})/h_N] \tag{11}$$

where K is the kernel function and $\{h_N\}$ is a sequence of bandwidths satisfying $Nh_N^6 \to \infty$ and $Nh_N^8 \to \infty 0$ as $N \to \infty$.

In establishing the asymptotic distributional properties of the estimator a trimming function is necessary that downweights observations near the boundary of the support of $x\beta$. However, because trimming appears to have little effect on the numerical results obtained in applications I do not consider trimming in this paper (see also Klein and Spady, 1993, p. 406).

I use the normal density function as the kernel function. In determining the smoothing parameter h_N I follow Newey *et al.* (1990) and use the method of generalized cross-validation (Craven and Wahba, 1979), taking into account the restrictions on the bandwidth. With this method, the smoothing parameter is chosen to minimize the residual sum of squares divided by the square of the 'degrees of freedom' of the residual, which is defined as the sample size minus the trace of the matrix (depending upon the regressors and smoothing parameters) that transforms the dependent variables into their nonparametrically fitted values. Generalized cross-validation appears to work reasonally well in this setting. The results do not vary substantially when bandwidths close to the optimal bandwidth are used. The bandwidth selection method used by Horowitz (1993) yields very similar values for the bandwidths.

The Klein–Spady estimator is computationally costly because of the nonparametric kernel estimation to be conducted at each iteration. However, it is well known that there are methods to speed up kernel estimation, e.g. the fast Fourier transform (FFT). In this paper I employ the FFT version of the Klein–Spady estimator. In Gerfin (1993b) I show that the FFT version can be more than two hundred times faster than the direct method and yields very accurate results.

2.4. Smoothed Maximum Score (Horowitz, 1992)

The maximum score model consists of equation (1) with the auxiliary assumptions that $\text{median}(u \mid x) = 0$ and that the distribution of u satisfies certain regularity conditions. In other respects the distribution of u is assumed to be unknown, i.e. u is allowed to have virtually arbitrary heteroscedasticity of unknown form. With respect to choice probabilities all we know is that $P(y=1 \mid x) = 0{\cdot}5$ if $x\beta = 0$ and that $P(y=1 \mid x) - 0{\cdot}5$ has the same sign as $x\beta$ if $x\beta \neq 0$. The maximum score estimator was proposed by Manski (1975, 1985). Horowitz (1992) developed a smoothed maximum score estimator that avoids the discontinuity of the objective function of the original maximum score estimator. Smoothed maximum score estimation consists of selecting the estimator b_{sms} to maximize

$$S_N(b_{sms}) = N^{-1} \sum_{n-1}^{N} [2 \cdot 1(y_n - 1) - 1]\Phi(x_n b_{sms}/h_N) \tag{12}$$

where $1(\cdot)$ denotes an indicator function, i.e. $1(\cdot) = 1$ if the expression in parentheses is true. The cumulative normal distribution function Φ is the smoothing function, and $\{h_N\}$ is a sequence of bandwidths that converge to zero at the rate $N^{-1/5}$. Horowitz (1992) gives the asymptotic theory of the estimator, methods for removing its asymptotic bias and for estimating asymptotic standard errors, and a plug-in method for selecting the bandwidth.[1] $N^{2/5}$ times the centred, bias-corrected estimator obtained from equation (12) is asymptotically normal.

The objective function S_N has many local maxima and requires a global optimization algorithm. I follow Horowitz (1993) and use the simulated annealing algorithm proposed by Szu and Hartley (1987). Simulated annealing yields a value of b_{sms} that is sufficiently near the global maximum of S_N to enable the maximum to be found by the Newton–Raphson algorithm.[2]

3. DATA

The models described in Section 2 are applied to the problem of the labour market participation of married women. This exercise is performed using cross-section data from two countries, Switzerland and Germany. The Swiss data consist of a sample of 873 married women drawn from the first representative health survey for Switzerland (SOMIPOPS) for 1981. The German data consist of a sample of 1564 married women drawn from the first wave of the German Socioeconomic Panel. Descriptive statistics are presented in the Appendix. The data are described in detail in Gerfin (1993a) and Wagenhals (1989). I use the same explanatory variables in both cases, namely *AGE* (age in years divided by 10), *AGESQ* (age squared divided by 1000), *EDUC* (years of formal education), *NYC* (Number of young children), *NOC* (number of older children), and *NLINC* (log of yearly non-labour income). In the case of Switzerland I also included a dummy variable *FOREIGN*, taking the value one if the woman is a permanent foreign resident. The age categories for children differ somewhat: the cutoff age is five in the case of Switzerland and six for the German data. Based on preliminary experiments I decided not to include interaction terms.

This specification represents a reduced form of the labour market participation equation because wages are not observed for non-participants. The estimation of the wage equation requires a selectivity correction term which is computed using the reduced-from participation equation. Therefore, in a first step the reduced form has to be examined carefully.

4. RESULTS

4.1. Estimation and Test Results

Table I displays estimation and test results for the probit model (with normalization $\sigma_u = 1$). For both data sets the estimated coefficients are as expected. In the Swiss case all explanatory

[1] The plug-in method provides an estimate of the asymptotically optimal bandwidth based on an initial bandwidth. The estimated bandwidth was in a small interval for a wide range of initial bandwidths, and all bandwidths within that interval yielded almost identical results.

[2] I experimented with the number of annealing iterations: 500 and 750 iterations yielded the same results for a variety of starting values.

variables except education have a significant effect on the participation probability. The strong positive effect of nationality on the participation probability is worth noting. In the German case all coefficients except the intercept are significant.

The lower half of Table I presents the parametric specification tests for the probit model. The diagnostic tests considered here are versions of Lagrange Multiplier tests suggested by White (1982) and are based on the theory of quasi-maximum likelihood. In the Monte Carlo results reported in Lechner (1991) the corresponding quasi-Lagrange Multiplier (QLM) version was found to have superior small sample properties in comparison with the standard LM tests. The specific formulas for the probit model are given in Blundell *et al.* (1993).

The first three QLM statistics concern the distributional assumption of the probit model. Normality is accepted for the Swiss data, but clearly rejected for the German data. The test statistics indicate that the distribution is asymmetric and leptokurtic in the German case.

The next QLM test statistics relate to heteroscedasticity. I test whether the elements of x are correlated with u individually as well as simultaneously. In the Swiss case homoscedasticity cannot be rejected for individual elements of x. However, homoscedasticity is rejected when all elements of x are considered simultaneously. For the German data homoscedasticity is rejected for the two children variables as well as for all elements of x simultaneously.

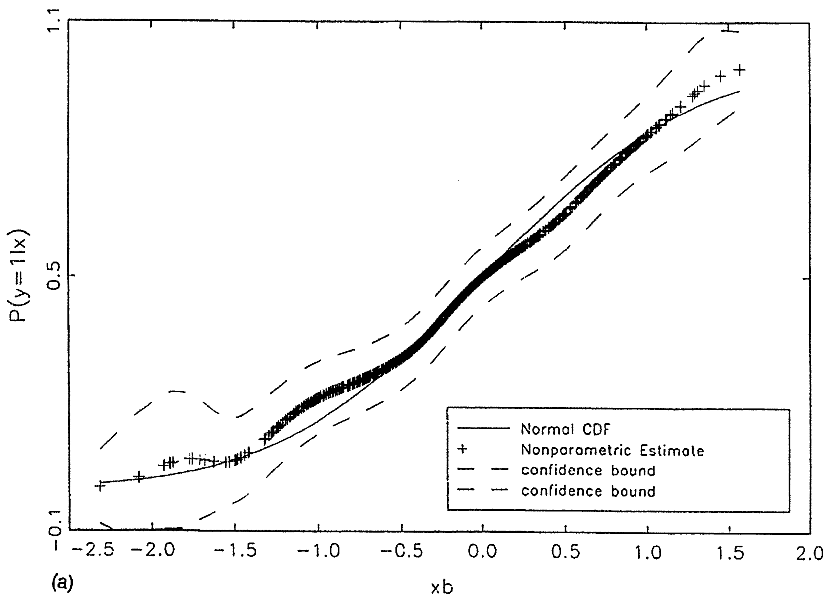
Table I. Probit estimation results (asymptotic standard errors in parentheses)

	Switzerland (N = 873)		Germany (N = 1564)	
Variable	Coefficient	Standard error	Coefficient	Standard error
Intercept	3·75	(1·41)	1·02	(1·13)
AGE	2·08	(0·41)	1·21	(0·35)
AGESQ	−0·29	(0·05)	−0·19	(0·04)
EDUC	0·02	(0·02)	0·08	(0·02)
NYC	−0·71	(0·10)	−0·72	(0·07)
NOC	−0·15	(0·05)	−0·29	(0·04)
NLINC	−0·67	(0·13)	−0·30	(0·10)
FOREIGN	0·71	(0·12)		
−Log Likelihood	509·4		968·1	

QLM specification tests	Test statistic	dof	Significance	Test statistic	dof	Significance
Skewness	1·52	1	21·76	6·51	1	1·08
Kurtosis	0·26	1	61·08	5·74	1	1·66
Normality	4·01	2	10·03	7·00	2	3·09
Heteroscedasticity						
AGE	3·70	1	5·43	0·00	1	97·48
AGESQ	3·11	1	7·74	0·00	1	97·40
EDUC	0·01	1	91·65	0·41	1	52·15
NYC	1·88	1	17·06	7·42	1	0·64
NOC	0·37	1	54·14	5·48	1	1·93
NLINC	0·04	1	83·76	0·14	1	70·63
FOREIGN	0·21	1	64·68	–		–
All	15·38	7	3·14	13·39	6	3·73

Note: Scale normalization achieved by setting s_u equal to 1.

I also computed the version of the information matrix test proposed by Orme (1988) which can be interpreted as an overall specification test. According to this test the probit cannot be rejected for the Swiss data (the test statistic has a significance level of 25%), but is severely rejected for the German data (a significance level of 0·5%). Furthermore, I computed the Chi-square specification test proposed by Andrews (1988). This test accepted the probit as well as



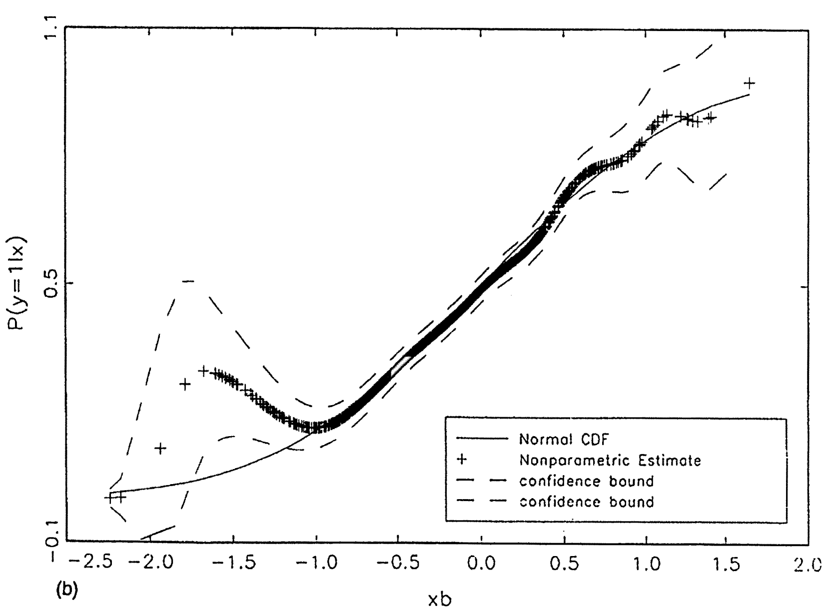


Figure 1. Nonparametric specification test, probit. (a) Switzerland; (b) Germany

the semi-nonparametric and the Klein–Spady models at significance levels between 65% and 75%, so the power of this test appears to be small in the present context.[3]

Summarizing, the specification tests indicate that the assumptions underlying the probit specification cannot be rejected for the Swiss data (with one exception), but for the German data they are rejected (cf. Lechner, 1991 who obtained very similar results for the logit model using the same data). This result is confirmed by the non-parametric specification test of Horowitz (1993). This test is based on a 95% uniform confidence band around the non-parametric regression of y on xb_{pr} (see Horowitz, 1993, p. 60). The probit specification cannot be rejected when the normal distribution function lies within the confidence band. Figure 1 shows the results of this test. In the case of the Swiss data the normal distribution function lies within the uniform confidence band. For the German data sets this is not the case. The bandwidths are chosen by generalized cross-validation.

Table II presents the estimation results for the semi-nonparametric model. For the Swiss data the distribution parameters α are not significantly different from zero. Consequently a likelihood ratio test does not reject the probit specification which is nested in the semi-nonparametric specification. Taking the different scaling into account the estimates of β are similar to the probit estimates.

In the German case the branching problem mentioned above leads to a different normalization than in the Swiss case. Only α_1 is estimated and turns out to be highly significant. A likelihood ratio test rejects the probit specification. I also estimated the semi-nonparametric model with $K=4$ and $K=5$. In the former case none of the estimated coefficients is significant and the

Table II. SNP estimation results (asymptotic standard errors in parentheses)

	Switzerland (N = 873)		Germany (N = 1564)	
Variable	Coefficient	Standard error	Coefficient	Standard error
Intercept	6·26	(2·50)	1·12	(0·99)
AGE	3·37	(0·85)	1·21	(0·34)
AGESQ	−0·49	(0·12)	−0·19	(0·04)
EDUC	0·02	(0·03)	0·07	(0·02)
NYC	−1·31	(0·27)	−0·77	(0·12)
NOC	−0·27	(0·10)	−0·30	(0·05)
NLINC	−1·04	(0·26)	−0·27	(0·09)
FOREIGN	1·02	(0·25)		
α_1	−0·21	(0·12)	−0·36	(0·08)
α_2	0·19	(0·11)	−0·21	—
α_3	0·05	(0·03)	0	—
Distribution				
Standard deviation	1·52		1·29	
Skewness	0·43		1·08	
Kurtosis	−0·62		0·50	
−Log Likelihood	506·8		961·2	

Note: The Swiss results are based on a different restriction on the distribution parameters α from the German results. K is set equal to 3 in both cases.

[3] The test is based on the two cells defined by the value of y.

likelihood ratio test does not reject the specification with $K = 3$. In the latter case the maximization algorithm did not converge. This is in line with Gabler *et al.* (1993), who found that very large sample sizes seem to be necessary to identify orders of K higher than 3. Score tests still indicate heteroscedasticity with respect to the variables *NYC* (significance level 3·12) and *NOC* (significance level 4·67), i.e. the flexible distribution does not solve this problem.

Tables III and IV present the estimation results for the Klein–Spady and the maximum score models as well as the rescaled estimates of the two previously discussed models. For the Swiss data (Table III) the estimated coefficients do not vary much across models when their standard errors are taken into account. Note that in the smoothed maximum score model only *AGE* and *AGESQ* are significant at the 5% level. But this should be interpreted with caution because

Table III. Estimation results, Switzerland (N = 873)

	Klein–Spady[a]		Smoothed maximum score[a]		Probit[a]		SNP[b]	
Variable	Coeff.	Std.err	Coeff.	Std.err	Coeff.	Std.err	Coeff.	Std.err
Intercept	—	—	5·83	(1·78)	5·62	(1·35)	5·99	(2·20)
AGE	2·98	(0·90)	2·84	(0·98)	3·11	(0·77)	3·23	(0·87)
AGESQ	−0·44	(0·12)	−0·40	(0·13)	0·44	(0·10)	−0·47	(0·12)
EDUC	0·02	(0·03)	0·03	(0·05)	0·03	(0·03)	0·02	(0·03)
NYC	−1·32	(0·33)	−0·80	(0·43)	−1·07	(0·26)	−1·26	(0·24)
NOC	−0·25	(0·11)	−0·16	(0·20)	−0·22	(0·09)	−0·26	(0·10)
NLINC	−1·0	—	−1·0	—	−1·0	—	−1·0	
FOREIGN	1·06	(0·32)	0·91	(0·57)	1·07	(0·29)	0·98	(0·26)
Bandwidth	0·40[c]		0·70					

[a] Results based on scale normalization $b_{NLINC} = -1$.
[b] Obtained by dividing the respective coefficients by the absolute value of the coefficient of NLINC. Standard errors computed by Delta method.
[c] Multiplied by the standard deviation of the index xb_{ks}.

Table IV. Estimation results, Germany (N = 1564)

	Klein–Spady[a]		Smoothed maximum score[a]		Probit[a]		SNP[b]	
Variable	Coeff.	Std.err	Coeff.	Std.err	Coeff.	Std.err	Coeff.	Std.err
Intercept	—	—	4·59	(2·25)	3·42	(3·10)	4·11	(3·29)
AGE	3·19	(1·28)	2·91	(1·34)	4·03	(1·72)	4·44	(1·11)
AGESQ	−0·53	(0·18)	−0·42	(0·16)	−0·64	(0·25)	−0·70	(0·13)
EDUC	0·25	(0·08)	0·15	(0·08)	0·28	(0·10)	0·26	(0·06)
NYC	−2·54	(0·77)	−1·32	(0·19)	−2·39	(0·90)	−2·82	(0·48)
NOC	−0·89	(0·28)	−0·64	(0·20)	−0·97	(0·39)	−1·09	(0·17)
NLINC	−1·0	—	−1·0	—	−1·0	—	−1·0	
Bandwidth	0·30[c]		0·60					

[a–c] As Table III

Monte Carlo simulations by Horowitz (1992) show that very large samples are needed to make the approximations of the asymptotic theory accurate. However, in contrast to the German data (see below) there is no obvious bias of the estimated standard errors. Overall, for the Swiss data the four models seem to yield very similar information on the individual behaviour that is governed by the index $x\beta$.

The coefficient estimates vary to some extent in the German case (Table IV). Apart from the intercept the estimates are largest in the semi-nonparametric model and smallest in the smoothed maximum score model (in absolute values). The difference is most pronounced for the coefficient of *NYC*. The differences between the estimates of the probit, the semi-nonparametric and the Klein–Spady estimators are not substantial. The smoothed maximum score standard errors seem to be biased, since they are smaller than the probit standard errors despite the slower convergence rate of the smoothed maximum score estimator. In this paper I do not correct for this by using bootstrap methods, as has been proposed by Horowitz.

Before turning to the predicted choice probabilities according to the different models I present some specification tests. It should be noted that these tests have an informal character because their asymptotic properties have only been derived for the fully parametric case. The semi-nonparametric and the Klein–Spady model are estimated within a quasi-maximum likelihood framework and can be tested with the information matrix test. However, due to numerical problems I was only able to compute the OPG version of the test. According to this test both models have to be rejected in both samples (the statistics have significance levels of 2·6 and 3·4 for the Swiss data and 3·1 and 2·9 for the German data in the Klein–Spady and the SNP model, respectively). This result is perhaps not too surprising given the known sensitivity of the OPG version of the IM test. Furthermore, because of the nonparametric estimation involved it is possible that the sample sizes are too small in the case of the Klein–Spady estimator.

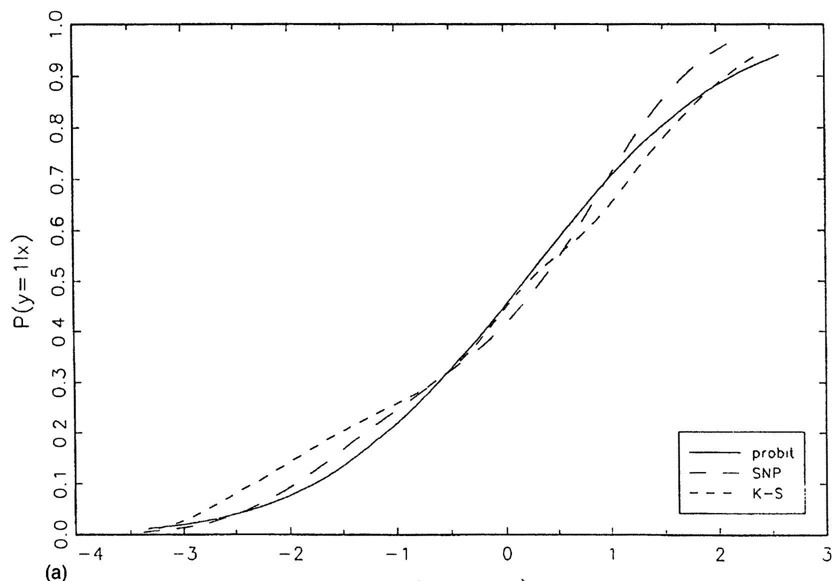
Another possibility to discrimate between the likelihood-based models is Vuong's (1989) test for non-nested models. This is a special form of the likelihood ratio test to test the hypothesis that two specified models are equally distant from the true model against the alternative that one model is closer to the true model. According to this test the Klein–Spady estimator is in both cases preferred over the probit and the semi-nonparametric estimators, but the test is significant at the 5% level for the comparison with probit in the German case only. Finally, the estimated smoothed maximum score models pass the specification test for this kind of model suggested in Horowitz (1993).

4.2. Predicted Choice Probabilities

It is difficult to evaluate the parameter estimates in terms of their effects on choice probabilities. For illustration purposes I plot the predicted choice probabilities for the probit, the semi-nonparametric, and the Klein–Spady estimators in Figure 2. For the Swiss data the three curves are close. The plots for the German data show that the predicted choice probabilities from the Klein–Spady estimator are not a monotonic function of xb for low values of xb, but for greater xb values it resembles closely the cumulative normal distribution. The choice probabilities according to the semi-nonparametric model lie between the two others in the problematic region, but almost coincide for larger values of xb. This problematic region is created by a number of women with low values of xb who are working.

It is difficult to compare the smoothed maximum score model with the other models because it does not accurately provide choice probabilities, when x is multidimensional. One way of comparison is to compute the probit and Klein–Spady choice probabilities at values of x where xb_{sms} equals 0. However, it turns out that this experiment is not informative. The computed choice probabilities are close to 0·5 at these points for all models in both samples. This is in contrast to Horowitz (1993), who found large differences in choice probabilities in this experiment. One possible explanation for this might be that in the samples analysed in this paper the sample proportion of observations with $y = 1$ is close to 0·5, whereas in Horowitz's sample this proportion is 0.84.

One possibility to examine the fit of the estimators is computing within sample predictions.



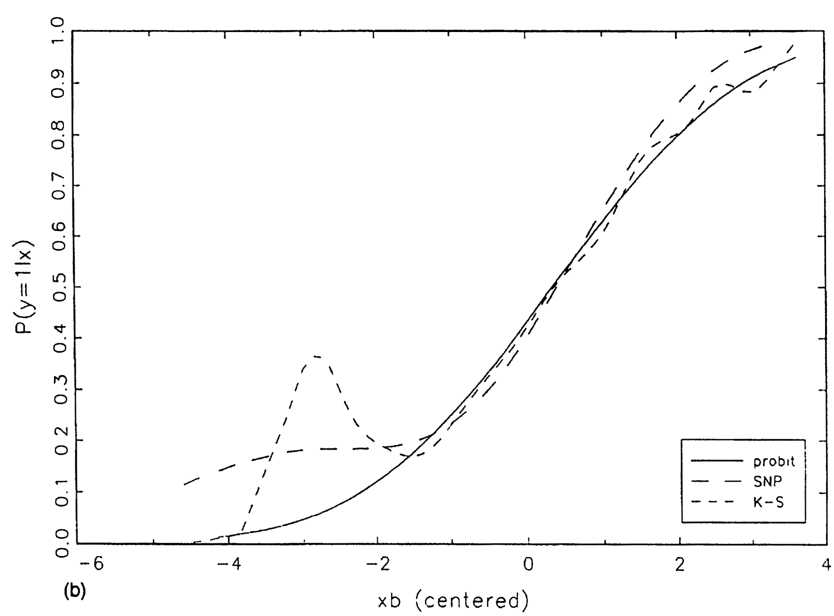


Figure 2. $P(y = 1 \mid x)$. (a) Switzerland; (b) Germany

Two prediction methods are considered. The first is based on the assumption that $y=1$ if $P(y=1 \mid x)$ is greater than 0·5, otherwise $y=0$. This yields the predicted number of participants for each model which can be compared to the actual number of participants. The second method simply sums the predicted choice probabilities, yielding the predicted sample proportion of working women. This second method cannot be applied with the smoothed maximum score estimator.

Table V presents the results of both prediction methods. In the first case the probit model performs best in predicting the number of participants. The semi-nonparametric and the smoothed maximum score estimators do not perform too well in this prediction. This is especially the case for the smoothed maximum score estimator with the German data. Since this is the only possible prediction with the smoothed maximum score model, it raises some doubts about the usefulness of this estimator in applied work. According to the second method the aggregate choice probabilities are very close to the sample proportion of participants for all considered models.

In a final prediction I only consider observations with $xb_{pr} < -1$ for the German data, i.e. observations in the problematic region. With the first method each of the models predicts zero participants. In the second method the Klein–Spady model performs very well in predicting the expected number of participants. This prediction reveals that the number of observations that cause the problems is surprisingly small (there are 80 observations, 17 of which participate).

The next question is to what extent simulations based on the different models will yield different results. I do not consider a full-scale simulation in this paper, but rather a demonstration of how the choice probability of a base case individual changes when her characteristics are changed. For space reason I only present one interesting example in which the number of children of this base case woman is altered. The other characteristics remain unchanged (they are given at the bottom of Table VI). This is the only case in which the different models yield different results. In all other cases I examined the models predict very

Table V. Within-sample prediction

	Swiss data (N = 873) number of participants	German data (N = 1564) number of participants	German data ($xb_{pb} < -1$) (N = 80) number of participants
Actual	401	686	17
1. Method			
$\Sigma 1(P_{pb} > 0{\cdot}5)$	389	575	0
$\Sigma 1(P_{snp} > 0{\cdot}5)$	338	540	0
$\Sigma 1(P_{ks} > 0{\cdot}5)$	382	560	0
$\Sigma 1(xb_{sms} > 0)$	355	423	0
2. Method			
$\Sigma(P_{pb})$	401·7	687·7	8·7
$\Sigma(P_{snp})$	401·0	685·8	14·8
$\Sigma(P_{ks})$	399·3	683·5	17·0

Notes: 1 indicates the indicator function. P_{pb} denotes the predicted choice probabilities according to the probit model. P_{snp} is defined analogously for the semi-nonparametric model and P_{ks} for the Klein–Spady estimator. xb_{sms} denotes the smoothed maximum score estimate of the index xb.

similar changes of the participation probabilities.[4] In the first four rows of Table VI the direction of the change in choice probabilities is the same, but the size of the changes differs to some extent (especially between the first and second rows). When the total number of children exceeds two the models predict different patterns of probability changes. The choice probability remains almost constant using the semi-nonparametric estimator, whereas the Klein–Spady estimator predicts an increase of the participation probability when the number of children exceeds two. Since the participation probability of the Klein–Spady model is a nonparametric estimate I also compute its standard error. These standard errors suggest that the estimate of the participation probability becomes more imprecise as the number of children increases because of the sparsity of observations.

The last column of Table VI presents the predicted choice probabilities according to a heteroscedastic probit (see next section). This model also predicts an increase in the participation probability when the number of children exceeds two. The heteroscedastic probit also picks up the peculiar feature of this data set that the participation rate of women with more than two children is relatively high.

4.3. Parametric Extensions

Another way to take account of detected misspecifications of the probit model is to extend it parametrically. Heteroscedasticity can be modelled explicitly by a parametric specification of the variance of u, $V(u)$, e.g $V(u)=\sigma_u^2[\exp(z\delta)]^2$. z is a vector of variables assumed to influence the variance of u and δ is the corresponding parameter vector. z could either be equal to x or a subset of x consisting of those variables that were detected by the score tests against heteroscedasticity. I estimated this kind of model for the German data. Estimation results are presented in the Appendix (Table A.II). z is defined to include the variables *NYC* and *NOC*, which were detected as the sources of heteroscedasticity by the score tests. The δ-coefficients are significant and the homoscedastic probit is rejected against this specification by a likelihood ratio test. The heteroscedastic probit model seems to describe the data much better than the simple probit (see Horowitz, 1993, for a similar result). This is confirmed by the nonparametric

Table VI. Participation probabilities, German data

	Participation probability			
	Probit	SNP	Klein–Spady (Standard error)	Heteroscedastic probit
NYC = 0 NOC = 0	0·73	0·78	0·75 (0·03)	0·79
NYC = 1 NOC = 0	0·46	0·41	0·41 (0·02)	0·36
NYC = 1 NOC = 1	0·35	0·29	0·31 (0·02)	0·31
NYC = 2 NOC = 0	0·21	0·20	0·18 (0·03)	0·20
NYC = 2 NOC = 1	0·13	0·19	0·23 (0·05)	0·21
NYC = 2 NOC = 2	0·08	0·18	0·39 (0·09)	0·23

Note: The base case woman has the following characteristics: AGE = 40, EDUC = 10, NLINC = 10. The children variables vary as given above. The standard error of the Klein–Spady participation probability is computed by the formula given in Härdle (1990, p. 100) for pointwise confidence intervals.

[4] A full set of results is available on request.

specification test (Figure A1 in the Appendix). This test does not reject the heteroscedastic probit. However, it is still rejected by the information matrix test (significance level 3·7%).

A possibility to correct for non-normality is to include polynomials and interactions of the original elements of x. This makes the index $x\beta$ a more flexible function of the explanatory variables. In the present case, however, it turns out that including third-order polynomials of age and education and interactions with the children variables does not solve the problems according to score tests. On the other hand, Klein (1993) gives an example where a probit model without polynomials is rejected, whereas the probit model including the polynomials is not rejected by the specification test proposed in this paper by Klein. This test appears to be more powerful in order to detect departures from normality than the usual score tests. An exploration of Klein's test is left for future research.

Overall, the results of this section suggest that parametric extensions of the probit are a promising alternative to semiparametric models.

4.4. Computational Issues

Computation time is an important issue in the present context. For the German sample (1564 observations, seven parameters) the SNP estimation took about 7 minutes on a 486 66 MHz PC. The FFT version of the Klein–Spady with 2048 Fourier points took about 5 minutes (only five parameters due to the necessary normalization and the excluded intercept). The final iteration using the direct method (i.e. no FFT), which seems to yield a better estimate of the covariance matrix, took about 6 minutes. The estimation of the smoothed maximum-score model (six parameters) took about 8 minutes. These numbers clearly show that at least in the present case computation time is not a restriction for the semiparametric models.

5. CONCLUSIONS

In this paper I compared a parametric and three semiparametric estimates of a binary choice model of labour market participation. The parametric model is the familiar probit specification. The semiparametric models are the semi-nonparametric estimator proposed by Gabler *et al.* (1993), the single-index estimator proposed by Klein and Spady (1993), and the smoothed maximum score estimator proposed by Horowitz (1992). This exercise was performed with two different cross-section data sets from Switzerland and Germany. Specification tests rejected the probit specification for the German data. For the Swiss data the probit model passed the specification tests.

The coefficient estimates do not differ substantially across models for both samples. The resulting predicted choice probabilities in the Swiss case are also similar. For the German data the differences are large for small values of the index $x\beta$, but for values larger than -1 the predicted choice probabilities almost coincide.

Several formal and informal tests gave mixed results. According to the Information matrix test the Klein–Spady and the semi-nonparametric models are rejected for both samples. The Vuong (1989) test for non-nested models indicated that the Klein–Spady model performed best compared to the probit and the semi-nonparametric models, but the test statistic was only significant for the comparison of probit and Klein–Spady with the German data. The smoothed maximum score model is accepted by a test designed for this model. In general, more work is necessary on the performance of existing specification tests within semiparametric models and on the development of specific tests for these models.

In order to examine the fit of the models I performed two different sets of within-sample predictions. The semiparametric models did not perform too well in the first method, which was based on the assumption that $y = 1$ if $P(y = 1 \mid x)$ is greater than 0·5. In the second prediction method I aggregated predicted choice probabilities to get an estimate of the sample proportion of participants. In this simulation all considered models performed equally well.

Finally, I examined the differences in predicted participation probabilities when the characteristics of a base case woman are altered. Only for the German data and only with respect to the number of children do the models predict substantially different probabilities. The probit model predicts continuously decreasing probabilities as the number of children increases, whereas the Klein–Spady model predicts probabilities that increase again when the total number of children exceeds two. The semi-nonparametric model predicts almost constant probabilities for these cases. Interestingly, a heteroscedastic probit also predicts increasing probabilities when the number of children exceeds two. So this feature of the data can also be captured by a modified probit model.

What recommendations for applied research follow from the results of this paper? First, testing parametric binary response models should become standard practice. If the parametric model is rejected the recommendations depend on the objective of the analysis. When the behavioural information contained in β is the only required result the smoothed maximum score estimator seems to be an attractive choice because it is the least restrictive estimator. When the estimation of the binary response model is the first step in the estimation of a censored model the Klein–Spady estimator seems to be a good solution. It is relatively easy to compute and has the required property of root-N consistency. On the other hand, when the simulation of sample expectations and changes thereof is the aim of the analysis it seems to be more promising to specify parametric extensions of the probit or possibly of the semi-nonparametric model. A systematic simulation analysis as well as an analysis of the effects of a misspecified binary response model in censoring models is left for future research.

APPENDIX

Table A.I. Descriptive statistics

	Switzerland (N = 873)		Germany (N = 1564)	
Variable	Mean	Standard deviation	Mean	Standard deviation
Y	0·46	0·51	0·44	0·48
AGE	4·00	1·05	4·14	0·98
AGESQ	1·71	0·87	1·81	0·83
EDUC	9·30	3·09	9·61	2·07
NYC	0·31	0·61	0·31	0·62
NOC	0·98	1·09	0·74	0·87
NLINC	10·67	0·55	6·30	0·35
FOREIGN	0·25	0·43	—	—

Notes: Y is the indicator of labour market participation. AGE is age in years divided by 10, AGESQ is age squared divided by 1000, EDUC is years of formal education, NYC is number of young children, NOC is number of older children, and NLINC is the log of yearly non-labour income. In the case of Switzerland I also included a dummy variable FOREIGN, taking the value one if the woman is a permanent foreign resident. The age categories for children differ somewhat: the cutoff age is five in the case of Switzerland and six for the German data.
Source for Swiss data: SOMIPOPS, the first representative health survey for Switzerland (1981)
Source for German data: German Socioeconomic Panel (GSOEP), first wave (1984)

Table A.II. Estimation results for heteroscedastic probit, German data

Variable	Coefficient	Standard error
Intercept	4·62	3·13
AGE	3·55	1·66
AGESQ	−0·61	0·25
EDUC	0·34	0·13
NYC	−4·02	1·61
NOC	−1·47	0·61
NLINC	−1·0	—
σ_0	3·09	1·14
σ_1 (NYC)	0·37	0·14
σ_2 (NOC)	0·29	0·11
−Log Likelihood	958·7	

Note: Scale normalization through setting the coefficient of NLINC equal to −1. The vector z includes the elements (NYC NOC).

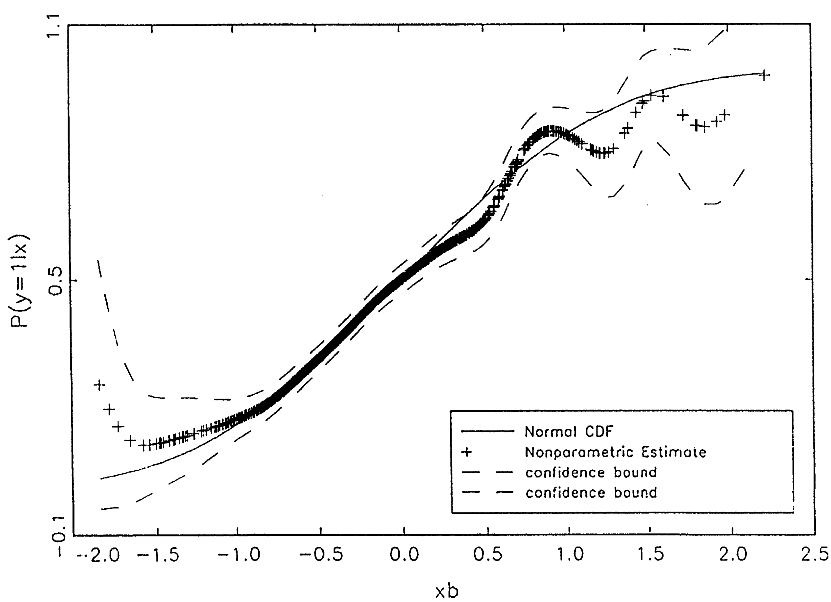


Figure A1. Nonparametric test, heteroscedasticity, probit, Germany

ACKNOWLEDGEMENTS

The first draft of this paper was written while I was a visitor at the Department of Economics of the University of Wisconsin-Madison and at the Institute for Research on Poverty. I want to thank both institutions for their generous hospitality. This stay was financed by the Swiss National Science Foundation. I am grateful to Michael Lechner and Joel Horowitz for providing me with the GAUSS codes for the semi-nonparametric and the smoothed maximum score estimators, respectively, and to Gerhard Wagenhals and the DIW for the German data. Three anonymous referees, Michael Lechner and participants at the ESEM 93 in Uppsala provided very helpful comments. All remaining errors are mine.

REFERENCES

Andrews, D. W. K. (1988), 'Chi-Square diagnostic tests for econometric models', *Journal of Econometrics*, **37**, 135–56.

Blundell, R. (ed.) (1987), 'Specification testing in limited and discrete dependent variable models', *Journal of Econometrics* (Annals 1987–1), **34**.

Blundell, R., F. Laisney and M. Lechner (1993), 'Alternative interpretations of hours information in an econometric model of labour supply', *Empirical Economics*, **18**, 393–415.

Chamberlain, G. (1986), 'Asymptotic efficiency in semiparametric models with censoring', *Journal of Econometrics*, **32**, 189–218.

Cosslett, S. R. (1987), 'Efficiency bounds for distribution free estimators of the binary choice and censored regression models', *Econometrica*, **55**, 559–85.

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.

Das, S. (1991), 'A semiparametric structural analysis of the idling of cement kilns', *Journal of Econometrics*, **50**, 235–56.

Gabler, S., F. Laisney and M. Lechner (1993), 'Semi-nonparametric maximum likelihood estimation of of binary choice models with an application to labour force participation', *Journal of Business and Economic Statistics*, **11**, 61–80.

Gallant, A. R. and D. N. Nychka (1987), 'Semi-nonparametric maximum likelihood estimation,' *Econometrica*, **55**, 363–390.

Gerfin, M. (1993a), 'A simultaneous discrete choice model of female labour supply and wages in Switzerland', *Empirical Economics*, **18**, 337–56.

Gerfin, M. (1993b), 'Computational aspects of the semiparametric binary response estimator of Klein and Spady', mimeo, University of Bern.

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

Horowitz, J. (1992), 'A smoothed Maximum Score estimator for the binary response model', *Econometrica*, **60**, 505–31.

Horowitz, J. (1993), 'Semiparametric estimation of a work-trip mode choice model', *Journal of Econometrics*, **58**, 49–70.

Ichimura, H. (1993), 'Semiparametric Least Squares (SLS) and weighted SLS estimation of Single-Index models', *Journal of Econometrics*, **58**, 71–120.

Klein, R. L (1993), 'Specification tests for binary choice models based on index quantities', *Journal of Econometrics*, **59**, 343–76.

Klein, R. L. and R. H. Spady (1993), 'An efficient semiparametric estimator for binary response models', *Econometrica*, **61**, 387–421.

Lechner, M. (1991), 'Testing Logit models in practice', *Empirical Economics*, **16**, 177–98.

Manski, C. F. (1975), 'Maximum Score estimation of the stochastic utility model of choice', *Journal of Econometrics*, **3**, 205–28.

Manski, C. F. (1985), 'Semiparametric analysis of discrete response: asymptotic properties of the Maximum Score estimator', *Journal of Econometrics*, **27**, 313–34.

Manski, C. F. and T. S. Thompson (1986), 'Operational properties of the Maximum Score estimator', *Journal of Econometrics*, **32**, 313–34.

Melenberg, B. and A. van Soest (1991), 'Parametric and semiparametric modelling of vacation expenditures', Center for Economic Research, University Tilburg, Working Paper 9144.

Melenberg, B. and A. van Soest (1993), 'Semiparametric estimation of the sample selection model', Center for Economic Research, Tilburg University, Working Paper 9334.

Newey, W. K, J. L. Powell and J. R. Walker (1990), 'Semiparametric estimation of selection models: some empirical results', *American Economic Review*: *Papers and Proceedings*, **80**, 324–28.

Orme, C. (1988), 'The calculation of the Information Matrix test for binary data models', *The Manchester School*, **56**, 370–76.

Powell, J. L., J. H. Stock and T. M. Stoker (1989), 'Semiparametric estimation of index coefficients', *Econometrica*, **57**, 474–523.

Ruud, P. (1983), 'Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of distribution in multinomial discrete choice models', *Econometrica*, **51**, 225–8.

Stone, C. J. (1980), 'Optimal rates of convergence of nonparametric estimators', *Annals of Statistics*, **8**, 1348–60.

Szu, H. and R. Hartley (1987), 'Fast simulated annealing', *Physics Letters A*, **122**, 157–62.

Vuong, Q. H. (1989), 'Likelihood ratio tests for model selection and non-nested hypotheses', *Econometrica* **57**, 307–33.

Wagenhals, G. (1989), 'Einkommensverteilung und Frauenerwerbstätigkeit', Discussion Paper 139, University of Heidelberg.

White, H. (1982), 'Maximum likelihood estimation of misspecified models', *Econometrica*, **50**, 1–25.