Parametric and semi-parametric modelling of vacation expenditures

Melenberg, B.; van Soest, A.H.O.

Publication date:
1991

Citation for published version (APA):
Melenberg, B., & van Soest, A. H. O. (1991). *Parametric and semi-parametric modelling of vacation expenditures*. (CentER Discussion Paper; Vol. 1991-44). Unknown Publisher.

entER
for
omic Research

Discussion paper

[redacted]

TILBURG UNIVERSITY

No. 9144

PARAMETRIC AND SEMI-PARAMETRIC MODELLING OF VACATION EXPENDITURES

by Bertrand Melenberg
and Arthur van Soest

August 1991

ISSN 0924-7815

Parametric and semi-parametric modelling of vacation expenditures

Bertrand Melenberg and Arthur van Soest*

Tilburg University, Dept. of Econometrics, P.O. Box 90153,
5000 LE Tilburg, The Netherlands

Abstract

We analyse several limited dependent variable models explaining the budget share that Dutch families spend on vacations. To take account of the substantial number of zero shares in our cross-section data, two types of models are used. The first is the single-equation censored regression model, of which the Tobit model is a special case. We estimate and test several parametric and semi-parametric extensions of the Tobit model. Although the parametric models appear to be misspecified, estimated income elasticities of vacation expenditures according to the Tobit model and its extensions do not differ substantially. Secondly, we consider two-equation models, in which the decision whether or not to go on vacation and the decision on the amount to spend are treated separately. The first decision is modelled as a binary choice equation, whereas the second is treated as a conditional regression equation. For both equations, we estimate, test and compare parametric and semi-parametric specifications. Simulations are used to investigate whether apparently misspecified specifications indeed lead to biased conclusions about income elasticities of vacation expenditures.

* We are grateful to Joel Horowitz, Arie Kapteyn, and Theo Nijman for stimulating discussions and helpful comments and to the Netherlands Central Bureau of Statistics (CBS) for providing the data. The views expressed in this paper are those of the authors and do not necessarily reflect the policies of the CBS. The first author is grateful to the Netherlands Organisation of Scientific Research (NWO) for financial support.

1. Introduction

The aim of this paper is to analyse total annual vacation expenditures of Dutch families. Using cross-section data, we study the impact of family income, household composition, and several other family characteristics on the decision which share of the family budget is spent on vacations.

A typical characteristic of expenditures on vacations compared to other expenditure categories such as food, clothing, etc., is that many families do not spend anything on vacations. In our sample, this is the case for 37% of all observations. Various limited dependent variable models can be used to account for these zero observations. See, for example, Maddala (1983). In particular, for the problem at hand, two types of models seem adequate: The one-equation censored regression model, of which the familiar Tobit model is the standard special case, and a two-equation model, which consists of a binary choice part explaining the participation decision,[1)] and a conditional regression equation explaining the expenditure level if it is positive.

Standard estimators for these types of models are based on maximum likelihood or a Heckman type two-step procedure. Consistency of these estimators generally requires a complete and correct specification of a parametric family of the error distribution. As a consequence, specification testing and model selection are more crucial than in the standard regression model. In this paper, we focus on a thorough treatment of the choice of the econometric model. Following the strategy which Horowitz (1991) uses for his analysis of the binary choice model, we study various parametric and semi-parametric models, comparing estimates and performing formal and graphic tests.

In section 2 we consider the single-equation censored regression model. Our starting point is the Tobit model, which is characterised by the assumptions that error terms are homoskedastic and normally distributed. Pseudo maximum likelihood estimates for the parameters of the censored regression model obtained by maximising the Tobit likelihood are generally inconsistent if either the homoskedasticity or the normality assumptions are violated. Appropriate tests for these assumptions are discussed by, for instance, Chesher and Irish (1987) and Newey (1985). In our case, both homoskedasticity and normality are rejected. Our next step is to account for parametric forms of heteroskedasticity and nonnormality, but neither of the extensions are accepted by chi-squared diagnostic tests. We then turn to semi-parametric estimation and present two estimators which allow for both heteroskedasticity and asymmetry: The censored least absolute deviations estimator (Powell, 1984) and an efficient two-step estimator given by Newey and Powell (1987). Parametric and semi-parametric estimates are compared by looking at estimated confidence bounds for resulting income elasticities.

Apart from many observations with zero expenditures on vacations, our data are also characterised by a lack of observations with small positive vacation expenditures. This is not captured in the usual (zero threshold) censored regression model. In Van Soest and Kooreman (1987), a model with unobserved random thresholds was used to take account of this phenomenon, similar to Nelson (1977). For such a model, no semi-parametric estimators are yet available. In this paper, we therefore use an alternative model which has the same flexibility as the Nelson model. In section 3, the vacation expenditures decision is modelled in two steps. We use a binary choice equation to model the participation decision. For the non-zero observations, we use a regression equation to explain the (positive) level.

We compare and test various parametric and semi-parametric specifications of the binary choice model. Using standard specification tests, we hardly find any evidence against a simple Probit model, contrary to, for instance, Horowitz (1991). The Probit model is accepted against parametric alternatives allowing for heteroskedasticity or nonnormality, and by nearly all chi-squared diagnostic tests. It is also accepted by the specification test against a nonparametric alternative suggested by Horowitz (1991).

Subsequently, we estimate the binary choice model using the estimator proposed by Klein and Spady (1989). We assume that the model can be written as a single index model: $P(y=1|X) = F(X'\alpha)$ and present estimates of α as well as (nonparametric) estimates of the unknown function F. The Probit model and the single index model estimates can thus be compared by comparing corresponding predicted probabilities. We find that the outcomes according to the Klein-Spady estimates are quite close to those obtained for the Probit model.

In case of the regression equation, which is used to explain the positive budget share spent on vacation, the hypotheses of normality and homoskedasticity are rejected. Imposing the weak moment restriction $E(\varepsilon|X) = 0$ ensures consistency but not efficiency of the OLS-estimator. We therefore also estimate the regression equation using the estimator proposed by Robinson (1987), which achieves the asymptotic efficiency bound corresponding to the condition $E(\varepsilon|X) = 0$. This specification of the regression equation is tested in several ways and cannot be rejected.

Parametric and semi-parametric specifications of the complete two-equation model are compared by calculating predicted budget shares and by performing some simulations. We do not find substantial differences between the various specifications.

In section 4, we briefly evaluate the results. Although the models in section 3 are more flexible than their section 2 counterparts, neither the parametric nor the semi-parametric models are nested. A formal test to compare the semi-parametric models in sections 2 and 3 is not available. Using less formal arguments, we motivate our preference for the two-equations model.

2. The Censored Regression Model

In this section we use a censored regression model to explain family expenditures on vacations as a function of total expenditures and family characteristics. The specification of the model is as follows.

$$y_i^* = X_i'\alpha + \varepsilon_i,\ y_i = \max(y_i^*, 0). \tag{2.1}$$

Here y_i is the budget share spent on vacations for family i (i=1,...,N), i.e., annual expenditures on vacations as a percentage of total annual family expenditures, and $X_i=(X_{1i},\ldots,X_{7i})'$ is a vector of covariates. In our case, we use $X_{1i}=1$, X_{2i} is the logarithm of total annual family expenditures, X_{3i} is the logarithm of family size, X_{4i} is an age indicator, X_{5i} indicates education, X_{6i} is the degree of urbanisation, and $X_{7i}=X_{2i}^2$.[2)] $\alpha=(\alpha_1,\ldots,\alpha_7)'$ is a vector of unknown parameters. Note that if $\alpha_7=0$, then the Engel curves correspond to the Almost Ideal Demand System (cf. Deaton and Muellbauer, 1980). ε_i denotes the error term. The models discussed in this section differ with respect to the assumptions on the conditional distribution of ε_i, given X_i. Throughout this paper, we assume that the (X_i,ε_i), i=1,...,N, are i.i.d.

We use a cross-section of N = 1815 families with at least two adults, taken from the Consumer Expenditure Survey drawn in 1981 by the Netherlands Central Bureau of Statistics. In appendix 1, we present details on the definitions of the variables involved and statistics for the whole sample and for the subsamples with zero and positive vacation expenditures. Basically, the same data were used by Van Soest and Kooreman (1987).

Parametric models

The most common assumptions about the distribution of error terms in (2.1) are

$$\varepsilon_i | X_i \sim N(0,\sigma^2). \tag{2.2}$$

(2.1) and (2.2) yield the well-known Tobit model. Error terms are normally distributed and independent of the regressors. Asymmetry and heteroskedasticity are not allowed for.

Maximum likelihood estimates for the Tobit model and corresponding standard errors are mentioned in the left panel of table 2.1. The estimates for α_2 and α_7 imply that the budget share spent on vacations is increasing as a function of total expenditures as long as total expenditures do not exceed Dfl 111,000, i.e., in the whole sample range, implying that vacation is a luxury. The estimated *ceteris paribus* difference between $X_i'\alpha$ for families with total expenditures Dfl 50,000 and Dfl 20,000 is 0.052. This number can be interpreted as the difference between budget shares spent on vacations for such families, assuming that ε_i and $X_{2i},\ldots,X_{6i}$ are such that, for both families, vacation expenditures are positive. Young families spend significantly more on vacations than older families. The education level of the family head has a significant positive impact on vacation expenditures. Families in big cities spend more than those living in the country. The impact of family size is insignificant.

The ML estimates under the Tobit assumptions will generally be inconsistent if the error terms in the censored regression model are heteroskedastic or non-normally distributed (see, e.g., Hurd (1979) and Goldberger (1983), respectively). These assumptions can be tested using, for example, the score or Lagrange Multiplier tests based on generalised residuals described by Chesher and Irish (1987). The test results are as follows.

<u>Heteroskedasticity</u>: realisation 137.0 with critical values $\chi^2_{27;0.05}$ = 40.1 and $\chi^2_{27;0.01}$ = 47.0. This leads to rejection of the null that errors are homoskedastic (under the maintained assumption of normality).

<u>Non-normality</u>: realisation 22.3 with critical values $\chi^2_{2;0.05}$ = 6.0 and $\chi^2_{2;0.01}$ = 9.2. This leads to rejection of the null of normally distributed errors (under the maintanied hypothesis of homoskedasticity).

It thus becomes clear that at least one of the two assumptions is violated.

Table 2.1. Estimation results parametric models

(standard errors in parentheses)

	I		II		III		IV	
α_1	-292.11	(94.83)	-515.21	(119.01)	-608.87	(125.64)	-492.87	(102.36)
α_2	50.43	(18.35)	92.37	(22.55)	110.18	(23.78)	88.49	(19.37)
α_3	-0.91	(0.56)	-0.38	(0.49)	-0.82	(0.38)	-0.39	(0.43)
α_4	0.22	(0.06)	0.14	(0.06)	0.14	(0.06)	0.14	(0.05)
α_5	0.27	(0.17)	0.34	(0.15)	0.29	(0.13)	0.32	(0.13)
α_6	0.56	(0.10)	0.39	(0.09)	0.43	(0.09)	0.39	(0.08)
α_7	-2.16	(0.88)	-4.12	(1.07)	-4.96	(1.14)	-3.95	(0.92)
σ	6.23	(0.10)						
β_1			18.46	(12.45)	1568.26	(269.99)	9.69	(14.02)
β_2			-2.86	(2.36)	-26.48	(4.69)	-1.21	(2.66)
β_3			-0.387	(0.065)	-8.23	(1.17)	-0.315	(0.078)
β_4			0.014	(0.006)	0.13	(0.04)	0.016	(0.008)
β_5			-0.047	(0.020)	-0.51	(0.14)	-0.035	(0.024)
β_6			0.044	(0.011)	0.29	(0.10)	0.045	(0.014)
β_7			0.124	(0.112)	0.11	(0.02)	0.044	(0.126)
γ_1							-0.061	(0.010)
loglik.	-4233.86		-4155.45		-4157.73		-4144.15	

Explanation

The censored regression model (equation (2.1)) and the regressors are described at the beginning of this section.

I: Tobit model (2.2)

II: Exponential heteroskedasticity and normality (2.3)

III: Random coefficients and normality (2.4)

IV: Exponential heteroskedasticity and nonnormality (2.5)

Explicit incorporation of parametric forms of heteroskedasticity is straightforward. Estimation results for the following two parametric forms are presented in table 2.1:

$$\varepsilon_i | X_i \sim N(0, \exp(X_i'\beta)^2), \tag{2.3}$$

$$\varepsilon_i | X_i \sim N(0, |X_i'\Sigma X_i|), \text{ where } \Sigma = \mathrm{Diag}(\beta_1, \ldots, \beta_7). \tag{2.4}$$

In (2.3) heteroskedasticity is modelled in a multiplicative way. This specification has the advantage that the variance is guaranteed to be positive for all values of $\beta=(\beta_1,\ldots,\beta_7)'$. Gabler et al. (1990) use (2.3) to test for the presence of heteroskedasticity in a binary choice model.

From an economic point of view, (2.4) is more attractive than (2.3), at least, if all components of $\beta=(\beta_1,\ldots,\beta_7)'$ are nonnegative. In this case, the latent model can be interpreted as a random coefficient model, with diagonal covariance matrix of the vector of coefficients. (2.4) is used by Horowitz (1991) to allow for heteroskedasticity in a binary choice model.

Estimation results for the model with exponential heteroskedasticity (2.3) and the random coefficients model (2.4) are mentioned in the second and third panel of table 2.1. The signs of the estimates for the α_i's are the same as in the Tobit model. Differences in magnitude and significance level with the Tobit model estimates do not seem too large either. The estimated difference between the budget share spent on vacations by a family with total expenditures Dfl 50,000 and Dfl 20,000 is 0.064 and 0.068, for the exponential heteroskedasticity and the random coefficients model, respectively.

In the model with exponential heteroskedasticity, four out of six of the heteroskedasticity slope parameters (the β_j's) are significantly different from 0 on a 5% level. In the random coefficients model, all heteroskedasticity parameters are significantly different from 0. Three of them are negative, contradicting the random coefficients interpretation. For two observations, the estimated value of $X_i'\Sigma X_i$ is negative and taking the absolute value in (2.4) becomes necessary.

According to the log-likelihoods mentioned at the bottom of table 4.1, the Tobit model is strongly rejected as a special case of either one of the two more general models. In terms of likelihoods, the exponential heteroskedasticity model does slightly better that the random coefficients model.

Both heteroskedasticity and nonnormality can be incorporated in a parametric framework, using, for example, the following family of distributions for $\varepsilon_i|X_i$:

$$P[\varepsilon_i<t|X_i,\beta] = F(t|X_i,\beta) = G(t/h(X_i,\beta)), \tag{2.5}$$

with

$$G(s) = \Phi(\gamma_0+s+\gamma_1 s^2+\gamma_2 s^3). \tag{2.6}$$

Here $h(X_i,\beta)$ is a twice differentiable nonnegative function. For the exponential case we have $h(X_i,\beta)=\exp(X_i'\beta)$. The distribution function G was proposed by Ruud (1984) as a family of probability distributions generalising the standard normal. Φ denotes the standard normal distribution function. If $\gamma=(\gamma_0,\gamma_1,\gamma_2)'=0$ then the conditional distribution of ε_i is normal with zero mean and standard deviation $h(X_i,\beta)$. In general, γ must be such that the probability density function corresponding to G can only take on nonnegative values. This density is given by

$$g(s) = (1+2\gamma_1 s+3\gamma_2 s^2)\varphi(\gamma_0+s+\gamma_1 s^2+\gamma_2 s^3), \tag{2.7}$$

where φ denotes the standard normal density function. Thus, $g(s)\geq 0$ for all $s\in\mathbb{R}$, is equivalent to

$$\gamma_2\geq \gamma_1^2/3. \tag{2.8}$$

An identifying restriction on the conditional distribution of ε_i (or $\varepsilon_i/h(X_i,\beta)$) is necessary, implying that γ_0 can be written as a function of γ_1 and γ_2. Two natural identifying restrictions can be imposed:

a. $F(0|X_i,\beta) = 1/2$ (zero conditional median): $\gamma_0=0$.

b. $E(\varepsilon|X_i,\beta) = 0$ (zero conditional mean): $\gamma_0= \psi(\gamma_1,\gamma_2)$, where ψ is some intricate function with the properties

$$\psi(0,0) = 0, \quad \frac{\partial\psi}{\partial\gamma_1}(0,0) = -1, \quad \frac{\partial\psi}{\partial\gamma_2}(0,0) = 0. \tag{2.9}$$

These conditions can easily be verified by substituting $\psi(\gamma_1,\gamma_2)$ for γ_0 and differentiating the equality $\int s\, g(s)\, ds = 0$ with respect to γ_1 and γ_2.

In the homoskedastic case, <u>a.</u> en <u>b.</u> are equivalent and necessary and sufficient to identify the constant term α_1 in (2.1). In case of heteroskedasticity, the equivalence no longer holds. The log-likelihood contribution (conditional on X_i) of an observation (y_i,X_i) is as follows.

$$\text{If } y_i=0\text{: } \log F(-X_i'\alpha|X_i,\beta)=\log G(s_i),\text{ with } s_i=-X_i'\alpha/h(X_i,\beta).$$

$$(2.10)$$

$$\text{If } y_i>0\text{: } \log f(y_i-X_i'\alpha|X_i,\beta)=-\log h(X_i'\beta)+\log g(s_i),\text{ with } s_i=(y_i-X_i'\alpha)/h(X_i,\beta).$$

The Chesher and Irish (1987) test for nonnormality in the Tobit model can be interpreted as a Lagrange multiplier test on $(\gamma_1,\gamma_2)'=(0,0)'$ for the homoskedastic case of (2.5)-(2.6). To obtain a test for nonnormality which remains appropriate in case of a parametric form of heteroskedasticity (e.g. exponential heteroskedasticity), it is possible to estimate the model (2.5)-(2.6) by maximum likelihood, without imposing homoskedasticity but imposing normality, and subsequently perform a Lagrange multiplier test on $(\gamma_1,\gamma_2)'=(0,0)'$. The test statistic can straightforwardly be derived and, as in Chesher and Irish (1987), be rewritten in terms of generalised standardised residuals

$$e_i^{(k)}= E_{\gamma=0}\{(\epsilon_i/h(X_i'\beta))^k|y_i,X_i,\beta\}-\mu(k),\text{ where } \mu(k)=\Gamma(k)/\{2^{k/2-1}\Gamma(k/2)\}.$$

The result depends on which restriction is imposed, zero conditional median (case <u>a</u>) or zero conditional mean (case <u>b</u>). In either case, the test statistic can be obtained as the explained sum of squares in a regression of a vector $(1,\ldots,1)'\in\mathbb{R}^N$ on the columns of an Nxm matrix, where N is the number of observations and m is 2 plus the number of free parameters in α and β. For the exponential heteroskedasticity case, the typical row entries in this matrix are

$$e_i^{(1)}\frac{X_{ji}}{\exp(X_i'\beta)},\qquad e_i^{(2)}X_{ji}\qquad (j=1.,,,.7),$$

either $-e_i^{(3)}+2e_i^{(1)}$ (case <u>a.</u>) or $-e_i^{(3)}+3e_i^{(1)}$ (case <u>b.</u>),

and $\quad -e_i^{(4)}+3e_i^{(2)}$,

where the unknown parameters are replaced by their ML estimates under the null (γ=0). Under the null, the test statistic asymptotically follows a χ^2_2 distribution.

In case of homoskedastic error terms (i.e., $\beta_j=0$ for j>1) the two variants a. and b. of the test statistic coincide and are identical to the standard test statistic in, for example, Chesher and Irish (1987) and Newey (1985). This test statistic can also be derived using White's information matrix comparison (cf. Chesher and Irish, 1987), starting from moment restrictions as in Newey (1985), or as a Lagrange multiplier test obtained by embedding the standard normal distribution in a Pearson family (cf. Bera et al., 1984).

In case of heteroskedastic error terms however, variants a. and b. are not identical, since the transformed covariates $X_{ji}/\exp(X_i'\hat{\beta})$ will generally not contain a constant term.

The empirical results of the test for the exponential heteroskedasticity case are given below. In order to take account of the inequality restriction (2.9), we also present the test results for one parameter restriction only. The conclusion is clear: normality, and in particular the restriction $\gamma_1=0$, are rejected at the usual significance levels. The difference in the outcomes of variants a. and b. is very small. Since the conditional distribution is symmetric if and only if $\gamma_1=0$, the results suggest that the error terms follow an asymmetric distribution. Note that for the third test $\gamma_1=0$ is a maintained hypothesis and the results for the second test strongly indicate that this assumption is violated, implying that the third test is not feasible.

Tests on normality assumptions in the exponential heteroskedasticity model

H_0	degrees of freedom	test statistic: test a.	test statistic: test b.	critical value: level 0.05	critical value: level 0.01
$\gamma_1=\gamma_2=0$	2	13.1	13.0	6.0	9.2
$\gamma_1=0$	1	12.3	12.2	3.8	6.6
$\gamma_2=0$	1	1.0	1.0	3.8	6.6

It is straightforward to estimate (2.5)-(2.6) allowing for exponential heteroskedasticity and without imposing normality, maximising the log-likelihood based on (2.10). Results are mentioned in the righthand panel of table 2.1. Restriction (2.8) was imposed and appeared to be binding.

According to the standard t-test, the restriction $\gamma_1=0$ is rejected at the usual significance levels. The same conclusion is obtained using a likelihood ratio test. Still, the estimates for both α and β and their estimated standard errors are quite close to those in the second panel of the table.

The implied *ceteris paribus* pattern of expenditures on vacations as a function of total expenditures is quite similar to the pattern found before: the budget share spent on vacations is maximal if total expenditures are Dfl 73,000. The difference between the predicted shares if total expenditures are Dfl 50,000 and Dfl 20,000 is 0.061.

To test whether the parametric models fit the data, without having in mind a specific parametric alternative, we performed several chi-squared goodness of fit tests developed by Andrews (1989). These tests are based upon classifying observations into cells and comparing estimated cell probabilities (conditional on the covariates) with empirical probabilities based upon the sample data. Andrews (1989) indicates how the cells can be chosen for the Tobit model and his strategy can straightforwardly be generalised to censored regression models with more general (parametric) specifications of the error distribution. We used a partitioning based upon the endogenous variable into five cells: $y_i=0$, and four cells with $y_i>0$, distinguished by the value of the transformed error term, where the transformation is such that the transformed error is standard normal. We also used products of these cells with a partitioning into four cells of the space of covariates, based upon the value of X_{2i} (total expenditures) only, yielding a partitioning into 20 cells.

In case of maximum likelihood estimation, the test statistics can easily be obtained as the explained sum of squares of a regression of a vector $(1,\dots,1)'\in\mathbb{R}^N$ on the vectors of scores and the vectors of differences between predicted and sample probabilities for each of the cells. Under the null hypothesis of no misspecification, the test statistics follow a χ^2 distribution, with 16 or 4 degrees of freedom, depending on whether 5 or 20 cells are used.

The two tests were performed for each of the four parametric specifications. In all cases, the null hypothesis of no misspecification was strongly rejected. The values of the test statistic for the 20 cells case varied from 84.1 (the normal exponential heteroskedasticity model) to 181.6 (the Tobit model). Allowing for heteroskedasticity reduces the chi-squared statistics and the exponential heteroskedasticity model does better than the random coefficients model in this respect. Allowing for the chosen parametric form of nonnormality in the exponential heteroskedasticity model does not reduce the values of the chi-squared statistics.

Semi-parametric models

In this subsection we consider semi-parametric estimation of the censored regression model. At present, a number of semi-parametric estimators for model (2.1) are available. The estimators vary with respect to the assumptions on the conditional distribution of ϵ_i given X_i necessary for consistency. For most of these estimators, these assumptions include independence between the error term and the covariates, i.e., the conditional distribution does not depend on X_i. Examples are the estimators given by Duncan (1986), Fernandez (1986), Horowitz (1986), and Ruud (1986). The empirical results of the previous subsection, however, strongly suggest the presence of heteroskedasticity, which the independence assumption does not allow for. Moreover, the significance of the nonnormality parameter γ_1 suggests that imposing symmetry of the error terms distribution should be avoided. This excludes the use of Powell's symmetrically trimmed least squares estimator (Powell, 1986).

The estimators we do apply are characterised by the weak identifying restriction that the conditonal distribution of ϵ_i has zero median:

$$\mathrm{Med}(\epsilon_i | X_i) = 0. \tag{2.11}$$

Two estimators thus are appropriate. The first one is Powell's Censored Least Absolute Deviations (CLAD-) estimator (Powell, 1984). The CLAD estimate for α solves the minimisation problem

$$\min_{\alpha} S_N(\alpha), \text{ where } S_N(\alpha) = \frac{1}{N} \sum_{i=1}^{N} |y_i - \max\{0, X_i'\alpha\}| \tag{2.12}$$

This estimator is consistent and asymptotically normal if condition (2.11) together with some mild regularity conditions on the distribution of the covariates are satisfied. Powell (1984) also derives a consistent estimator for its asymptotic covariance matrix. The CLAD estimator is not efficient: its asymptotic covariance matrix does not attain the asymptotic efficiency bound.[3)]

The second estimator is proposed by Newey and Powell (1987). They show that, under mild regularity conditions, the efficient score corresponding to (2.11) satisfies

$$s(Z_i) = 2f(0|X_i)\ I_{(0,\infty)}(-\alpha'X_i)\ \mathrm{sgn}(y_i-\alpha'X_i).X_i. \tag{2.13}$$

Here $Z_i=(y_i,X_i')'$, $f(.|X_i)$ denotes the conditional density function of ε_i given X_i with respect to the Lebesgue measure on $\mathbb{R}$, and $\mathrm{sgn}(\varepsilon_i)=1$ if $\varepsilon_i>0$, $\mathrm{sgn}(\varepsilon_i)=0$ if $\varepsilon_i=0$, and $\mathrm{sgn}(\varepsilon_i)=-1$ if $\varepsilon_i<0$.

Let

$N_1 = \mathrm{Int}(N/2)$, the largest integer $\leq N/2$, $N_2 = N - N_1$,

$I_1 = \{1,\dots,N_1\}$, $I_2 = \{N_1+1,\dots,N\}$,

$\tilde{\alpha}_1,\tilde{\alpha}_2$: CLAD-estimators (or other feasible preliminary estimators) based on subsamples I_1 and I_2, respectively,

$\hat{s}_1(Z),\hat{s}_2(Z)$: estimators of the efficient score, based on I_1 and I_2,

$V_1^* = [\Sigma_{i\in I_2}\hat{s}_1(Z_i)\hat{s}_1(Z_i)'/N_2]^{-1}$, $V_2^* = [\Sigma_{i\in I_1}\hat{s}_2(Z_i)\hat{s}_2(Z_i)'/N_1]^{-1}$.

Then the estimator $\hat{\alpha}$ proposed by Newey and Powell is given by

$$\hat{\alpha} = \tfrac{1}{2}(\tilde{\alpha}_1+\tilde{\alpha}_2) + \tfrac{1}{N}[V_2^*\ \Sigma_{i\in I_1}\hat{s}_2(Z_i) + V_1^*\ \Sigma_{i\in I_2}\hat{s}_1(Z_i)]. \tag{2.14}$$

This estimator is based on the efficient score. Newey and Powell prove that, under some regularity conditions, it is semi-parametrically efficient, and derive a consistent estimator for its asymptotic covariance matrix.

CLAD-estimates require minimisation of the non-differentiable expression in (2.12). For this purpose, we used the simplex algorithm introduced by Nelder and Mead (1965) and extended by O'Neill (1971). Estimation results are mentioned in table 2.2. Estimating the covariance matrix of the CLAD estimator involves the choice of smoothness parameters c_0 and γ, cf. Powell (1984). Following Powell (1984) we chose $\gamma=0.2$ and we tried various values of c_0. Reported smoothing parameters minimised the covariance matrix. Thus the resulting standard errors might be interpreted as (estimated) lower bounds for the true standard errors. In case of the

Newey-Powell estimator, the choice of the smoothness parameters not only affects the covariance matrix but also the estimates themselves. We tried different values of the smoothing parameters. The choice hardly affected the parameter estimates, but had some impact on the estimated covariance matrix. The reported smoothing parameters yield standard errors which are smaller than the corresponding CLAD-standard errors.

According to the point estimates, the budget share spent on vacations is maximal if total expenditures are Dfl 57,000 (CLAD) or 79,000 (Newey-Powell). The estimated difference between the shares for families with total expenditures Dfl 50,000 and Dfl 20,000 is 0.066 (CLAD) or 0.047 (Newey-Powell). These numbers do not seem to be too much out of line with those based upon the parametric estimates. This is also the case for the estimates of the slope parameters corresponding to family size, age, education, and degree of urbanisation.

Table 2.2. Estimation results semi-parametric model
(standard errors in parentheses)

	I		II	
α_1	-731.30	(180.65)	-350.72	(175.75)
α_2	133.99	(33.68)	62.92	(33.05)
α_3	-0.23	(0.90)	-1.03	(0.44)
α_4	0.14	(0.10)	0.13	(0.06)
α_5	0.14	(0.18)	0.01	(0.13)
α_6	0.49	(0.11)	0.31	(0.08)
α_7	-6.12	(1.57)	-2.79	(1.55)

Explanation
The model assumptions are given by (2.1) and (2.11). The regressors are described at the beginning of this section.
I: CLAD estimator with smoothness parameters c_0=0.73875 and γ=0.20.
II: Newey-Powell estimator with smoothness parameters c=0.7 and k=40.

The estimated standard errors for the CLAD-estimator are larger than those for the Newey-Powell estimator. Still, the estimated difference between the two covariance matrices fails to be positive definite. As a consequence, a Hausman type specification test could not be performed. This may be a consequence of the choice of smoothness parameters.

In order to compare the results of parametric and semi-parametric model estimates, we used the estimates to compute confidence intervals for elasticities of vacation expenditures with respect to total expenditures (TE). Results are given in table 2.3. Two elasticities were considered: the elasticity of $y_i^*TE_i$ with respect to TE_i for the average family (EL1), and the elasticity of the average value of $y_i^*TE_i$ with respect to TE_i with the same percentage for all families (EL2). The censoring was not taken into account, because computation of elasticities of y_iTE_i would require a complete specification of the error distribution, which is not provided by the semi-parametric model.

The elasticities are calculated 500 times, for 500 independent draws of the parameter values from the estimated asymptotic distribution of the estimator of the parameter vector. For each case, we present the mean and the median elasticity, and the 0.05 and 0.95 quantiles. The latter two can be interpreted as the bounds of a two-sided 90% confidence interval.

The difference between elasticities based on CLAD and Newey-Powell estimates are relatively large, although these estimators are consistent under the same assumptions. In particular, the elasticities according to Newey-Powell estimates are small compared to the others. This estimator also yields the smallest confidence intervals. All confidence intervals overlap and in each case the elasticity is significantly larger than one on a 5% level.

Table 2.3: Elasticities

	EL1				EL2			
	mean	median	Q05	Q95	mean	median	Q05	Q95
Tobit model	3.00	3.13	2.05	4.72	2.67	2.78	1.95	3.84
Expon. Heterosk.	2.95	3.10	2.05	4.64	3.75	4.11	2.40	7.03
Exp. Het. & Nonnorm.	2.82	2.89	2.14	3.98	3.52	3.68	2.15	5.47
CLAD	2.53	2.54	1.74	3.44	3.45	3.62	2.00	5.85
Newey-Powell	1.70	1.70	1.17	2.25	1.89	1.95	1.05	2.81

3. A Two Equations Model

In the previous section, a single latent variable equation was used to model both the choice between zero and a positive budget share, and the level of the budget share. In this section we consider a model in which these two decisions are separated:

$$P(a_i=1|X_i) = F(X_i'\alpha_a);\ P(a_i=0|X_i) = 1-F(X_i'\alpha_a), \tag{3.1}$$

$$y_i^* = X_i'\alpha_b + \varepsilon_{bi}, \tag{3.2}$$

$$y_i = 0 \text{ if } a_i<0;\ y_i = y_i^* \text{ if } a_i>0. \tag{3.3}$$

Here F is an unknown function (not necessarily a distribution function) with range contained in [0,1]. (3.1) is a single index binary choice equation for the decision whether or not to go on holiday. X_i only enters (3.1) through the 'single index' $X_i'\alpha_a$. In order to identify α_a, some normalisation must be imposed. If it is assumed that $F(z)=\Phi(z/\sigma_a)$, where Φ denotes the standard normal distribution function, (3.1) becomes the familiar Probit model.

The regression equation (3.2) explains the budget share spent on vacation, conditional upon the decision to go on holiday. We cannot think of economic arguments for excluding regressors from either (3.1) or (3.2). Therefore, in principle, (3.2) and (3.3) may include the same regressors, although in the final empirical specifications, the two sets of regressors will include different cross products. Instead of imposing exclusion restrictions on the regressors in (3.2), we make the following identifying assumption:

$$E(\varepsilon_{bi}|X_i,a_i=1) = 0 \tag{3.4}$$

This assumption makes it possible to estimate (3.2) separately from (3.1), using only those observations with $y_i>0$.[4)]

In the remainder of this section, we discuss and compare parametric and semi-parametric estimation results of (3.1) and (3.2) separately. Finally, the estimation results of both equations are used to perform some simulations.

The Binary Choice equation

The endogenous variable a_i in (3.1) represents the decision to go on holiday ($a_i=1$) or not ($a_i=0$). We first consider the Probit model, i.e.

$$F(z) = \Phi(z/\sigma_a), \tag{3.5}$$

for some $\sigma_a>0$. In the vector of covariates X_i we included $X_{1i},\dots,X_{6i}$, described at the beginning of section 2. On the basis of preliminary estimation results with squares and cross products of the regressors $X_{1i},\ldots,X_{6i}$, we also included the cross term $X_{4i}X_{5i}$, whereas there appeared to be no reason to include X_{2i}^2, which was included in the model in the previous section.

The maximum likelihood estimates of the Probit model are presented in the left panel of table 3.1. In order to be able to compare the Probit estimation results with the semi-parametric estimation results discussed below, we have normalised the coefficient α_{a2}, corresponding to log total expenditures, to be 1. According to the Probit estimates, the probability of going on vacation is an increasing function of family size and degree of urbanisation, although family size has an insignificant influence. The cross term between age class and education level implies that a *ceteris paribus* increase in age has a positive effect on the probability of going on holiday for small education levels and a negative effect for high levels of education. A similar result holds for a *ceteris paribus* change in the education level.

Since the Probit estimator for α_a may be inconsistent if the underlying distributional assumptions are incorrect, we performed various specification tests. Hardly any of these tests rejected the null hypothesis of no misspecification. For instance, neither normality nor homoskedasticity was rejected using the score tests given by Chesher and Irish (1987). In addition, most of the Andrew's chi-squared tests we performed, did not result in rejection of the Probit model. There was only one exception: Partitioning the support of the explanatory variable 'family size' into two classes, family size less than or equal to three, and family size larger than three, the Andrew's chi-square test statistic was 8.04, leading to rejection of the null hypothesis of no misspecification at the 5% level $(8.04>5.99=\chi^2_{2;0.05})$.

The problem with the chi-squared diagnostic tests is that the results may strongly depend on the arbitrary choice of the partitions. We, therefore, also tested the Probit specification using a test proposed by Horowitz (1991). This test consists of performing a non-parametric regression of a_i on $X_i'\hat{\alpha}_a/\hat{\sigma}_a$, and computing a uniform confidence band for the regression function. If the standard normal distribution function Φ lies within this band, the null-hypothesis that the Probit specification is correct is accepted. For technical details, see proposition 1 of Horowitz (1991). The outcome of the test is presented in Figure 3.1, which includes a plot of Φ, the uniform 95 % confidence band, and the values of the kernel regression[5)] of a_i on $X_i'\hat{\alpha}_a/\hat{\sigma}_a$, evaluated in the points $X_{ai}'\hat{\alpha}_a/\hat{\sigma}_a$,

Table 3.1 Estimation results of the binary choice model
(Standard errors between parentheses)

parameter	I		II	
α_{a1} (constant term)	-10.896	(0.227)	0.	
α_{a2} (log total expenditures)	1.		1.	
α_{a3} (log family size)	0.050	(0.083)	0.052	(0.067)
α_{a4} (age class)	0.060	(0.019)	0.093	(0.012)
α_{a5} (education level)	0.180	(0.063)	0.236	(0.043)
α_{a6} (degree of urbanisation)	0.045	(0.015)	0.051	(0.012)
α_{a7} $(X_{4i}X_{5i})$	-0.017	(0.007)	-0.023	(0.005)
σ_a^2	0.606	(0.104)		
log L	-1082.3		-1090.4	

Explanation:

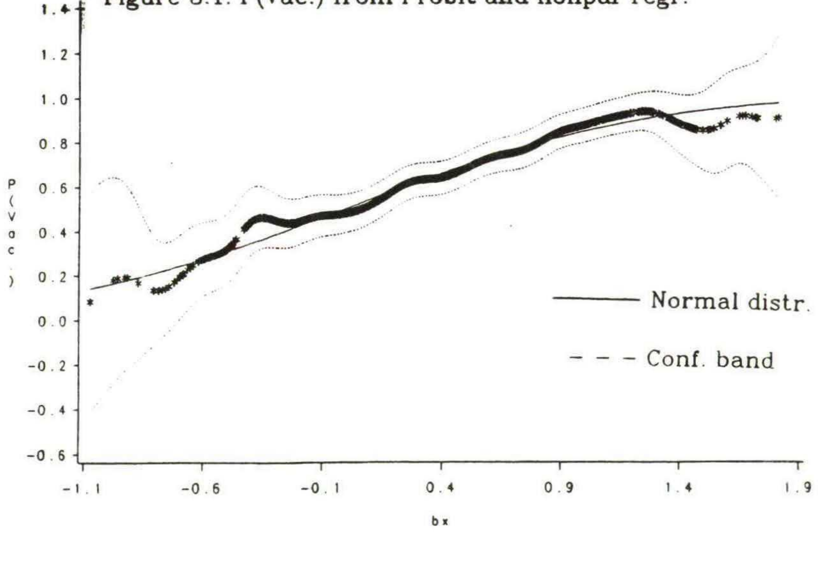
The first six regressors are the same as those in section 2; $X_{7i}=X_{4i}X_{5i}$.

I : ML estimates Probit model (3.1), (3.5) (normalisation: $\alpha_{a2}=1$);

II : Klein-Spady estimates single index model (3.1) (normalisation: $\alpha_{a1}=0$, $\alpha_{a2}=1$; smoothing parameter $h_N=0.2$);

log L: log-likelihood value in case of Probit; quasi-log-likelihood value (without trimming) in case of the Klein-Spady estimator.

Figure 3.1: P(Vac.) from Probit and nonpar regr.



$i=1,\ldots,N$. Since the normal cumulative distribution function lies entirely within the uniform 95 % confidence band, the Probit specification is accepted at the 5% level.

A semi-parametric estimator for the single-index model (3.1) is derived by Klein and Spady (1989). α_a can be estimated by maximising the (estimated) quasi-log-likelihood,[6] given by

$$Q(\alpha_a;F_N)=\frac{1}{N}\sum_{i=1}^{N}\{a_i\log[F_N(\alpha_a'X_i)]+(1-a_i)\log[1-F_N(\alpha_a'X_i)]\}, \tag{3.6}$$

where F_N represents a nonparametric estimate of F, specified as follows: Let P_N denote the sample frequency estimating the unconditional probability that $a_i=1$, i.e.,

$$P_N = N^{-1}\Sigma_{i=1}^{N} a_i. \tag{3.7}$$

Next, let $g_N(.|a_i)$ denote the density estimates for $X_i'\alpha_a$ conditional on a_i (for $a_i=1$ and $a_i=0$), given by

$$g_N(z|a_i=1) = \Sigma_{j=1}^{N} a_j\ K((z-X_j'\alpha_a)/h_N)/(h_N NP_N), \tag{3.8}$$

$$g_N(z|a_i=0) = \Sigma_{j=1}^{N}(1-a_j).K((z-X_j'\alpha_a)/h_N)/(h_N N(1-P_N)), \tag{3.9}$$

with K a Kernel function, and $\{h_N\}$ a sequence of bandwidths which have to satisfy $Nh_N^6 \to \infty$ and $Nh_N^8 \to 0$, for $N \to \infty$. Then the estimate F_N, evaluated at $z \in \mathbb{R}$, is given by

$$F_N(z) = P_N g_N(z|a_i=1)/[P_N g_N(z|a_i=1)+(1-P_N)g_N(z|a_i=0)]. \tag{3.10}$$

We made use of the kernel given by

$$K(z) = (3/2 - (1/2)z^2)\varphi(z), \tag{3.11}$$

with φ the standard normal density function.[7]

Klein and Spady show that, under some regularity conditions, the estimator $\hat{\alpha}_a$ of α_a obtained by maximising (3.6), satisfies

$$\sqrt{N}(\hat{\alpha}_a-\alpha_a) \to_d N(0,V_a), \tag{3.12}$$

where V_a is given by

$$V_a = [E(\partial P(\alpha_a)/\partial a_a)(\partial P(\alpha_a)/\partial a_a)'\{1/[P(\alpha_a)(1-P(\alpha_a))]\}]^{-1}, \qquad (3.13)$$

with $P(\alpha_a)$ a shorthand notation for $F(X_i'\alpha_a)$. The covariance matrix V_a can be estimated consistently by using its sample analogue.

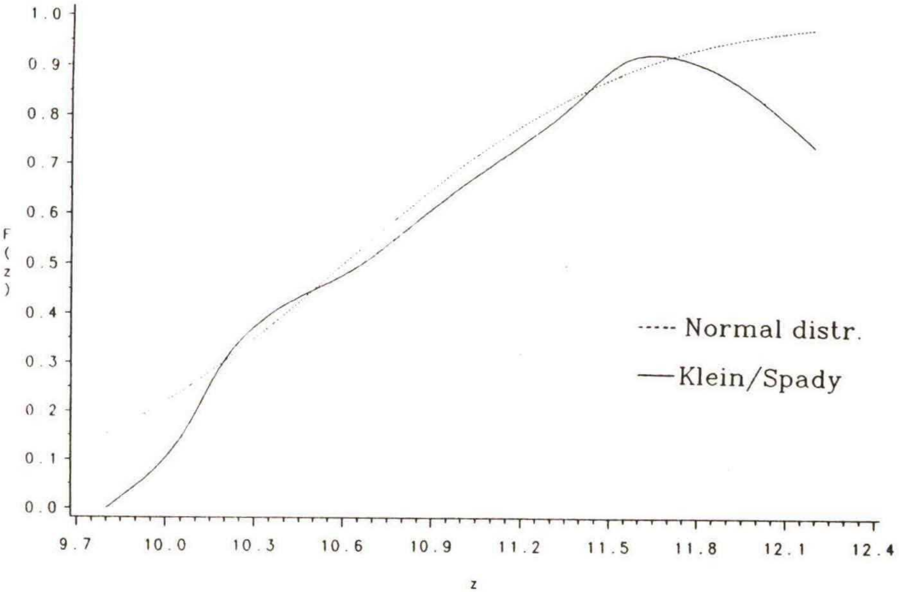
In order to guarantee identification of α_a in the single index model, one has to impose some normalisation condition. First of all, the coefficient corresponding to the constant term is set equal to zero, since the constant is absorbed in the function F. In addition, we have fixed the coefficient corresponding to the variable X_{2i}, i.e., log expenditures, to be equal to one. The choice of this normalisation is based upon the fact that log expenditures is the only explanatory variable which can reasonably be assumed to be continuously distributed.

The estimation results for the regression coefficients of the Klein-Spady estimator are presented in the right panel of Table 3.1. The function F_N, the estimate of F given by (3.10), is presented in Figure 3.2, together with the function $G(z) = \Phi((z-10.90)/0.77)$, the corresponding function in case of Probit. The results are obtained using the value 0.2 for the smoothing parameter h_N in (3.8) and (3.9)[8].

The Klein-Spady estimates of the regression coefficients are quite close to those obtained using Probit: The mean function is an increasing function of family size and the degree of urbanisation; it is increasing and decreasing as a function of age class for small and large values of the education level, respectively. According to figure 3.2, the Klein-Spady estimate F_N differs somewhat from the corresponding function in case of Probit. F_N is initially steeper and decreases if the value of the mean function becomes high. Relatively large differences between the two functions only occur in the region where observations are sparse. The conclusions from this figure are thus well in line with those from figure 3.1.

In order to get some more insight in the performance of the Klein-Spady estimation results, we applied an informal graphical test, suggested by Horowitz (1991). This test consists of assigning observations to cells, according to the predicted probability of going on vacation. This predicted probability is given by $F_N(X_i'\hat{\alpha}_a)$, with $\hat{\alpha}_a$ the Klein-Spady estimates of α_a and with F_N given by (3.10). We used the cells [0;0.3), [0.3;0.4), [0.4;0.5), ..., [0.8;0.9), [0.9;1.0]. The first cell is chosen larger than the others in order to obtain a comparable number of observations per cell.

Figure 3.2: Estimated distribution functions



For each cell, one can compute the quantities

$$Q_{pred} = \Sigma_{i \in cell} F_N(X_i'\hat{\alpha}_a), \tag{3.14}$$

$$Q_{obs} = \Sigma_{i \in cell} a_i/N_{cell}, \tag{3.15}$$

with N_{cell} the number of observations in the cell. If the single-index model is correctly specified, Q_{obs} and Q_{pred} will be close. The same test can also be applied to the Probit model.

Figures 3.3 and 3.4 represent graphs of Q_{obs} against Q_{pred} for the Probit and the Klein-Spady specification, respectively. In both cases, the outcomes are quite close to the 45°-line, suggesting that the model fits the data well.

In conclusion, formal and informal tests suggest that the Probit specification works quite well for the data at hand. Not surprisingly, the single index model, of which Probit is a special case, yields quite similar results and also works quite well.

Figure 3.3: Obs. and pred. probabilities for Probit model

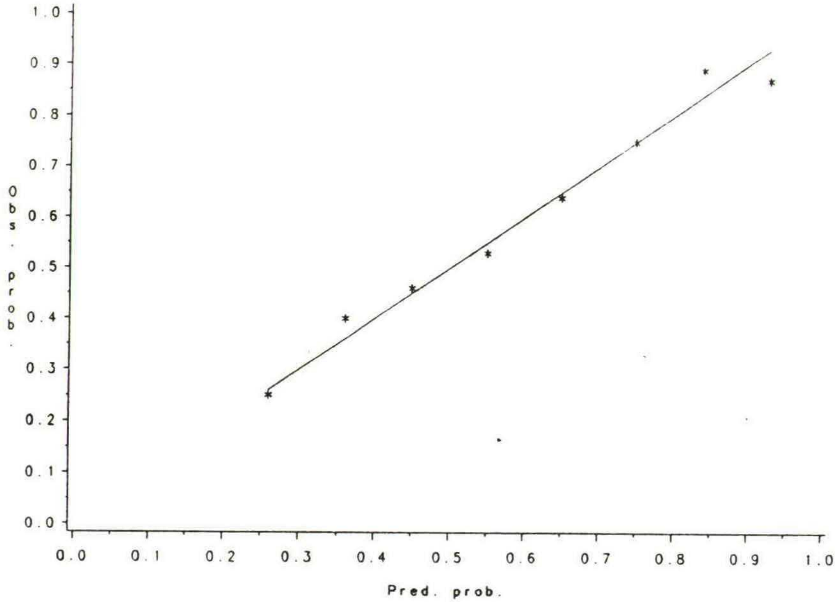
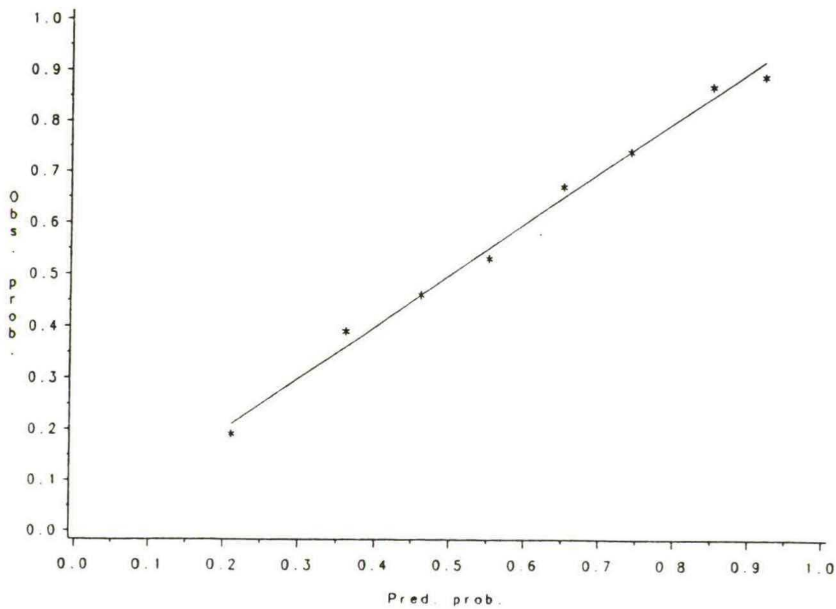


Figure 3.4: Obs. and pred. probabilities for Kl./Sp.-model



The Regression part of the model

Once the decision to go on holiday has been made, one has to decide how much to spend on it. We model this decision by means of the following conditional regression equation for those families with $y_i>0$:

$$E(y_i|a_i=1, X_i) = X_i'\alpha_b. \tag{3.16}$$

In terms of the error term $\varepsilon_{bi}=y_i-X_i'\alpha_b$, (3.16) can also be written as

$$y_i= X_i'\alpha_b+ \varepsilon_{bi};\ E(\varepsilon_{bi}|a_i=1, X_i) = 0. \tag{3.17}$$

As in section 2, y_i is the budget share of vacation expenditures (in %). In the vector of covariates X_i we included $X_{1i},\ldots,X_{6i}$, described in section 2. On the basis of preliminary estimation results with squares and cross products of the regressors $X_{1i},\ldots,X_{6i}$, we also included the cross product $X_{7i}=X_{2i}X_{6i}$.

The parameter α_b can simply be estimated by Ordinary Least Squares of y_i on X_i, using the subsample for which $y_i>0$ only. The estimation results are presented in the left panels of table 3.2. The first set of standard errors is computed in the standard way, assuming the error terms to be homoskedastic and independent of the covariates. The second set of standard errors is based upon an estimator of the covariance matrix of the OLS-estimator which, under weak regularity conditions, remains consistent in case of heteroskedasticity. This estimator will be discussed below (equation (3.27)). We also present the estimation results using the mean function without cross term. According to the OLS-estimation results without cross term, only the variable log family size has a negative influence on the budget share. The other variables have a positive influence, although the coefficients corresponding to log expenditures and education level are insignificant. The results with the cross term $X_{2i}X_{6i}$ included imply that the *ceteris paribus* effect of log expenditures is large and positive for people living in the country, but small or even negative for people living in cities.

The OLS estimator is asymptotically efficient if the errors are independent of the covariates and normally distributed. In order to test these distributional assumptions, we performed various specification tests. Using the same score test as in section 2, normality was strongly rejected.

To test against the alternative of heteroskedastic error terms, we used the same two parametric forms of heteroskedasticity as in section 2 and again applied similar score tests. In each case, the hypothesis of homoskedastic error terms was strongly rejected.

The outcomes of these tests suggest that either the normality or the homoskedasticity assumption or both do not hold. In this case, given (3.16), the OLS estimator is consistent, but not asymptotically efficient. A semi-parametric asymptotically efficient estimator for α_b is presented by Robinson (1987). It is given by

$$\hat{\alpha}_b = [\Sigma_{i=1}^{N} X_{bi} X_{bi}' \hat{\sigma}_i^{-2}]^{-1} [\Sigma_{i=1}^{N} X_{bi} y_i \hat{\sigma}_i^{-2}], \tag{3.18}$$

where now N=1143, the number of observations used in the regression, and $\hat{\sigma}_i^2$ is the so-called uniform k-NN estimator of σ_i^2, given by

$$\hat{\sigma}_i^2 = \Sigma_{j=1}^{N} \hat{\varepsilon}_{bj}^2 W_{ij}, \tag{3.19}$$

Here $\hat{\varepsilon}_{bi}$ is the OLS-residual, and where W_{ij} are weights, given by (if there are no ties):

$$\begin{aligned} W_{ij} &= k_N^{-1}, \text{ if } X_j \text{ is one of the } k_N \text{ nearest neighbors of } X_i, \\ W_{ij} &= 0, \quad \text{otherwise.} \end{aligned} \tag{3.20}$$

The distance between X_j and X_i is determined on the basis of the distance function ρ, given by

$$\rho(X_i, X_j) = \Sigma_{\ell=2}^{L} (X_{\ell i} - X_{\ell j})^2 / s_\ell, \tag{3.21}$$

where s_ℓ is the (univariate) sample variance of $X_{\ell i}$, which is included to correct for differences in measurement units of the $X_{\ell i}$'s, and where L denotes the number of covariates.

Table 3.2 Estimation results of the regression model

Parameter	I		II		III	IV	V
α_{b1} (constant)	-2.16		-22.00		-3.78	-18.67	-6.39
	(4.34)	[4.37]	(10.04)	[9.21]	(4.01)	(8.54)	(18.29)
α_{b2} (log total exp's)	0.67		2.57		0.87	2.28	1.06
	(0.44)	[0.44]	(0.97)	[0.09]	(0.40)	(0.83)	(1.56)
α_{b3} (log family size)	-1.42		-1.46		-1.60	-1.56	-1.47
	(0.42)	[0.40]	(0.42)	[0.39]	(0.38)	(0.37)	(0.39)
α_{b4} (age class)	0.16		0.15		0.15	0.15	0.16
	(0.05)	[0.05]	(0.05)	[0.05]	(0.05)	(0.04)	(0.05)
α_{b5} (education)	0.005		0.002		-0.007	0.02	0.03
	(0.13)	[0.12]	(0.13)	[0.12]	(0.11)	(0.11)	(0.15)
α_{b6} (deg. of urban.)	0.45		5.37		0.40	4.28	0.40
	(0.08)	[0.07]	(2.25)	[2.28]	(0.07)	(2.12)	(0.10)
α_{b7} ($X_{2i}X_{6i}$)	-		-0.47		-	-0.37	-
			(0.22)	[0.22]		(0.20)	
λ	-		-		-	-	0.48
							(2.39)

Explanation:
The model specification is given in (3.16). The first six regressors are the same as in section 2; $X_{7i}=X_{2i}X_{6i}$; λ is the coefficient corresponding to the inverse of Mill's ratio.
I, II: Ordinary Least Squares estimates;
(.): standard errors, computed in the standard way,
[.]: heteroskedasticity corrected standard errors, based upon (3.27).
III, IV: Robinson-estimates; smoothing parameter k_N=150,
(.): standard errors, computed using (3.24).
V : Robinson-estimates with the inverse of Mill's ratio included (without cross term),
(.): standard errors, computed using (3.24).

It is shown by Robinson that under some regularity conditions the limiting distribution of the estimator $\hat{\alpha}_b$ satisfies

$$\sqrt{N}(\hat{\alpha}_b - \alpha_b) \to_d N(0,V_b), \tag{3.22}$$

where V_b is given by

$$V_b = [E(X_i X_i' \sigma^{-2}(X_i))]^{-1}. \tag{3.23}$$

Here $\sigma^2(X_i) = E(\epsilon_{bi}^2 | X_i)$. V_b can be consistently estimated by

$$\hat{V}_b = [N^{-1}\Sigma_{i=1}^N X_i X_i' \hat{\sigma}_i^{-2}]^{-1}. \tag{3.24}$$

Estimation results using the Robinson estimator are mentioned in the third and fourth panel of table 3.2. The smoothing parameter k_N (c.f. equation (3.20)), is set equal to 150. Changing k_N hardly affected the outcomes. Because the coefficient of the cross term turned out to be insignificant on the 5% level, we also estimated the coefficients of the mean function without the cross term.[9] The resulting estimates are quite close to those obtained using OLS. The coefficient of the education level changes sign but remains quite insignificant. Estimated standard errors of the Robinson-estimator are smaller than (heteroskedasticity corrected) OLS standard errors. Surprisingly, however, the differences are quite small.

Comparing the asymptotically efficient Robinson estimates and the consistent OLS-estimates, a Hausman specification test for (3.16) can be performed. Under the assumptions given by Robinson (1987), it follows that the OLS-estimator of the regression coefficients, $\hat{\alpha}_{b,OLS}$, satisfies,

$$\surd N(\hat{\alpha}_{b,OLS}-\alpha) \rightarrow_d N(0,V_{b,OLS}), \tag{3.25}$$

with

$$V_{b,OLS} = [EX_iX_i']^{-1}[EX_iX_i'\sigma^2(X_i)][EX_iX_i']^{-1}. \tag{3.26}$$

Using the proofs of Robinson, it is easy to show that a consistent estimate of $V_{b,OLS}$ is given by

$$\hat{V}_{b,OLS} = [N^{-1}\Sigma_{i=1}^N X_iX_i']^{-1}[N^{-1}\Sigma_{i=1}^N X_iX_i'\hat{\sigma}_i^2][N^{-1}\Sigma_{i=1}^N X_iX_i']^{-1}, \tag{3.27}$$

where $\hat{\sigma}_i^2$ is the estimator of $\sigma^2(X_i)$ given in (3.19). The Hausman test statistic can thus easily be calculated. The outcome (in the case without cross term) is 7.1, which is less than $\chi^2_{6;0.05} = 12.6$. Thus, on the basis of the Hausman test we cannot reject specification (3.16).[10]

Finally, we tested the assumption that the error term ϵ_{bi} of the regression equation (3.17) is independent of the error term occurring in the latent equation underlying the binary choice equation, generalising the Heckman (1979) test for selectivity bias. If the function F in (3.1) is a distribution function, then (3.1) can be rewritten as

$$a_i^* = X_i'\alpha_a + \varepsilon_{ai},\ a_i=0 \text{ if } a_i^* \leq 0,\ a_i=1 \text{ if } a_i^*>0, \tag{3.28}$$

where the error term ε_{ai} has distribution function F. The regression model (3.17) can then be embedded in the following, more general, setting, which allows for dependence between ε_{ai} and ε_{bi},

$$E(\varepsilon_{bi}|X_i,\ \varepsilon_{ai}) = \lambda\varepsilon_{ai}. \tag{3.29}$$

Notice that (3.29) is the usual assumption imposed in the standard Heckman two-step procedure. Obviously, (3.29) implies

$$E(\varepsilon_{bi}|X_i,a_i=1) = \lambda E(\varepsilon_{ai}|\varepsilon_{ai}>-X_i'\alpha_a). \tag{3.30}$$

In case of Probit, the expectation on the righthand side is the inverse of Mill's ratio. Since Probit appears to describe the binary choice problem quite well, the hypothesis H_0: $\lambda=0$, which corresponds to model (3.17), can easily be tested by including the inverse of Mill's ratio (with α_a replaced by its Probit ML estimate) as an extra regressor in the regression equation. We estimated the resulting regression equation (without cross-term) using the Robinson-estimator. The estimates are presented in the righthand panel of Table 3.2. It follows that the null hypothesis $\lambda=0$ is not rejected.

Simulations

Following Van Soest and Kooreman (1987), we performed various simulations on the same sample of 1815 households used for estimation and testing. According to the two-equation model, the expected value of y_{bi}, the budget share spent on vacation expenditures, conditional on the covariates, is given by,

$$E(y_i|X_i) = E(y_i|a_i=1,\ X_i)\ P(a_i=1|X_i). \tag{3.30}$$

$P(a_i=1|X_i)$ can be estimated using the Probit estimates or the Klein-Spady estimates of α_a, together with the non-parametric estimate F_N drawn in figure 3.2. According to (3.17), $E(y_i|a_i=1,X_i)$ is equal to $X_i'\alpha_b$ and can be estimated by replacing α_b by its OLS or its Robinson estimate. We estimated $E(y_i|a_i=1,X_i)$ using the estimation results without cross term.

Table 3.3. Effects of a change of all total expenditures by the same factor

factor	I	II	III	IV	V	VI
0.90	0.583	0.586	1841.2	1830.4	1188.0	1185.6
0.99	0.625	0.627	2046.9	2041.4	1402.4	1399.6
1.00	0.630	0.631	2069.8	2065.0	1426.6	1423.9
1.01	0.634	0.635	2092.8	2088.7	1450.9	1448.2
1.10	0.671	0.670	2300.8	2302.7	1672.4	1669.9

Explanation

I : $P(y_i>0|\ X_i)$ according to Probit;
II : $P(y_i>0|\ X_i)$ according to Klein-Spady;
III: $E(y_iTE_i|\ y_i>0,\ X_i)$ according to OLS;
IV : $E(y_iTE_i|\ y_i>0,\ X_i)$ according to Robinson-estimator;
V : $E(y_iTE_i|\ X_i)$ according to Probit and OLS;
VI : $E(y_iTE_i|\ X_i)$ according to Klein-Spady and Robinson-estimator.

The first simulations concern changes of all family incomes by the same factor. Table 3.3 presents the consequences of different overall changes, ranging from -10% to +10%.[11] Comparing the various estimation results on the basis of the outcomes of these macro-economic simulation results, we hardly find any difference between Probit in combination with OLS and the Klein-Spady estimator in combination with the Robinson estimator. According to the last two columns in the table, the elasticity of aggregate vacation expenditures with respect to total expenditures is 1.70. This outcome is much smaller that the one in table 2.3. The main reason is that in table 2.3 the truncation was not taken into account.

In addition to overall changes in household incomes, we also consider the impact of a redistribution of incomes in the sample. Following Van Soest and Kooreman (1987), we used a redistribution of incomes such that the sample standard deviation of the logarithm of incomes decreases with 10 %, while the average income remains constant.[12] As a result, the ratio of the maximum and minimum income in the sample falls from 8.4 to 6.8. According to the Probit estimates, the participation probability rises from 0.630 to 0.637. Using the Klein-Spady estimator, we find a rise from 0.631 to 0.637. On the other hand, expected expenditures would slightly fall (with Dfl 4.6 or Dfl 5.1, according to the parametric and semi-parametric estimates, respectively). Thus, again, we hardly find any differences between parametric and the semi-parametric outcomes.

Thus, although, for example, figure 3.2 suggests that for at least some families, substantial differences between Probit and Klein-Spady estimates of participation probabilities exist, these are not reflected in the particular macro-economic simulations we consider.

4. Evaluation and conclusions

In recent years, the use of limited dependent variable models in empirical micro-econometrics has become quite popular. In the literature on structural labour supply models, for example, more and more complicated models are now being used, with various error terms reflecting different sources of random variation. Most of these models are estimated by maximum likelihood under the assumptions of normality and homoskedasticity. People have by now started to realise the consequences of the fact that this estimation procedure may yield inconsistent estimates if the rather strong assumptions are violated.

One implication of this is the need to test model assumptions thoroughly. A number of specification tests is available in a rather general framework. See, e.g., the studies in Blundell (1987). If the model then appears to be misspecified, the next step may be to relax the model assumptions and find estimators which remain consistent under these more general assumptions. Although various semi-parametric models and estimation techniques have been developed for this goal in the recent literature (cf. Robinson (1988) for a survey), applications are still sparse. The aim of this paper has been to analyse models explaining vacation expenditures, with emphasis on specification testing and comparing parametric and semi-parametric techniques.

We have considered two types of models. In section 2, we considered various specifications of the single-equation censored regression model. On the basis of specification tests, such as chi-squared diagnostics, all parametric specifications we considered are rejected. We estimated a semi-parametric specification allowing for heteroskedasticity, using two different estimators. One of the drawbacks of this approach is that the distribution function of the error terms cannot be estimated. Thus, we were not able to compute, for example, the income elasticity of vacation expenditures. This elasticity seems, from a practical point of view, much more interesting than the elasticity of the underlying latent variable which we did compute. A second drawback is that hardly any methods are available to test whether this semi-parametric model fits the data.

In section 3, we considered two-equation models, in which the decisions whether or not to go on holiday and how much to spend are modelled recursively. The participation decision was modelled by a binary choice model. The semi-parametric single index specification, and even the Probit model, a special case of the single index model, appeared to fit the data quite well. Predicted probabilities according to these models differ for sparse values of the covariates only.

Conditional on the decision to participate, the decision on how much to spend was modelled by a regression equation. This equation was estimated using both OLS and a semi-parametrically efficient estimator, with nearly identical results. The assumption of independence between the error in the regression equation and the error in the Probit model (written as a latent variable model) was tested and not rejected.

The two-equations model has the advantage that, even in case of the semi-parametric specification, it can be used to compute expected vacation expenditures for each family. It thus also allows for the computation of, for example, the elasticity of vacation expenditures with respect to total expenditures. The reason is that the probabilities in the binary choice model can be estimated satisfactorily. Problems would arise if we would try to estimate the distribution of the error term in the regression equation, but because of the linearity of this equation, this distribution is not needed. We therefore think that, at least in our case, the two-equations model is more appropriate than the single equation censored regression model.

If we repeat the simulations discussed in section 3 with some of the parametric models in section 2, we find similar income elasticities (cf. appendix 2). On the other hand, it appears that the models in section 3 capture the average value of vacation expenditures in the data much better than those in section 2. This seems one more reason to prefer the two-equations model.

Semi-parametric estimators are at this moment only available for some specific, relatively simple, univariate models. The recursive model in section 3 serves to illustrate that such simple models can be combined into a model which in some sense captures the complicated economic phenomena of interest.

Appendix 1: The Data

The data stem from the Consumer Expenditure Survey drawn in 1981 by the Netherlands Central Bureau of Statistics. Sample statistics for the exogenous variables are given in table A.1.

Table A.1. Sample statistics

	All observations		Observations with zero vacation expenditures		Observations with non-zero vacation expenditures	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
X_{2i}	10.37	0.36	10.22	0.34	10.46	0.35
X_{3i}	1.10	0.37	1.05	0.36	1.13	0.37
X_{4i}	7.09	3.18	7.32	3.44	6.96	3.01
X_{5i}	2.31	1.06	2.06	0.96	2.45	1.09
X_{6i}	3.86	1.74	3.76	1.77	3.91	1.72

Explanation:

X_{2i}: logarithm of total family expenditures in 1981 (in Dfl)

X_{3i}: logarithm of family size ($2 \leq$ family size ≤ 7)

X_{4i}: age class family head; $X_{4i}=1$: < 20 years old; $X_{4i}=2$: 20-24 years old; $X_{4i}=3$: 25-29 years old; ...; $X_{4i}=13$: > 74 years old

X_{5i}: education level family head, ranging from 1 (low) to 5 (high)

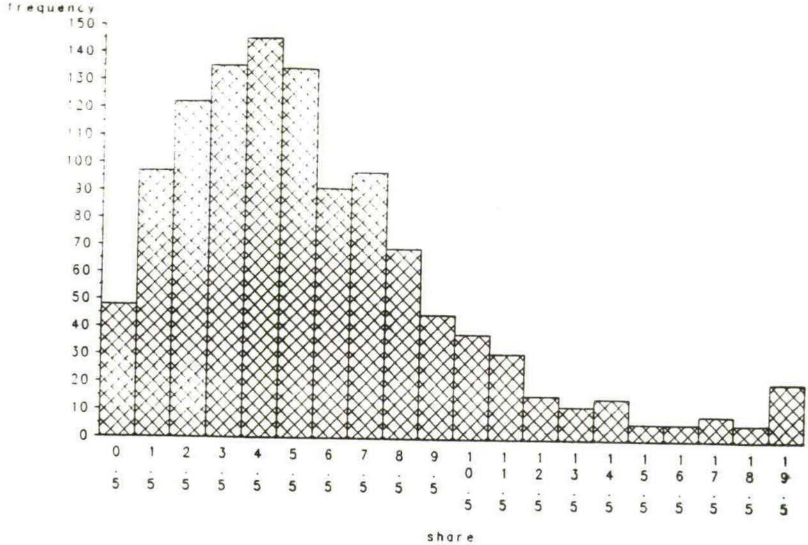
X_{6i}: degree of urbanisation, ranging from 1 (country village) to 6 (big city)

Vacation expenditures are defined as expenditures of any member of the family on a vacation, which is defined as a 'stay away from home for recreation purposes for at least four successive nights'. The average annual amount spent on vacations per family is Dfl 1415.4, zeroes included, and Dfl 2247.5 if zero expenditures are not included.

The distribution of positive budget shares of vacation expenditures is presented in Figure A.1.

Appendix 2: Table 3.3 for two censored regression specifications

The corresponding outcomes of table 3.3 for the Tobit specification and the censored regression specification assuming normality and exponential heteroskedasticity are represented in table A.2.



zero shares excluded; 1143 observations

Figure A.1: Distribution of budget shares spent on vacations (in %)

Table A.2 Effects of a change of all total expenditures by the same factor

factor	I	II	III	IV	V	VI
0.90	0.562	0.590	1764.3	1922.2	1100.8	1244.6
0.99	0.595	0.627	2013.2	2056.2	1311.3	1468.6
1.00	0.599	0.631	2041.1	2182.3	1335.1	1493.6
1.01	0.602	0.635	2069.1	2208.5	1359.0	1518.7
1.10	0.629	0.666	2322.5	2444.2	1576.3	1744.7

Explanation

I : $P(y_i>0|X_i)$ according to Tobit;

II : $P(y_i>0|X_i)$ according to censored regression model with exponential heteroskedasticity and normality;

III: $E(y_i TE_i|y_i>0,X_i)$ according to Tobit;

IV : $E(y_i TE_i|y_i>0,X_i)$ according to censored regression model with exponential heteroskedasticity and normality;

V : $E(y_i TE_i|X_i)$ according to Tobit;

VI : $E(y_i TE_i|X_i)$ according to censored regression model with exponential heteroskedasticity and normality.

References.

Andrews, D. (1989), "Chi-square diagnostic tests for econometric models", Journal of Econometrics, 37, pp. 135-156.

Bera, A.K., C.M. Jarque, and C.F. Lee, (1984), "Testing the normality assumption in limited dependent variable models", International Economic Review, 25, pp. 563-578.

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

Chamberlain, G. (1986), "Asymptotic efficiency in semi-parametric models with censoring", Journal of Econometrics, 32, pp. 189-218.

Chesher, A., and M. Irish, (1987), "Residual analysis in the grouped and censored normal linear model", Journal of Econometrics, 34, pp. 33-61.

Cosslett, S.R. (1987), "Efficiency bounds for distribution-free estimators of the binary choice and censored regression models", Econometrica, 55, pp. 559-585.

Deaton, A., and J. Muellbauer, (1980), "Economics and consumer behavior", New York, Cambridge University Press.

Duncan, G.M. (1986), "A semi-parametric censored regression estimator", Journal of Econometrics, 32, pp. 5-34.

Fernandez, L. (1986), "Non-parametric maximum likelihood estimation of censored regression models", Journal of Econometrics, 32, pp. 35-57.

Gabler, S., F. Laisney, and M. Lechner (1990), "Semi-nonparametric maximum likelihood estimation of binary choice models with an application to labour force participation", working paper, University of Mannheim.

Goldberger, S. (1983), "Abnormal selection bias", in: S. Karlin, T. Amemiya, and L. Goodman, eds., "Studies in econometrics, time series, and multivariate statistics", New York, Academic Press.

Heckman, J.J. (1979), "Sample selection bias as a specification error", Econometrica 47, pp. 153-161.

Horowitz, J.L. (1986), "A distribution-free least squares estimator for censored linear regression models", Journal of Econometrics, 32, pp 59-84.

Horowitz, J.L. (1990), "A smoothed maximum score estimator for the binary response model", Manuscript, Department of Economics, University of Iowa.

Horowitz, J.L. (1991), "Semiparametric estimation of a work-trip mode choice model", Manuscript, Department of Economics, University of Iowa.

Hurd, M. (1979), "Estimation in truncated samples when there is heteroskedasticity", Journal of Econometrics, 11, pp. 247-258.

Klein, R.L., and R.H. Spady, (1989), "An efficient semiparametric estimator for discrete choice models", Manuscript, Bell Communications Research Corp.

Maddala, G. (1983), "Limited dependent and qualitative variables in econometrics", Cambridge: Cambridge University Press.

Moon, C.G. (1989), "A Monte Carlo comparison of semiparametric Tobit estimators", Journal of Applied Econometrics, 4, pp. 361-382.

Nelder, J.A., and R. Mead, (1965), "A simplex method for function minimization", Computer Journal, 7, pp. 308-313.

Nelson, F. (1977), "Censored regression models with unobserved, stochastic censoring thresholds", Journal of Econometrics, 6, pp. 309-327.

Newey, W.K. (1985), "Maximum likelihood specification testing and conditional moment tests", Econometrica, 53, pp. 1047-1069.

Newey, W.K., and J.L. Powell, (1987), "Efficient estimation of linear and type I censored regression models under conditional quantile restrictions", Research memorandum No. 331, Princeton University.

O'Neill, R. (1971), "Algorithm AS 47: Function minimization using a Simplex procedure", Applied Statistics, 20, pp. 337-345.

Powell, J.L. (1984), "Least absolute deviations estimation for the censored regression model", Journal of Econometrics, 25, pp. 303-325.

Powell, J.L. (1986), "Symmetrically trimmed least squares estimation for Tobit models", Econometrica, 54, pp. 1435-1460.

Robinson, P.M. (1987), "Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form", Econometrica, 55, pp. 875-891.

Robinson, P.M. (1988), "Semiparametric econometrics: a survey", Journal of Applied Econometrics, 3, pp. 35-51.

Ruud, P.A. (1984), "Tests of specification in econometrics", Econometric Reviews, 3, pp. 211-242.

Ruud, P.A. (1986), "Consistent estimation of limited dependent variables models despite misspecification of distribution", Journal of Econometrics, 32, pp. 157-187.

Van Soest, A., and P. Kooreman, (1987), "A micro-econometric analysis of vacation behaviour", Journal of Applied Econometrics, 2, pp. 215-226.

Notes

1) 'To go on holiday', 'to participate' and 'positive budget share of vacation expenditures' are used as synonyms. Vacation expenditures include expenditures of all members of the family. Thus a family is said to participate or to go on holiday if at least one member of the family goes on holiday.

2) Preliminary estimation results with squares and cross products of the regressors suggested that only X_{2i}^2 should be incorporated.

3) Moon (1989) presents a Monte-Carlo comparison of the finite sample performance of the CLAD-estimator and several other semi-parametric estimators.

4) A test for (3.4) which generalises the usual test for selection bias in this type of two equation models (cf. Heckman, 1979), will be discussed below.

5) Following Horowitz we used the normal density. The smoothing parameter δ was set equal to 4/3, and the smoothing parameter h_N was chosen such that ω_N, the kernel-bandwidth, was equal to the standard deviation of $X'_{ai}\hat{\alpha}_a/\hat{\sigma}$. Other choices of h_N yield slightly different curves, but the conclusion that the Probit specification is accepted, remains.

6) Klein and Spady (1989) suggest that the finite sample performance of their estimator can be improved by multiplying the quasi-log-likelihood contributions with a trimming function τ_{Ni}, which is used to downweight observations with $X'_i\hat{\alpha}_a$ near the support boundary, where density estimates may not be reliable. The parameter estimation results we obtained using such trimming functions were virtually identical to those presented in table 3.1. Only the value of the quasi-log-likelihood increased, as might be expected.

7) An alternative kernel, which we used for comparison, can be found in Horowitz (1991).

8) Almost identical results were obtained using the kernel used by Horowitz (1991) (with correspondig smoothness parameter h_N=0.1). Other values of the smoothing parameters resulted in slightly different outcomes. The (informal) graphical test to be presented later suggests that the present choice fits the data reasonably well.

9) In addition, we tried other cross products and squares of $X_{2i},\dots,X_{6i}$, using the Robinson-estimator. None of these turned out to be significant on the 5% level.

10) The Hausman test may not be very powerful, since, for example, it is not easy to think of alternative specifications under which the Robinson estimator is not consistent while OLS remains consistent.

11) For comparison, we present in appendix 2 the corresponding results for the Tobit model as well as the censored regression model with normality and exponential heteroskedasticity.

12) For details we refer to Van Soest and Kooreman (1987), footnote p. 224.

Discussion Paper Series, CentER, Tilburg University, The Netherlands:

(For previous papers please consult previous discussion papers.)

No.	Author(s)	Title
9025	K. Kamiya and D. Talman	Variable Dimension Simplicial Algorithm for Balanced Games
9026	P. Skott	Efficiency Wages, Mark-Up Pricing and Effective Demand
9027	C. Dang and D. Talman	The D_1-Triangulation in Simplicial Variable Dimension Algorithms for Computing Solutions of Nonlinear Equations
9028	J. Bai, A.J. Jakeman and M. McAleer	Discrimination Between Nested Two- and Three-Parameter Distributions: An Application to Models of Air Pollution
9029	Th. van de Klundert	Crowding out and the Wealth of Nations
9030	Th. van de Klundert and R. Gradus	Optimal Government Debt under Distortionary Taxation
9031	A. Weber	The Credibility of Monetary Target Announcements: An Empirical Evaluation
9032	J. Osiewalski and M. Steel	Robust Bayesian Inference in Elliptical Regression Models
9033	C. R. Wichers	The Linear-Algebraic Structure of Least Squares
9034	C. de Vries	On the Relation between GARCH and Stable Processes
9035	M.R. Baye, D.W. Jansen and Q. Li	Aggregation and the "Random Objective" Justification for Disturbances in Complete Demand Systems
9036	J. Driffill	The Term Structure of Interest Rates: Structural Stability and Macroeconomic Policy Changes in the UK
9037	F. van der Ploeg	Budgetary Aspects of Economic and Monetary Integration in Europe
9038	A. Robson	Existence of Nash Equilibrium in Mixed Strategies for Games where Payoffs Need not Be Continuous in Pure Strategies
9039	A. Robson	An "Informationally Robust Equilibrium" for Two-Person Nonzero-Sum Games
9040	M.R. Baye, G. Tian and J. Zhou	The Existence of Pure-Strategy Nash Equilibrium in Games with Payoffs that are not Quasiconcave
9041	M. Burnovsky and I. Zang	"Costless" Indirect Regulation of Monopolies with Substantial Entry Cost

No.	Author(s)	Title
9042	P.J. Deschamps	Joint Tests for Regularity and Autocorrelation in Allocation Systems
9043	S. Chib, J. Osiewalski and M. Steel	Posterior Inference on the Degrees of Freedom Parameter in Multivariate-t Regression Models
9044	H.A. Keuzenkamp	The Probability Approach in Economic Methodology: On the Relation between Haavelmo's Legacy and the Methodology of Economics
9045	I.M. Bomze and E.E.C. van Damme	A Dynamical Characterization of Evolutionarily Stable States
9046	E. van Damme	On Dominance Solvable Games and Equilibrium Selection Theories
9047	J. Driffill	Changes in Regime and the Term Structure: A Note
9048	A.J.J. Talman	General Equilibrium Programming
9049	H.A. Keuzenkamp and F. van der Ploeg	Saving, Investment, Government Finance and the Current Account: The Dutch Experience
9050	C. Dang and A.J.J. Talman	The D_1-Triangulation in Simplicial Variable Dimension Algorithms on the Unit Simplex for Computing Fixed Points
9051	M. Baye, D. Kovenock and C. de Vries	The All-Pay Auction with Complete Information
9052	H. Carlsson and E. van Damme	Global Games and Equilibrium Selection
9053	M. Baye and D. Kovenock	How to Sell a Pickup Truck: "Beat-or-Pay" Advertisements as Facilitating Devices
9054	Th. van de Klundert	The Ultimate Consequences of the New Growth Theory; An Introduction to the Views of M. Fitzgerald Scott
9055	P. Kooreman	Nonparametric Bounds on the Regression Coefficients when an Explanatory Variable is Categorized
9056	R. Bartels and D.G. Fiebig	Integrating Direct Metering and Conditional Demand Analysis for Estimating End-Use Loads
9057	M.R. Veall and K.F. Zimmermann	Evaluating Pseudo-R^2's for Binary Probit Models
9058	R. Bartels and D.G. Fiebig	More on the Grouped Heteroskedasticity Model
9059	F. van der Ploeg	Channels of International Policy Transmission

No.	Author(s)	Title
9060	H. Bester	The Role of Collateral in a Model of Debt Renegotiation
9061	F. van der Ploeg	Macroeconomic Policy Coordination during the Various Phases of Economic and Monetary Integration in Europe
9062	E. Bennett and E. van Damme	Demand Commitment Bargaining: - The Case of Apex Games
9063	S. Chib, J. Osiewalski and M. Steel	Regression Models under Competing Covariance Matrices: A Bayesian Perspective
9064	M. Verbeek and Th. Nijman	Can Cohort Data Be Treated as Genuine Panel Data?
9065	F. van der Ploeg and A. de Zeeuw	International Aspects of Pollution Control
9066	F.C. Drost and Th. E. Nijman	Temporal Aggregation of GARCH Processes
9067	Y. Dai and D. Talman	Linear Stationary Point Problems on Unbounded Polyhedra
9068	Th. Nijman and R. Beetsma	Empirical Tests of a Simple Pricing Model for Sugar Futures
9069	F. van der Ploeg	Short-Sighted Politicians and Erosion of Government Assets
9070	E. van Damme	Fair Division under Asymmetric Information
9071	J. Eichberger, H. Haller and F. Milne	Naive Bayesian Learning in 2 x 2 Matrix Games
9072	G. Alogoskoufis and F. van der Ploeg	Endogenous Growth and Overlapping Generations
9073	K.C. Fung	Strategic Industrial Policy for Cournot and Bertrand Oligopoly: Management-Labor Cooperation as a Possible Solution to the Market Structure Dilemma
9101	A. van Soest	Minimum Wages, Earnings and Employment
9102	A. Barten and M. McAleer	Comparing the Empirical Performance of Alternative Demand Systems
9103	A. Weber	EMS Credibility
9104	G. Alogoskoufis and F. van der Ploeg	Debts, Deficits and Growth in Interdependent Economies
9105	R.M.W.J. Beetsma	Bands and Statistical Properties of EMS Exchange Rates

No.	Author(s)	Title
9106	C.N. Teulings	The Diverging Effects of the Business Cycle on the Expected Duration of Job Search
9107	E. van Damme	Refinements of Nash Equilibrium
9108	E. van Damme	Equilibrium Selection in 2 x 2 Games
9109	G. Alogoskoufis and F. van der Ploeg	Money and Growth Revisited
9110	L. Samuelson	Dominated Strategies and Commom Knowledge
9111	F. van der Ploeg and Th. van de Klundert	Political Trade-off between Growth and Government Consumption
9112	Th. Nijman, F. Palm and C. Wolff	Premia in Forward Foreign Exchange as Unobserved Components
9113	H. Bester	Bargaining vs. Price Competition in a Market with Quality Uncertainty
9114	R.P. Gilles, G. Owen and R. van den Brink	Games with Permission Structures: The Conjunctive Approach
9115	F. van der Ploeg	Unanticipated Inflation and Government Finance: The Case for an Independent Common Central Bank
9116	N. Rankin	Exchange Rate Risk and Imperfect Capital Mobility in an Optimising Model
9117	E. Bomhoff	Currency Convertibility: When and How? A Contribution to the Bulgarian Debate!
9118	E. Bomhoff	Stability of Velocity in the G-7 Countries: A Kalman Filter Approach
9119	J. Osiewalski and M. Steel	Bayesian Marginal Equivalence of Elliptical Regression Models
9120	S. Bhattacharya, J. Glazer and D. Sappington	Licensing and the Sharing of Knowledge in Research Joint Ventures
9121	J.W. Friedman and L. Samuelson	An Extension of the "Folk Theorem" with Continuous Reaction Functions
9122	S. Chib, J. Osiewalski and M. Steel	A Bayesian Note on Competing Correlation Structures in the Dynamic Linear Regression Model
9123	Th. van de Klundert and L. Meijdam	Endogenous Growth and Income Distribution
9124	S. Bhattacharya	Banking Theory: The Main Ideas
9125	J. Thomas	Non-Computable Rational Expectations Equilibria

No.	Author(s)	Title
9126	J. Thomas and T. Worrall	Foreign Direct Investment and the Risk of Expropriation
9127	T. Gao, A.J.J. Talman and Z. Wang	Modification of the Kojima-Nishino-Arima Algorithm and its Computational Complexity
9128	S. Altug and R.A. Miller	Human Capital, Aggregate Shocks and Panel Data Estimation
9129	H. Keuzenkamp and A.P. Barten	Rejection without Falsification - On the History of Testing the Homogeneity Condition in the Theory of Consumer Demand
9130	G. Mailath, L. Samuelson and J. Swinkels	Extensive Form Reasoning in Normal Form Games
9131	K. Binmore and L. Samuelson	Evolutionary Stability in Repeated Games Played by Finite Automata
9132	L. Samuelson and J. Zhang	Evolutionary Stability in Asymmetric Games
9133	J. Greenberg and S. Weber	Stable Coalition Structures with Uni-dimensional Set of Alternatives
9134	F. de Jong and F. van der Ploeg	Seigniorage, Taxes, Government Debt and the EMS
9135	E. Bomhoff	Between Price Reform and Privatization - Eastern Europe in Transition
9136	H. Bester and E. Petrakis	The Incentives for Cost Reduction in a Differentiated Industry
9137	L. Mirman, L. Samuelson and E. Schlee	Strategic Information Manipulation in Duopolies
9138	C. Dang	The D_2^*-Triangulation for Continuous Deformation Algorithms to Compute Solutions of Nonlinear Equations
9139	A. de Zeeuw	Comment on "Nash and Stackelberg Solutions in a Differential Game Model of Capitalism"
9140	B. Lockwood	Border Controls and Tax Competition in a Customs Union
9141	C. Fershtman and A. de Zeeuw	Capital Accumulation and Entry Deterrence: A Clarifying Note
9142	J.D. Angrist and G.W. Imbens	Sources of Identifying Information in Evaluation Models
9143	A.K. Bera and A. Ullah	Rao's Score Test in Econometrics
9144	B. Melenberg and A. van Soest	Parametric and Semi-Parametric Modelling of Vacation Expenditures

P.O. BOX 90153, 5000 LE TILBURG, THE NETHERLAND