Testing the normality assumption in the sample selection model with an application to travel demand

Klaauw, B. van der; Koning, R.H.

Testing the Normality Assumption in the Sample Selection Model with an Application to Travel Demand

Bas van der Klaauw [1]
Ruud H. Koning [2]

SOM-theme F Interactions between consumers and firms

[1]Corresponding author. Free University Amsterdam, Tinbergen Institute, and Scholar. Address: Department of Economics, Free University Amsterdam, De Boelelaan 1105, NL–1081 HV Amsterdam, The Netherlands, email `klaauw@tinbergen.nl`.

[2]University of Groningen. Address: Department of Econometrics, University of Groningen, P.O.Box 800,
NL–9700 AV Groningen, The Netherlands.

Abstract

In this paper we introduce a test for the normality assumption in the sample selection model. The test is based on a generalization of a semi-nonparametric maximum likelihood method. In this estimation method, the distribution of the error terms is approximated by a Hermite series, with normality as a special case. Because all parameters of the model are estimated both under normality and in the more general specification, we can test for normality using the likelihood ratio approach. This test has reasonable power as is shown by a simulation study. Finally, we apply the generalized semi-nonparametric maximum likelihood estimation method and the normality test to a model of car ownership and car use. The assumption of normal distributed error terms is rejected and we provide estimates of the sample selection model that are consistent.

Keywords: semi-nonparametric maximum likelihood, density estimation, Hermite series, sample selection.
JEL classification: C3, C5, D12, R41.

1 Introduction

Maximum likelihood is the most popular estimation method in micro-econometrics. The method yields consistent (in fact, asymptotically efficient) estimators if the model is specified correctly. However, correct specification may not be known beforehand. Two major sources of misspecification are incorrect specification of the functional form of the relationship under study (for example, omitting exogenous variables or inappropriately assuming linearity) and misspecification of the stochastic structure of the model (for example, neglecting heteroscedasticity or misspecification of the distribution of the random variables). The maximum likelihood estimator is generally inconsistent in these cases. In this paper we focus on one particular form of misspecification: misspecification of the distribution of the disturbances. We retain the assumption of correct specification of the functional form of the relationship.

The model we study is the sample selection model introduced by Heckman (1979). This type of model accounts for problems which arise because the outcome of the endogenous variable is observed only for a selective part of the sample. For example, sample selection models are used to study wage equations, where the wages of employed workers are observed only. Usually, the sample selection model is estimated by maximum likelihood, under the assumption that the error term of the regression equation and the error term of the selection equation follow a bivariate normal distribution. We introduce a formal test for this normality assumption. The test is derived from a generalization of the semi-nonparametric maximum likelihood estimation method of Gallant and Nychka (1987) in which the true distribution of the error terms is approximated by a Hermite series. Using this method the density of the disturbance terms is estimated together with all other parameters of the model. The test thus provides an alternative distribution of the disturbance terms in case normality is rejected. To examine the power of the normality test we perform a simulation study.

An advantage of the semi-nonparametric estimation used in this paper is that it is of general applicability: it is not specific to one particular econometric model. The approach taken in this paper can be used to examine the sensitivity of estimation results to the assumption of normality in other micro-econometric models where no formal test of normality is available. We will illustrate both the semi-nonparametric estimation method and the normality test in a model of car ownership and car use.

The plan of this paper is as follows. Section 2 discusses the sample selection model. In Section 3 we discuss the semi-nonparametric maximum likelihood method of Gallant and Nychka (1987). A generalization of this estimation method which naturally leads to a normality test is discussed in Section 4. Section 5 discusses an application of the estimation method and the normality test to the model of car ownership and car use. Finally, Section 6 concludes.

2 Identification and estimation of the sample selection model

In this section we discuss some identification and estimation issues of the sample selection model. Sample selection arises in case the outcome of the dependent variable is only observed for a (nonrandom) part of the sample. This may for example be caused by selective nonresponse or self-selection of individuals. An often used application of sample selection models is the estimation of wage equations (see for example Melenberg and Van Soest, 1993). In this case only the wages of employed workers are observed, which is a selective subsample of the whole population.

Let y denote the dependent variable and x a vector of exogenous variables. The binary variable z indicates whether the outcome of y is observed. For each individual in the sample the data reveal information on (z, x), but the realization of y is observed only if $z = 1$. We are interested in the probability distribution of y conditional on x, given by $\Pr(y|x)$. Using this distribution function we can compute the corresponding density function and the expectation of y given x (if these exist).

As mentioned above the data are not fully informative on $\Pr(y|x)$. Conditioning on z, we can write the distribution function as

$$\Pr(y|x) = \Pr(y|x, z = 1) \Pr(z = 1|x) + \Pr(y|x, z = 0) \Pr(z = 0|x)$$

The data identify the selection probability $\Pr(z = 1|x)$, the censoring probability $\Pr(z = 0|x)$, and the distribution of outcomes conditional on observing the outcome $\Pr(y|x, z = 1)$. But the data provide no information on the distribution of counterfactuals $\Pr(y|x, z = 0)$ (see for a more extensive overview Manski, 1989, 1995). Without any additional assumptions it is possible to establish bounds on $\Pr(y|x)$ (see Manski, 1995). Since $\Pr(y|x, z = 0)$ lies in the interval $[0, 1]$ we have

$$\Pr(y|x, z = 1) \Pr(z = 1|x) \leq \Pr(y|x) \leq \Pr(y|x, z = 1) \Pr(z = 1|x) + \Pr(z = 0|x)$$

The width of this interval is $\Pr(z = 0|x)$. In general, less censoring in the data causes smaller intervals. Note that if no censoring is observed, $\Pr(z = 0|x)$ reduces to 0 and thus $\Pr(y|x)$ equals $\Pr(y|x, z = 1)$.

Without any prior information or additional assumptions it is not possible to identify $\Pr(y|x)$. In economic literature, an assumption made frequently is that at least one of the regressors affects the selection probability but not the conditional probability of y (such a regressor is an instrumental variable). Without such an exclusion restriction identification hinges entirely on the functional form and distributional assumptions (see Manski, 1989). Exclusion restrictions have some identifying power as the interval in which $\Pr(y|x)$ lies becomes smaller. However, exclusion restrictions only identify $\Pr(y|x)$ in case an instrumental variable exists that perfectly predicts whether or not y is observed.

The most common used method to ensure identification of the sample selection model is by parameterization of the model. Usually linear specifications for both the regression equation and the selection equation are chosen, and the error terms are supposed to follow a probability distribution that is known except for certain parameters (see Heckman, 1979). We follow this approach and we assume to have a random sample of N individuals. For individual i $(i = 1, \ldots, N)$ the regression equation is given by

$$y_i = \beta_1' x_{1i} + \varepsilon_{1i} \tag{1}$$

However, the variable of interest y_i is observed for a nonrandom subsample only. The selection rule is given by

$$\begin{aligned} z_i^* &= \beta_2' x_{2i} + \varepsilon_{2i} \\ z_i &= \begin{cases} 1 & z_i^* > 0 \\ 0 & z_i^* \leq 0 \end{cases} \end{aligned} \tag{2}$$

If the conditional expectation of ε_{1i} given $z_i = 1$ does not equal 0, OLS-estimation of (1) will not yield consistent estimates for β_1.

Let $f(\cdot,\cdot)$ denote the bivariate density function of $\varepsilon_i = (\varepsilon_{1i}, \varepsilon_{2i})'$. The log-likelihood function for the sample selection model is

$$\begin{aligned} \log \mathcal{L}(\theta) = & \sum_{i=1}^{N} z_i \log \left(\int_{-\beta_2' x_{2i}}^{\infty} f(y_i - \beta_1' x_{1i}, \varepsilon_2) d\varepsilon_2 \right) \\ & + (1 - z_i) \log \left(\int_{-\infty}^{-\beta_2' x_{2i}} \int_{-\infty}^{\infty} f(\varepsilon_1, \varepsilon_2) d\varepsilon_1 d\varepsilon_2 \right) \end{aligned} \tag{3}$$

If one is willing to assume that $f(\cdot,\cdot)$ is the bivariate normal density function, an alternative to maximum likelihood is using Heckman's two-stage procedure to estimate the parameters (see Heckman, 1979). However, the estimates are rather sensitive to the distributional assumption (see for an overview Manski, 1989). So, one would like to test this distributional assumption or estimate the model under a less restrictive distributional assumption.

3 Semi-nonparametric maximum likelihood estimation

As an alternative to maximum likelihood estimation of the sample selection model, we discuss in this section semi-nonparametric estimation as introduced by Gallant and Nychka (1987). The semi-nonparametric estimation method is based on the approximation of the (unknown) density function $f(\cdot,\cdot)$ by a Hermite series (see Powell, 1994; for an overview of semi-parametric estimation). In the first part of this section we recapitulate the estimation approach of Gallant and Nychka (1987) and in the second part of this section we consider applying this method to the sample selection model.

Elaborating on Phillips (1983), Gallant and Nychka (1987) proposed approximating the unknown density in a model by a Hermite series. Phillips (1983) showed that an extended rational approximant (ERA) of the form

$$h(\varepsilon) = \frac{P^2(\varepsilon)}{Q^2(\varepsilon)}\phi^2(\varepsilon|\tau,\Sigma) \tag{4}$$

can approximate any density function satisfying certain regularity conditions arbitrarily well. In (4), $P(\varepsilon)$ and $Q(\varepsilon)$ are polynomials and $\phi(\varepsilon|\tau,\Sigma)$ is the multivariate normal density function with mean τ and covariance matrix Σ. Of course, (4) is not a proper density function if the polynomials $P(\varepsilon)$ and $Q(\varepsilon)$ are not restricted such that it integrates to 1.

Gallant and Nychka (1987) restrict the density $h(\varepsilon)$ to a subclass $\mathcal{H}_K$ which consists of densities of the Hermite form

$$h(\varepsilon) = P_K^2(\varepsilon-\tau)\phi^2(\varepsilon|\tau,\Delta) \tag{5}$$

with Δ a diagonal matrix. $P_K(\cdot)$ is a polynomial of degree K. Gallant and Nychka (1987) show that, by increasing the number of terms K of the polynomial, a large class $\mathcal{H}$ of density functions can be approximated arbitrarily well. Conditions defining $\mathcal{H}$ precisely are given in Gallant and Nychka (1987).

For our purposes it suffices to note that the fattest tails allowed are t-like tails and the thinnest tails allowed are thinner than normal-like tails. Any sort of skewness and kurtosis (especially in that part of the distribution where most probability mass is observed) is allowed, only very violently oscillatory densities are excluded from $\mathcal{H}$. Gallant and Nychka (1987) prove that densities in $\mathcal{H}$ can be estimated consistently by increasing the number of terms K in the approximation with the number of observations.

It is also possible to assume that the true density is a member of $\mathcal{H}_K$ and hence, to interpret $\mathcal{H}_K$ as a flexible class of density functions. The latter interpretation is especially appealing if one wants to examine the sensitivity of estimation results obtained by assuming normality to this distributional assumption because it allows one to use the standard framework of inference. In (5), the normal density is used as the base class for $\mathcal{H}_K$ but this is not necessary: any density with a moment generating function could be used (for example, see Cameron and Johansson, 1997).

Gallant and Nychka (1987) parameterize $h(\varepsilon)$ as

$$\begin{aligned} h^*(\varepsilon) &= \left(\sum_{i_1,\ldots,i_n=0}^{K} \alpha_{i_1\cdots i_n} (\varepsilon_1 - \tau_1)^{i_1} \cdots (\varepsilon_n - \tau_n)^{i_n} \right)^2 \\ &\qquad \times \exp\left(- \left[(\varepsilon_1 - \tau_1)^2/\delta_1^2 + \cdots + (\varepsilon_n - \tau_n)^2/\delta_n^2 \right] \right) \\ &= \sum_{i_1,\ldots,i_n,j_1,\ldots,j_n=0}^{K} \alpha_{i_1\cdots i_n} \alpha_{j_1\cdots j_n} (\varepsilon_1 - \tau_1)^{i_1+j_1} \cdots (\varepsilon_n - \tau_n)^{i_n+j_n} \\ &\qquad \times \exp\left(- \left[(\varepsilon_1 - \tau_1)^2/\delta_1^2 + \cdots + (\varepsilon_n - \tau_n)^2/\delta_n^2 \right] \right) \end{aligned} \tag{6}$$

Because of the squaring in (6), no additional restrictions on the parameters are necessary to ensure that $h^*(\varepsilon)$ is nonnegative. Additional restrictions on the parameters of the density are required for identification of other parameters in a model but these restrictions depend on the type of model at hand. The parameters cannot be chosen freely, one restriction will be needed to ensure integration to 1. These restrictions can take the form of explicit restrictions on the parameters of the density. However, for computational convenience we follow Gabler, Laisney and Lechner (1993) who ensure integration to 1 by scaling the density. Define S by

$$\begin{aligned} S &= \int_{R^n} \sum_{i_1,\ldots,i_n,j_1,\ldots,j_n=0}^{K} \alpha_{i_1\cdots i_n}\alpha_{j_1\cdots j_n}(\varepsilon_1-\tau_1)^{i_1+j_1}\cdots(\varepsilon_n-\tau_n)^{i_n+j_n} \\ &\qquad \times \exp\left(-\left[(\varepsilon_1-\tau_1)^2/\delta_1^2+\cdots+(\varepsilon_n-\tau_n)^2/\delta_n^2\right]\right) d\varepsilon_1\cdots d\varepsilon_n \\ &= \sum_{i_1,\ldots,i_n,j_1,\ldots,j_n=0}^{K} \alpha_{i_1\cdots i_n}\alpha_{j_1\cdots j_n}\int_R(\varepsilon_1-\tau_1)^{i_1+j_1}\exp\left(-(\varepsilon_1-\tau_1)^2/\delta_1^2\right)d\varepsilon_1 \\ &\qquad \cdots\int_R(\varepsilon_n-\tau_n)^{i_n+j_n}\exp\left(-(\varepsilon_n-\tau_n)^2/\delta_n^2\right)d\varepsilon_n \end{aligned}$$

See Appendix A for the recursion formulae, which can be used to explicitly determine S. Because of the definition of S, the following density integrates to 1:

$$h(\varepsilon) = h^*(\varepsilon)/S. \tag{7}$$

We will refer to densities of the type (7) as snp-densities. It is clear that α in (7) is identified up to a scale only, so a normalization is necessary. In particular applications, additional restrictions will be needed to achieve identification. For most applications it will be convenient to set τ to 0 which we will do from now on. The flexibility of snp-densities is illustrated in Figures 1–3 ($K = 2$). It is clear that the contour lines differ from the usual ellipsoids of the bivariate normal density.

This estimation approach to the sample selection model has been implemented by Melenberg and Van Soest (1993). Their choice of snp-density is $h(\varepsilon) = h^*(\varepsilon)/S$ with

$$h^*(\varepsilon) = \sum_{i,j,k,l=0}^{K} \alpha_{ij}\alpha_{kl}\varepsilon_1^{i+k}\varepsilon_2^{j+l}\exp(-[\varepsilon_1^2/\delta_1^2+\varepsilon_2^2/\delta_2^2])$$

Identification is achived by setting $\delta_2 = \sqrt{2}$ (to ensure identification of the scale of (2)), and $\alpha_{00} = 1$ to normalize the α's. For $K = 0$, $h(\varepsilon)$ now reduces to a bivariate normal density with zero correlation between ε_1 and ε_2. Finally, complex nonlinear restrictions on the parameters are needed to ensure that the means of ε_1 and ε_2 are 0 in case $K \geq 0$. Melenberg and Van Soest (1993) suggest not to impose restrictions on the parameters of the density function of ε to ensure a zero mean, but to restrict the intercepts of (1) and (2) instead. However, these are not a useful restrictions for the purpose of this paper. We return to this issue in the next section.

4 Testing the normality assumption

Semi-nonparametric maximum likelihood estimation as discussed in the previous section does not allow us to test for normality in the sample selection model (unless ε_1 and ε_2 are independent). The bivariate normal distribution is not a special case of the class of snp-densities (7). In this section we discuss a more general specification. This specification allows to test for normality, even if the error terms are correlated, by choosing another base class of density functions in the ERA approximation in (4).

Because any density function with a finite moment generating function can be used as the basis in approximation (4), we can consider the following family of functions:

$$\bar{h}^*(\varepsilon) = \sum_{i,j,k,l=0}^{K} \alpha_{ij}\alpha_{kl}\varepsilon_1^{i+k}\varepsilon_2^{j+l}\exp(-\varepsilon'\Sigma^{-1}\varepsilon)$$

and define a generalized snp-density by $\bar{h}(\varepsilon) = \bar{h}^*(\varepsilon)/S$ (again, S is the constant that ensures integration to 1). Even though the use of this generalized snp-density is not necessary to obtain consistent estimates of the parameters of the model (the parameters are estimated consistently if the model is identified and if the number of terms K increases with the number of observations), a clear advantage is that bivariate normality (with unrestricted correlation) is a special case of this family ($\alpha_{ij} = 0$ for all $i + j \geq 1$). This implies that it is possible to test for normality in this model. A test for normality in the sample selection model has not been derived in the literature. A disadvantage of the generalized snp-density in equation (4) is that it does not have the same computationally attractive properties, i.e. evaluation of the relevant integrals in the loglikelihood function 3 will involve evaluation of bivariate normal probabilities, in general. Because of the special structure of the sample selection model, we are able to avoid evaluations of bivariate normal integrals, so the computational cost of this generalization is limited (see Appendix B).

As mentioned in the previous section some restrictions are necessary to ensure identification. First, again we fix $\delta_2 = \sqrt{2}$ and $\alpha_{00} = 1$. Second, to ensure that the location of the distribution function is fixed, we optimize the loglikelihood function conditional to the restriction that the means of ε_1 and ε_2 equal 0. This allows us to test the null hypothesis that $(\varepsilon_1, \varepsilon_2)$ is distributed according to a normal distribution function against the alternative hypothesis that $(\varepsilon_1, \varepsilon_2)$ has some other bivariate distribution function in the class of distribution functions $\mathcal{H}_K$ for any fixed K. In other words: for some fixed K we test for joint significance of all α_{ij} with $i + j \geq 1$. The additional number of parameters under the alternative hypotheses compared to the null hypotheses is $(K+1)^2 - 1$. Note that there are two restrictions on these parameters to fix the location of the distribution. Therefore, the Likelihood Ratio test statistic is distributed according to a χ^2-distribution with $(K+1)^2 - 3$ degrees of freedom.

We conduct a limited simulation exercise to examine the power of this normality test. We consider the following simulation experiment:

$$\begin{aligned} y_i &= \beta_{10} + \beta_{11} x_i + \beta_{12} w_i + \varepsilon_{1i} \\ z_i^* &= \beta_{20} + \beta_{21} v_i + \beta_{22} w_i + \varepsilon_{2i} \qquad i = 1, \ldots, N \end{aligned}$$

with true parameters $\beta_{10} = 1$, $\beta_{11} = 0.5$, $\beta_{12} = -0.5$, $\beta_{20} = 1$, $\beta_{21} = -1$ and $\beta_{22} = 1$. The exogenous variables x_i and v_i are independently $\mathcal{N}(0, 3)$ distributed and w_i is distributed uniformly on $[-3, 3]$. We perform six experiments, where we vary the distribution of ε and the number of observations. Within each experiment, we draw 100 samples. We draw ε from either a bivariate normal distribution with mean 0, a bivariate t-distribution, and a centered χ^2-distribution. The t-distribution has fatter tails than the normal distribution, and the χ^2-distribution is asymmetric so both cases are a deviation from normality. For all three experiments we set $\text{var}(\varepsilon_1) = 4$, $\text{var}(\varepsilon_2) = 1$ and $\text{cov}(\varepsilon_1, \varepsilon_2) = 1$.[1] The sample size N is either 500 or 1000. The simulations were performed on Pentium workstations using the CML-library (Constrained Maximum Likelihood library) of GAUSS.

As the computer time required for optimization of the loglikelihood function increases quickly with K, we only estimate the model for $K = 1$ and $K = 2$ and for the bivariate normal distributed disturbances. Hence, we consider the normality test against the class of generalized snp-densities with $K = 1$ and $K = 2$ (denoted by $\mathcal{H}_1^*$ and $\mathcal{H}_2^*$). The case of the normal-disturbances is presented in detail in Tables 1 and 2, the case of the t-disturbances in Tables 3 and 4, and the case of the χ^2-disturbances in Tables 5 and 6.

It is remarkable how well standard Maximum Likelihood under the assumption of normally distributed disturbances performs. Even if the true disturbances follow a t-distribution all estimated parameters are within two standard deviations of their true values. This is even the case when the disturbances follow a transformed χ^2-distribution which is non-symmetric. The tables also report the number of rejections of normality for each of the simulation experiments. The simulation results indicate that normality test against both the class $\mathcal{H}_1^*$ and $\mathcal{H}_2^*$ performs well. The number of incorrect rejections is small when compared to the level of significance, and the number of correct rejections very high. The simulation results indicate that the test has more power when the true distribution is assumed to belong to $\mathcal{H}_2^*$ than when it is assumed to belong to $\mathcal{H}_1^*$.

[1] To be precise, in case of the bivariate t-distribution $\varepsilon_2 = u_1/\sqrt{3}$ and $\varepsilon_1 = \varepsilon_2 + u_2$ with u_1 and u_2 independent draw from a t_3-distribution, and in case of the bivariate χ^2-distribution $\varepsilon_2 = \frac{1}{2}(v_1^2 + v_2^2) - 1$ and $\varepsilon_1 = \varepsilon_2 + (v_3^2 + v_4^2 + v_5^2)/\sqrt{2} - 3/\sqrt{2}$ with v_1^2 to v_5^2 independent $\chi^2(1)$ random variates. Hence, $\mathrm{E}[\varepsilon_1] = \mathrm{E}[\varepsilon_2] = 0$ and $\text{var}(\varepsilon_1) = 4$ and $\text{var}(\varepsilon_2) = \text{cov}(\varepsilon_1, \varepsilon_2) = 1$.

The estimated parameters of the selection and the regression equations do not differ much between the experiments. As one would expect, the price of more flexibility in specification of the distribution of the error terms is that the parameters of the model are estimated with less precision. However, the increase of the variance of the parameters is small.

5 An application to a simultaneous model of car ownership and car use

5.1 The model

In this section we apply the generalized semi-nonparametric estimation method and the normality tests to a model describing the ownership and the use of cars. This section consists of three parts: first, we present a structural model, then we discuss the data and finally we present the empirical results.

The model assumes that, in a given year, households need to travel a stochastic number of kilometers and that these households minimize the costs of traveling. At the beginning of the year, a household makes a prediction of the number of kilometers to be traveled, which is denoted by y^*. The actual number of kilometers traveled in this year is related to the prediction by $y = y^* \cdot \varepsilon$, where ε is a random prediction error with a positive support and a finite mean μ. The household can choose to travel either by car or by other transport. The latter is most likely public transport, other possibilities may be bikes, car pooling, etc. The cost of using a car consists of fixed costs c_c and variable costs v_c. Fixed costs are depreciation of the car, maintenance, insurance and taxations for car ownership. Variable costs are the costs of fuel. The costs of other transport do not contain any fixed costs. The variable costs, denoted by v_o, are associated with the costs of public transport. However, also opportunity costs may be important, as traveling with public transport most often takes more time than traveling by car.

The household decision problem now involves choosing whether or not to use a car. Assuming that (risk-neutral) households minimize their expected costs, the decision rule implies that a car is used if

$$\mathrm{E}\left[c_c + v_c y\right] \leq \mathrm{E}\left[v_0 y\right] \tag{8}$$

Now assume that c_c, v_c and v_o are known by the household. This means that the household decision simplifies to using a car if

$$\mathrm{E}[y] \geq \frac{c_c}{v_o - v_c} \tag{9}$$

If we substitute the relation between the actual number of kilometers driven and the predicted number of kilometers, this implies that a household chooses to use a car if

$$y^* \geq \frac{c_c}{\mu(v_o - v_c)}$$

Now we parameterize the unknown components of the model as

$$y^* = \exp(x'\beta + v_1)$$

which is the distance function and a costs function

$$\frac{c_c}{v_o - v_c} = \exp(z'\gamma + v_2)$$

We normalize $\mu = 1$.

The model reduces to a regression equation, which has the following structure

$$\log(y) = x'\beta + \varepsilon_1$$

with $\varepsilon_1 = v_1 + \ln(\varepsilon)$. A selection equation indicates if the household owns a car

$$x'\beta - z'\gamma + \varepsilon_2 \geq 0$$

with the household specific term ε_2 equal to $v_1 - v_2$, which is known to the household but generally unknown to the econometrician. Because v_1 is included in both ε_1 and ε_2, these disturbances are not a priori independent.

To improve the identification of the model, we should impose an exclusion restriction, i.e. find a variable which is included in z, and excluded from x. According to the model, the predicted number of kilometers affects both the actual number of kilometers and the decision to buy a car, while the fixed and variable costs of using a car and the variable costs of other transport only affect the decision to use a car. The exclusion restriction should therefore be a variable which affects only the costs functions. Since the Dutch government provides free public transport passes to some students and some civil servants, we use the dummy variable that indicates if the household has such a free public transport pass as a variable which is included in z but not in x.

The data we use is a subset from the Dutch database on transportation behavior of Statistics Netherlands. This database contains 34454 households. To avoid complications of households owning more than one car, we focus on single-person households. Furthermore, we exclude individuals who are younger than 18 years old, as this is the legal age for obtaining a drivers license in The Netherlands. This restrict the database to 7404 individuals. To construct our final data set we also exclude 130 individuals, who own a car, but for which the number of kilometers driven in the past year is unknown and 756 individuals for which one or more explanatory variables are missing. We finally use a data set consisting of 6518 observations.

Table 7 provides some characteristics of the data set. The data contain 3110 individuals who own a car and 3408 who do not own a car. Except for region, all variables display differences in car-ownership rates. While 58% of the men are car-owners, only 41% of the women have a car. Until people reach the age of 65 the car-ownership rates increase with age, after that there is a large drop. Furthermore, car-ownership rates increase with income and the level of education. Finally, individuals living in areas with a low degree of urbanization, full-time employed workers and individuals who do not have a government-provided free public transport pass are more likely to own a car than their counterparts.

Using these data we estimate the structural model discussed previously. In the first step we estimate the model under the assumption that the disturbances ε_1 and ε_2 follow a bivariate normal distribution. Next we relax this assumption by assuming that the density of the disturbances belongs either to $\mathcal{H}_1^*$ or $\mathcal{H}_2^*$. For these two cases we will test whether the restriction of normality of the disturbances can be imposed or not.

Table 8 presents the estimation results under the assumption of normality. The most important covariates in the distance function are gender, age, income, level of education, and the individual labor market status, while age, degree of urbanization, and income are the main covariates determining the costs function. Although the availability of a free public transport pass provided by the government has the expected effect on the costs function, the corresponding parameter estimate is not significantly different from 0.

The only individual characteristics, which are important in both the distance and the costs function are age and income. Age affects both the distance function and the costs function negatively. Young people travel more kilometers than older people, while the costs of car use relative to other transport decrease over age. By comparing the coefficients we can see that people with age between 50 and 64 are most likely to own a car. Income has an opposite effect on the distance and the costs function. People with a higher income travel more kilometers, while the costs of car use relative to other transport decrease. Considering that using public transport takes more time, this latter effect can be explained by differences in opportunity costs. Time is more costly for individuals with a high income. The degree of urbanization only affects the costs function. Since there is less public transport available in areas with low degree of urbanization, the costs of using public transport are higher in these areas (e.g. waiting times are longer). On the other hand, car use in areas with a high degree of urbanization is more costly because individuals often have additional costs such as parking costs. Employed individuals, both full-time and part-time, travel more kilometers. Since we did not distinguish between traveling for private purposes and for professional purposes, this may be caused because they commute or because their work requires them to travel. Finally, people with a higher level of education travel more than lower educated individuals.

In the second estimation step we have relaxed the assumption of normality and have estimated the model using the generalized snp-density with $K = 1$ and $K = 2$. The likelihood ratio test statistics equal 283.7 and 299.6, which implies that we must reject that the disturbances follow a normal distribution. The parameter estimates do not seem to be very sensitive to the normality assumption, which is in line with the results from the simulation study. The main difference with the estimates under normality is that, under the assumption of normality, we do not find any correlation between ε_1 and ε_2. This suggests that there is no unobserved selection between car ownership and car use, which would imply that both the regression and the selection equation could be estimated consistently separately of each other by for example OLS and Probit, respectively. The generalized snp-densities show that there is correlation between both disturbance terms, which could not be captured by a normal density. The estimated covariance between ε_1 and ε_2 is -0.23 for $K = 1$ and -0.30 for $K = 2$.[2] The Figures 4 and 5 show the marginal densities of the disturbances ε_1 and ε_2 estimated under the generalized snp-densities and normality. Both generalized snp-densities have more mass close to the mode and slightly fatter tails than the normal density. The estimated standard deviation of ε_1 is 0.64 in both the $K = 1$ and the $K = 2$ generalized snp-density specification, which is almost

[2] In the context of our theoretical model this would imply a positive correlation between v_1 and v_2, since $\text{cov}(\varepsilon_1, \varepsilon_2) = \text{var}(v_1) - \text{cov}(v_1, v_2)$. Obviously, there are some unobserved covariates, which increase both the costs of car use relative to other transport and the expected number of kilometers traveled. Such a covariate thus has similar effects as for example age. Note that we maintained the assumption that the prediction error ε is independent of the individual specific effects v_1 and v_2.

similar to the estimate under normality. For the standard deviation of ε_2 we find 0.89 and 0.90 for ε_2 for the $K = 1$ and the $K = 2$ generalized snp-density, respectively. In Figures 6 and 7 we graph the estimated densities of both the normal model and the model estimated using an snp-density ($K = 2$). We see that the estimated snp-density is a bit more spread out that the normal density and the estimated contour lines in Figure 7 are not the ellipsoids of Figure 6.

6 Conclusion

In this paper we derived a test for normality in sample selection models. To our knowledge, no such test has been derived previously. The test is based on a generalization of the semi-nonparametric maximum likelihood method introduced by Gallant and Nychka (1987). It assumes that the number of terms in the series approximation is known in advance. The generalization exploits the special structure of the sample selection model to allow for a bivariate normal base density. The main advantage of this generalization is that the bivariate normal distribution is a special case of the class of generalized snp-densities, which allows us to test for normality. Also the generalized snp-density provides an alternative density function in case normality is rejected.

Although the simulation study provided in this paper is limited, we think that this test is promising. The test performed well in the simulations, the percentage of incorrect rejections is below the significance level, while the power is high. This latter is especially true for the normality test against the class of generalized snp-densities with $K = 2$. However, it should be noted that the parameter estimates do not seem to be very sensitive to the distributional assumptions of the disturbances. Even in case the disturbances do not follow a normal distribution, Maximum Likelihood under the assumption of normality provides estimates close to the true values.

Finally, we have applied the normality test to a model of car ownership and car use. The empirical results mimic those found in the simulation study: we reject the assumption of normality, but the parameter estimates turn out to be not very sensitive to the normality assumption. In this case, the generalized snp-density is capable to correct for the unobserved selectivity, which is not captured by the normal density.

References

Cameron, A.C. and P. Johansson (1997), Count data regression using series expansions: with applications, *Journal of Applied Econometrics* 12, 203–223.

Gabler, S., F. Laisney and M. Lechner (1993), Semi-nonparametric maximum likelihood estimation 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.W. Nychka (1987), Semi-nonparametric maximum likelihood estimation, *Econometrica* 55, 363-390.

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

Manski, C.F. (1989), Anatomy of the selection problem, *Journal of Human Resources* 24, 343–360.

Manski, C.F. (1995), *Identification Problems in the Social Sciences*, Harvard University Press, Cambridge.

Melenberg, B. and A. van Soest (1993), Semi-parametric estimation of the sample selection model, Working Paper, CentER, Tilburg.

Phillips, P.C.B. (1983), ERA's: A new approach to small sample theory, *Econometrica* 51, 1505-1525.

Powell, J.L. (1994), Estimation of semiparametric models, in R.F. Engle and D.L. McFadden (eds.), *Handbook of Econometrics, vol. 4*, North-Holland, Amsterdam.

A Recursion formulae of the snp-density

$I_k(a,b)$ is defined as the univariate integral

$$I_k(a,b) = \int_a^b u^k \exp(-u^2/\delta^2) du$$

Using partial integration one obtains the recursion formulae

$$I_k(a,b) = \frac{\delta^2}{2} \left(a^{k-1} \exp(-a^2/\delta^2) - b^{k-1} \exp(-b^2/\delta^2) \right) + \frac{(k-1)\delta^2}{2} I_{k-2}(a,b)$$

if $k \geq 2$, else

$$I_1(a,b) = \frac{\delta^2}{2} \left(\exp(-a^2/\delta^2) - \exp(-b^2/\delta^2) \right)$$

$$I_0(a,b) = \delta\sqrt{\pi} \left(\Phi\left(\frac{\sqrt{2}b}{\delta}\right) - \Phi\left(\frac{\sqrt{2}a}{\delta}\right) \right)$$

where $\Phi(\cdot)$ is the standard normal distribution function. In the special case where $a = -\infty$ and $b = \infty$, the recursion formulae simplify to

$$I_k(-\infty,\infty) = \begin{cases} \delta\sqrt{\pi} & k = 0 \\ 0 & k = 1,3,5,\ldots \\ \frac{(k-1)\delta^2}{2} I_{k-2}(-\infty,\infty) & k = 2,4,6,\ldots \end{cases}$$

B Relevant integrals of the generalized snp-density in the sample selection model

In the sample selection model where all relevant integrals are of the form

$$\int_a^\infty \bar{h}^*(\varepsilon) d\varepsilon_2$$

and

$$\int_{-\infty}^b \int_{-\infty}^\infty \bar{h}^*(\varepsilon) d\varepsilon_1 d\varepsilon_2$$

Substituting for $\bar{h}$ we obtain integrals of the type

$$\int_a^\infty \varepsilon_1^i \varepsilon_2^j \phi(\varepsilon_1, \varepsilon_2) d\varepsilon_2$$

and

$$\int_{-\infty}^{b}\int_{-\infty}^{\infty}\varepsilon_1^i\varepsilon_2^j\phi(\varepsilon_1,\varepsilon_2)d\varepsilon_1 d\varepsilon_2$$

where $\phi(\varepsilon_1,\varepsilon_2)$ is the bivariate normal density function. Because

$$\phi(\varepsilon_1,\varepsilon_2)=\phi(\varepsilon_2|\varepsilon_1)\phi(\varepsilon_1)$$

we can rewrite these integrals as

$$\varepsilon_1^i\phi(\varepsilon_1)\int_{a}^{\infty}\varepsilon_2^j\phi(\varepsilon_2|\varepsilon_1)d\varepsilon_2$$

and

$$\int_{-\infty}^{b}\varepsilon_2^j\phi(\varepsilon_2)\int_{-\infty}^{\infty}\varepsilon_1^i\phi(\varepsilon_1|\varepsilon_2)d\varepsilon_1 d\varepsilon_2=\int_{-\infty}^{b}\varepsilon_2^j\phi(\varepsilon_2)\mathrm{E}(\varepsilon_1^i|\varepsilon_2)d\varepsilon_2 \qquad (10)$$

The last integral can be solved easily because

$$\mathrm{E}(\varepsilon_1^i|\varepsilon_2)=a_0+a_1\varepsilon_2+\cdots+a_i\varepsilon_2^i.$$

The coefficients a depend on the other parameters of the density function only and they are independent of ε_2. Note that both integrals in (10) can be calculated using the recursion formulas in Appendix A.

$N = 500$	normal		generalized snp			
			$K = 1$		$K = 2$	
β_{10}	1.02	(0.13)	1.02	(0.13)	1.02	(0.13)
β_{11}	0.50	(0.032)	0.50	(0.032)	0.50	(0.032)
β_{12}	-0.50	(0.074)	-0.50	(0.072)	-0.50	(0.072)
β_{20}	1.01	(0.11)	1.01	(0.11)	1.01	(0.11)
β_{21}	-1.01	(0.075)	-1.01	(0.075)	-1.01	(0.075)
β_{22}	1.01	(0.081)	1.01	(0.081)	1.01	(0.082)
σ_1	1.98	(0.087)	1.98	(0.085)	1.97	(0.090)
σ_{12}	1.00	(0.088)	1.00	(0.090)	1.00	(0.094)
α_{00}			1		1	
α_{01}			0		-0.0036	(0.030)
α_{02}					-0.0042	(0.024)
α_{10}			0		-0.0014	(0.015)
α_{11}			-0.0012	(0.0092)	0.0067	(0.030)
α_{12}					0.0023	(0.028)
α_{20}					-0.0004	(0.0084)
α_{21}					-0.0006	(0.011)
α_{22}					-0.0022	(0.0066)
$\hat{\sigma}_1$			1.98	(0.091)	1.96	(0.11)
$\hat{\sigma}_2$			1.00	(0.018)	0.99	(0.030)
$\hat{\sigma}_{12}$			0.99	(0.15)	0.98	(0.16)
loglikelihood	-689.38		-689.12		-687.80	
rejections			1		0	

The standard errors are given in parentheses.

Table 1: Results of the simulation experiment with bivariate normal distributed disturbances, 100 replications, $N = 500$.

$N = 1000$	normal		generalized snp $K = 1$		$K = 2$	
β_{10}	1.00	(0.072)	1.00	(0.073)	1.00	(0.074)
β_{11}	0.50	(0.025)	0.50	(0.024)	0.50	(0.025)
β_{12}	-0.51	(0.051)	-0.51	(0.049)	-0.51	(0.049)
β_{20}	1.00	(0.060)	1.00	(0.062)	1.00	(0.061)
β_{21}	-0.99	(0.045)	-0.99	(0.043)	-0.99	(0.043)
β_{22}	0.99	(0.054)	0.99	(0.054)	0.99	(0.054)
σ_1	2.00	(0.060)	2.00	(0.059)	2.00	(0.059)
σ_{12}	1.00	(0.034)	1.00	(0.036)	1.00	(0.036)
α_{00}			1		1	
α_{01}			0		0.0001	(0.019)
α_{02}					0.0026	(0.012)
α_{10}			0		0.0003	(0.010)
α_{11}			-0.0002	(0.0064)	0.0034	(0.017)
α_{12}					0.0006	(0.017)
α_{20}					-0.0004	(0.0051)
α_{21}					-0.0005	(0.0067)
α_{22}					-0.0014	(0.0050)
$\hat{\sigma}_1$			2.00	(0.061)	2.00	(0.070)
$\hat{\sigma}_2$			1.00	(0.013)	1.00	(0.020)
$\hat{\sigma}_{12}$			1.00	(0.087)	1.01	(0.10)
loglikelihood	-1412.71		-1412.35		-1411.18	
rejections			0		1	

The standard errors are given in parentheses.

Table 2: Results of the simulation experiment with bivariate normal distributed disturbances, 100 replications, $N = 1000$.

$N = 500$	**normal**		**generalized snp** $K = 1$		$K = 2$	
β_{10}	0.98	(0.11)	0.98	(0.11)	0.98	(0.11)
β_{11}	0.50	(0.039)	0.50	(0.038)	0.50	(0.035)
β_{12}	-0.51	(0.074)	-0.51	(0.072)	-0.51	(0.067)
β_{20}	1.06	(0.20)	1.06	(0.20)	1.07	(0.20)
β_{21}	-1.06	(0.18)	-1.06	(0.18)	-1.08	(0.18)
β_{22}	1.07	(0.18)	1.07	(0.18)	1.08	(0.18)
σ_1	1.97	(0.33)	1.98	(0.34)	2.01	(0.34)
σ_{12}	0.97	(0.23)	1.02	(0.29)	0.97	(0.27)
α_{00}			1		1	
α_{01}			0		0.0099	(0.030)
α_{02}					-0.051	(0.14)
α_{10}			0		0.0056	(0.032)
α_{11}			-0.016	(0.048)	-0.031	(0.062)
α_{12}					0.0085	(0.041)
α_{20}					-0.040	(0.026)
α_{21}					-0.0062	(0.020)
α_{22}					0.011	(0.014)
$\hat{\sigma}_1$			1.96	(0.32)	1.98	(0.36)
$\hat{\sigma}_2$			0.99	(0.042)	1.00	(0.050)
$\hat{\sigma}_{12}$			0.94	(0.28)	0.98	(0.31)
loglikelihood	-678.87		-675.64		-652.50	
rejections			11		89	

The standard errors are given in parentheses.

Table 3: Results of the simulation experiment with bivariate t distributed disturbances, 100 replications, $N = 500$.

$N = 1000$	**normal**		**generalized snp** $K = 1$		$K = 2$	
β_{10}	0.99	(0.076)	1.03	(0.064)	0.98	(0.040)
β_{11}	0.50	(0.028)	0.50	(0.029)	0.50	(0.019)
β_{12}	-0.51	(0.055)	-0.51	(0.044)	-0.51	(0.026)
β_{20}	1.06	(0.11)	1.05	(0.071)	1.08	(0.055)
β_{21}	-1.05	(0.096)	-1.02	(0.065)	-1.08	(0.059)
β_{22}	1.07	(0.11)	1.04	(0.076)	1.09	(0.065)
σ_1	1.92	(0.22)	2.03	(0.19)	2.05	(0.096)
σ_{12}	0.95	(0.22)	1.18	(0.22)	0.97	(0.053)
α_{00}			1		1	
α_{01}			0		0.012	(0.019)
α_{02}					-0.078	(0.062)
α_{10}			0		0.0025	(0.018)
α_{11}			-0.12	(0.082)	-0.014	(0.037)
α_{12}					0.0099	(0.028)
α_{20}					-0.047	(0.012)
α_{21}					-0.0068	(0.013)
α_{22}					0.013	(0.0049)
$\hat{\sigma}_1$			1.88	(0.18)	2.00	(0.098)
$\hat{\sigma}_2$			0.93	(0.059)	1.00	(0.030)
$\hat{\sigma}_{12}$			0.72	(0.32)	0.94	(0.16)
loglikelihood	-1371.30		-1348.14		-1331.18	
rejections			69		97	

The standard errors are given in parentheses.

Table 4: Results of the simulation experiment with bivariate t distributed disturbances, 100 replications, $N = 1000$.

$N = 500$	normal		generalized snp			
			$K = 1$		$K = 2$	
β_{10}	0.99	(0.14)	0.96	(0.16)	0.97	(0.15)
β_{11}	0.49	(0.040)	0.50	(0.039)	0.50	(0.035)
β_{12}	-0.49	(0.081)	-0.50	(0.077)	-0.49	(0.071)
β_{20}	1.04	(0.12)	1.03	(0.11)	1.09	(0.15)
β_{21}	-1.03	(0.14)	-1.04	(0.14)	-1.06	(0.15)
β_{22}	1.03	(0.13)	1.04	(0.13)	1.06	(0.14)
σ_1	2.06	(0.13)	2.03	(0.12)	1.93	(0.23)
σ_{12}	1.20	(0.27)	1.24	(0.27)	1.17	(0.33)
α_{00}			1		1	
α_{01}			0		-0.074	(0.10)
α_{02}					-0.044	(0.12)
α_{10}			0		-0.060	(0.059)
α_{11}			-0.031	(0.058)	0.021	(0.10)
α_{12}					-0.034	(0.075)
α_{20}					-0.022	(0.060)
α_{21}					0.032	(0.043)
α_{22}					-0.0010	(0.012)
$\hat{\sigma}_1$			1.97	(0.12)	1.78	(0.27)
$\hat{\sigma}_2$			0.97	(0.045)	0.86	(0.092)
$\hat{\sigma}_{12}$			1.08	(0.33)	0.82	(0.38)
loglikelihood	-693.32		-684.63		-665.48	
rejections			73		100	

The standard errors are given in parentheses.

Table 5: Results of the simulation experiment with bivariate χ^2 distributed disturbances, 100 replications, $N = 500$.

	normal		generalized snp			
$N = 1000$			$K = 1$		$K = 2$	
β_{10}	0.94	(0.089)	0.93	(0.047)	0.95	(0.061)
β_{11}	0.50	(0.029)	0.50	(0.027)	0.50	(0.018)
β_{12}	-0.50	(0.054)	-0.50	(0.036)	-0.49	(0.021)
β_{20}	1.09	(0.075)	1.05	(0.039)	1.12	(0.041)
β_{21}	-1.04	(0.076)	-1.04	(0.037)	-1.06	(0.028)
β_{22}	1.03	(0.083)	1.04	(0.037)	1.06	(0.024)
σ_1	2.03	(0.084)	2.01	(0.053)	1.89	(0.045)
σ_{12}	1.28	(0.15)	1.27	(0.053)	1.19	(0.020)
α_{00}			1		1	
α_{01}			0		-0.11	(0.021)
α_{02}					-0.039	(0.030)
α_{10}			0		-0.088	(0.014)
α_{11}			-0.018	(0.032)	0.066	(0.031)
α_{12}					-0.0067	(0.032)
α_{20}					-0.046	(0.019)
α_{21}					0.037	(0.017)
α_{22}					0.0034	(0.0037)
$\hat{\sigma}_1$			1.94	(0.074)	1.61	(0.042)
$\hat{\sigma}_2$			0.97	(0.038)	0.85	(0.012)
$\hat{\sigma}_{12}$			1.12	(0.17)	0.56	(0.066)
loglikelihood	-1403.24		-1391.81		-1345.13	
rejections			89		100	

The standard errors are given in parentheses.

Table 6: Results of the simulation experiment with bivariate χ^2 distributed disturbances, 100 replications, $N = 1000$.

Car owner	yes	no
Gender		
Male	1523	1102
Female	1587	2306
Age		
18–24	106	518
25–29	416	415
30–39	746	485
40–49	467	296
50–64	692	451
65+	683	1243
Region		
West	1588	1881
North	211	207
East	895	959
South	416	361
Degree of urbanization		
Very high	764	1329
High	786	970
Average	694	497
Low	563	397
Very low	303	215

Car owner	yes	no
Net income (in guilders)		
0–15000	158	999
15000–23000	411	1046
23000–30000	539	574
30000–38000	695	402
38000–52000	773	278
52000+	534	109
Level of education		
Primary	237	688
Lower secondary	712	856
Higher secondary	1022	1075
University	1139	789
Labor market status		
Full-time work	1626	771
Part-time work	178	245
Student	37	406
Unemployed	105	250
Nonparticipant	1164	1736
Free public transport pass		
no	3074	2965
yes	36	443
Average number of kilometers		
	15093	
Observations	3110	3408

Table 7: Some characteristics of the data set.

	Distance function β		Costs function γ	
Intercept	9.41	(0.32)	10.81	(0.35)
Gender				
Female	−0.29	(0.038)	0.046	(0.053)
Age				
25–29	−0.13	(0.088)	−0.49	(0.13)
30–39	−0.24	(0.084)	−0.56	(0.13)
40–49	−0.28	(0.089)	−0.65	(0.13)
50–64	−0.33	(0.10)	−1.00	(0.15)
65+	−0.59	(0.086)	−0.76	(0.14)
Region				
North	0.0027	(0.050)	−0.11	(0.092)
East	−0.0007	(0.031)	0.049	(0.058)
South	−0.054	(0.040)	−0.16	(0.072)
Degree of urbanization				
High	0.025	(0.042)	−0.18	(0.068)
Average	0.064	(0.066)	−0.51	(0.086)
Low	0.077	(0.073)	−0.57	(0.096)
Very low	0.064	(0.082)	−0.65	(0.11)
Net income (in guilders)				
15000–23000	0.086	(0.075)	−0.21	(0.10)
23000–30000	0.23	(0.11)	−0.48	(0.14)
30000–38000	0.30	(0.14)	−0.72	(0.16)
38000–52000	0.37	(0.16)	−0.90	(0.18)
52000+	0.50	(0.18)	−1.04	(0.20)

Table 8: Estimation results for the model with normal distributed disturbance terms (continued).

	Distance function β		Costs function γ	
Level of education				
Lower secondary	0.10	(0.051)	−0.16	(0.079)
Higher secondary	0.17	(0.062)	−0.26	(0.089)
University	0.28	(0.064)	−0.15	(0.095)
Labor market status				
Part-time work	0.017	(0.058)	0.15	(0.095)
Student	−0.28	(0.11)	0.14	(0.19)
Unemployed	−0.29	(0.069)	0.058	(0.11)
Nonparticipant	−0.25	(0.046)	−0.023	(0.082)
Free public transport pass				
yes			0.21	(0.16)
σ_1	0.63	(0.0088)		
σ_{12}	0.041	(0.17)		
loglikelihood	-6386.34			

Table 8: Estimation results for the model with normal distributed disturbance terms.

	Distance function β		Costs function γ	
Intercept	9.77	(0.11)	10.86	(0.14)
Gender				
Female	−0.24	(0.022)	0.017	(0.038)
Age				
25–29	−0.14	(0.061)	−0.41	(0.10)
30–39	−0.23	(0.059)	−0.46	(0.10)
40–49	−0.30	(0.062)	−0.58	(0.11)
50–64	−0.35	(0.064)	−0.85	(0.11)
65+	−0.50	(0.067)	−0.61	(0.12)
Region				
North	0.0066	(0.045)	−0.075	(0.079)
East	0.013	(0.027)	0.060	(0.050)
South	−0.057	(0.034)	−0.13	(0.062)
Degree of urbanization				
High	−0.0070	(0.032)	−0.18	(0.056)
Average	−0.0088	(0.035)	−0.47	(0.057)
Low	−0.0043	(0.038)	−0.53	(0.064)
Very low	−0.011	(0.044)	−0.59	(0.078)
Net income (in guilders)				
15000–23000	−0.012	(0.061)	−0.26	(0.085)
23000–30000	0.089	(0.065)	−0.48	(0.089)
30000–38000	0.12	(0.066)	−0.68	(0.090)
38000–52000	0.17	(0.069)	−0.83	(0.093)
52000+	0.30	(0.072)	−0.90	(0.10)

Standard errors in parentheses.

Table 9: Estimation results for the model with generalized snp-distributed disturbances ($K = 1$) (continued).

	Distance function β		Costs function γ	
Level of education				
Lower secondary	0.043	(0.040)	−0.16	(0.066)
Higher secondary	0.097	(0.041)	−0.24	(0.068)
University	0.21	(0.043)	−0.12	(0.075)
Labor market status				
Part-time work	0.023	(0.050)	0.13	(0.080)
Student	−0.19	(0.10)	0.13	(0.16)
Unemployed	−0.18	(0.059)	0.11	(0.095)
Nonparticipant	−0.23	(0.037)	−0.054	(0.068)
Free public transport pass				
yes			0.14	(0.12)
σ_1	0.71	(0.018)		
σ_{12}	−0.45	(0.034)		
α_{00}	1			
α_{01}	0			
α_{10}	0			
α_{11}	0.49	(0.031)		
loglikelihood	-6244.49			

Standard errors in parentheses.

Table 9: Estimation results for the model with generalized snp-distributed disturbances ($K = 1$).

	Distance function β		Costs function γ	
Intercept	9.87	(0.15)	10.91	(0.29)
Gender				
Female	−0.23	(0.026)	0.047	(0.095)
Age				
25–29	−0.15	(0.066)	−0.40	(0.13)
30–39	−0.23	(0.064)	−0.46	(0.12)
40–49	−0.30	(0.067)	−0.58	(0.13)
50–64	−0.36	(0.071)	−0.86	(0.18)
65+	−0.50	(0.073)	−0.59	(0.13)
Region				
North	0.0010	(0.047)	−0.077	(0.088)
East	0.016	(0.028)	0.065	(0.055)
South	−0.061	(0.034)	−0.13	(0.070)
Degree of urbanization				
High	−0.012	(0.033)	−0.18	(0.076)
Average	−0.035	(0.039)	−0.49	(0.15)
Low	−0.033	(0.043)	−0.56	(0.17)
Very low	−0.037	(0.050)	−0.61	(0.19)
Net income (in guilders)				
15000–23000	0.022	(0.067)	−0.20	(0.11)
23000–30000	0.093	(0.075)	−0.46	(0.17)
30000–38000	0.11	(0.081)	−0.69	(0.24)
38000–52000	0.15	(0.086)	−0.86	(0.30)
52000+	0.27	(0.090)	−0.95	(0.37)

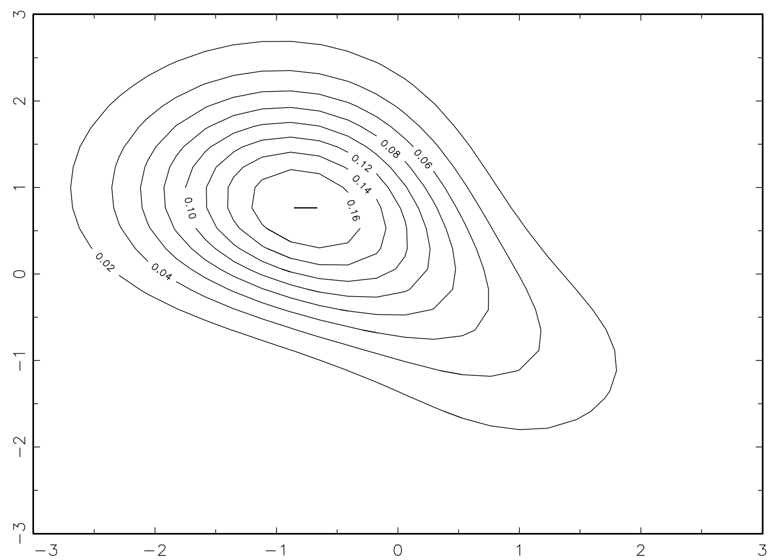
Standard errors in parentheses.

Table 10: Estimation results for the model with generalized snp-distributed disturbances ($K = 2$) (continued).

	Distance function β		Costs function γ	
Intercept	9.87	(0.15)	10.91	(0.29)
Level of education				
Lower secondary	0.035	(0.044)	−0.16	(0.089)
Higher secondary	0.085	(0.046)	−0.25	(0.12)
University	0.20	(0.047)	−0.13	(0.12)
Labor market status				
Part-time work	0.035	(0.052)	0.15	(0.094)
Student	−0.16	(0.12)	0.16	(0.19)
Unemployed	−0.17	(0.063)	0.13	(0.14)
Nonparticipant	−0.22	(0.039)	−0.040	(0.094)
Free public transport pass				
yes			0.15	(0.13)
σ_1	0.76	(0.23)		
σ_{12}	−0.54	(0.31)		
α_{00}	1			
α_{01}	−0.0035	(0.049)		
α_{02}	0.038	(0.12)		
α_{10}	−0.012	(0.045)		
α_{11}	0.38	(0.41)		
α_{12}	0.15	(0.12)		
α_{20}	−0.052	(0.14)		
α_{21}	0.16	(0.079)		
α_{22}	0.0005	(0.098)		
loglikelihood	-6236.54			

Standard errors in parentheses.

Table 10: Estimation results for the model with generalized snp-distributed disturbances ($K = 2$).



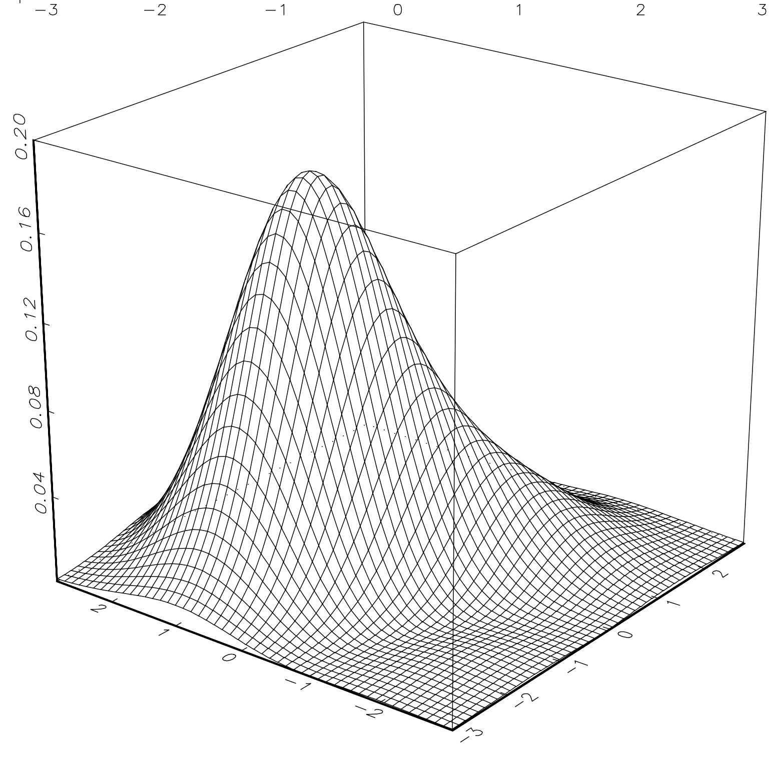


Figure 1: Bivariate snp-density, $\alpha_{01} = 0.1$, $\alpha_{10} = -0.1$ and $\alpha_{11} = -0.2$

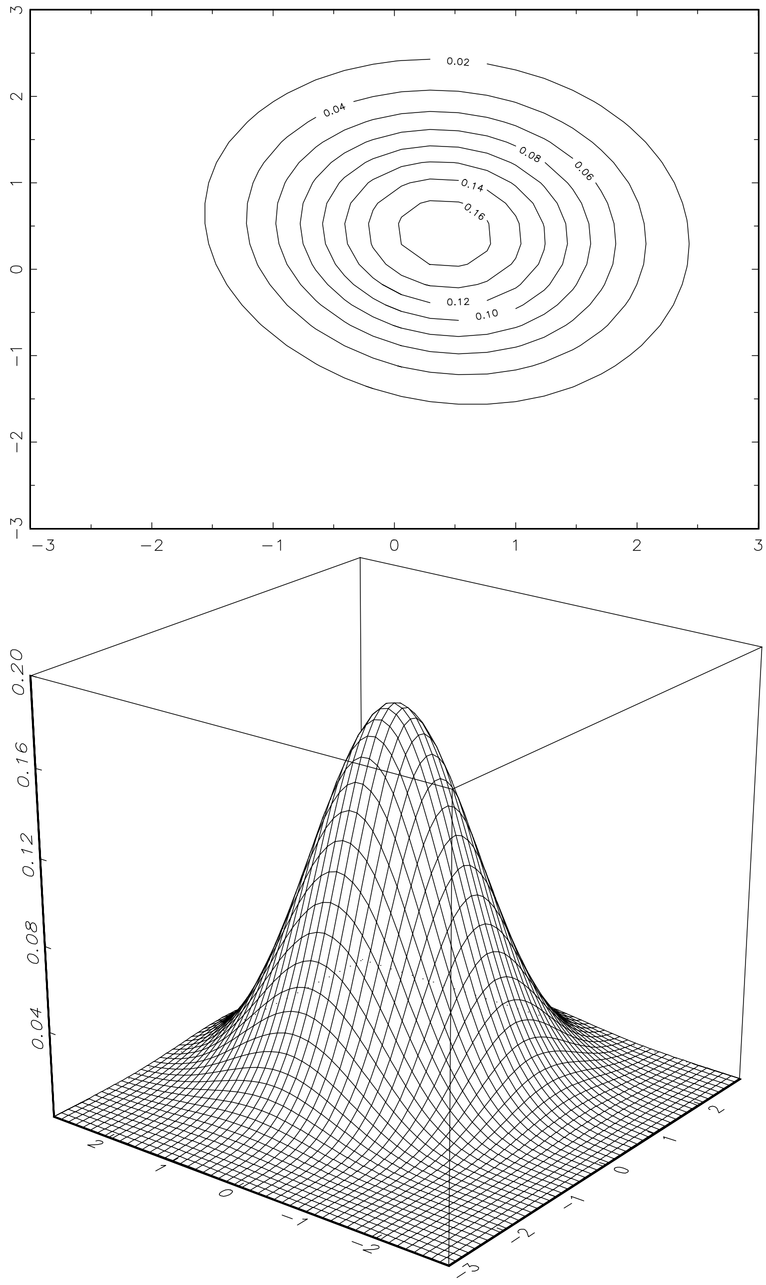


Figure 2: Bivariate snp-density, $\alpha_{01} = 0.1$, $\alpha_{10} = 0.1$ and $\alpha_{11} = 0$

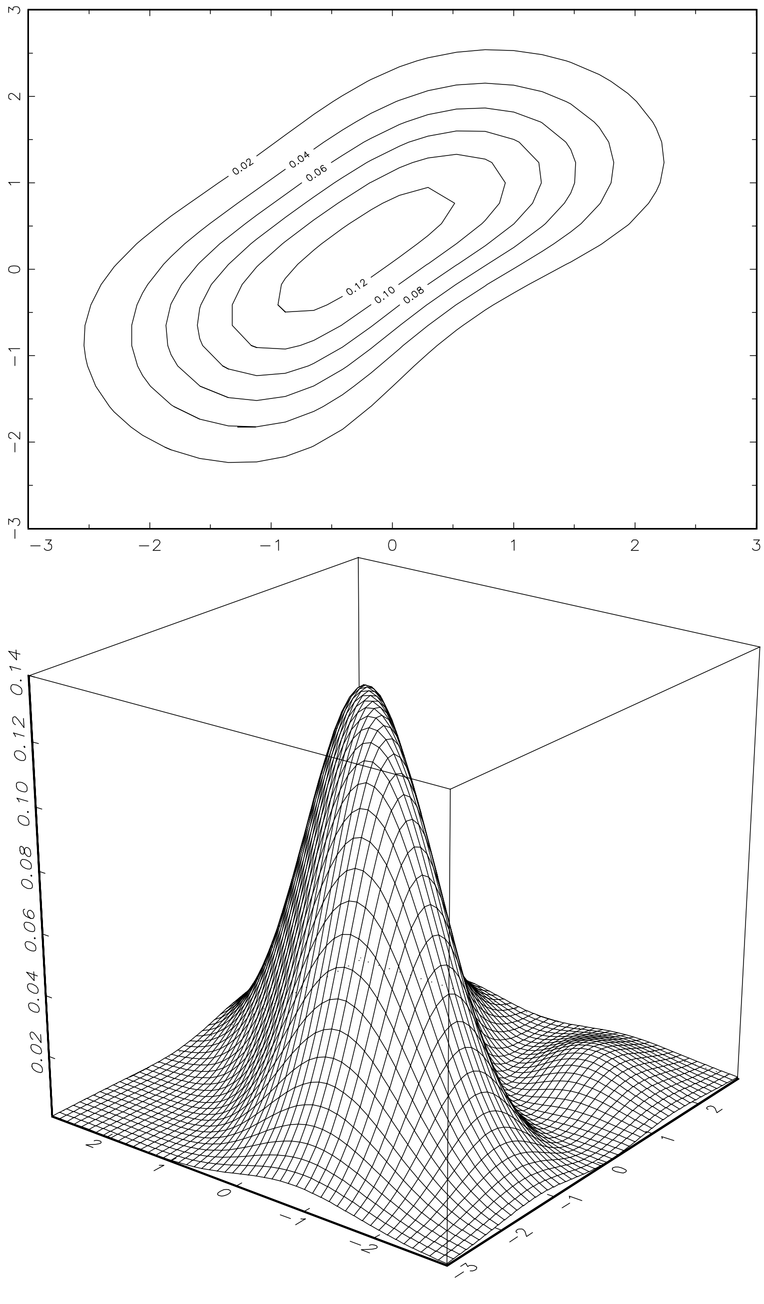


Figure 3: Bivariate snp-density, $\alpha_{01} = 0.1$, $\alpha_{10} = -0.1$ and $\alpha_{11} = 0.2$

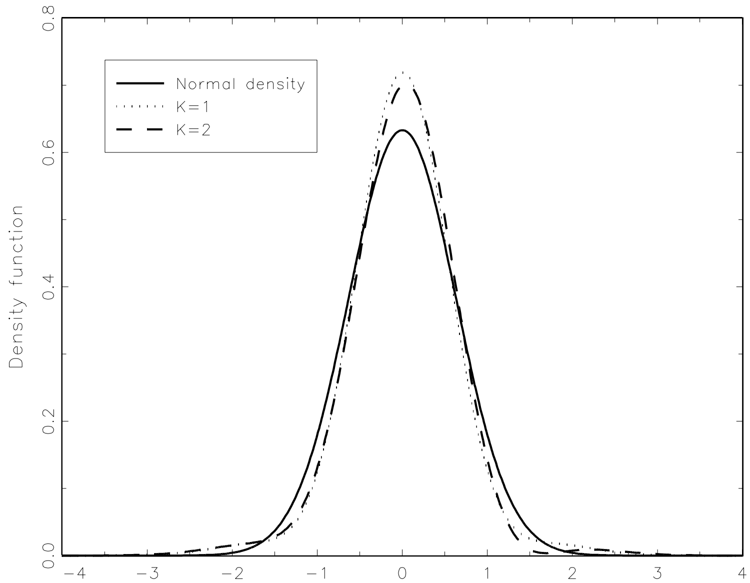


Figure 4: The estimated marginal density function of ε_1.

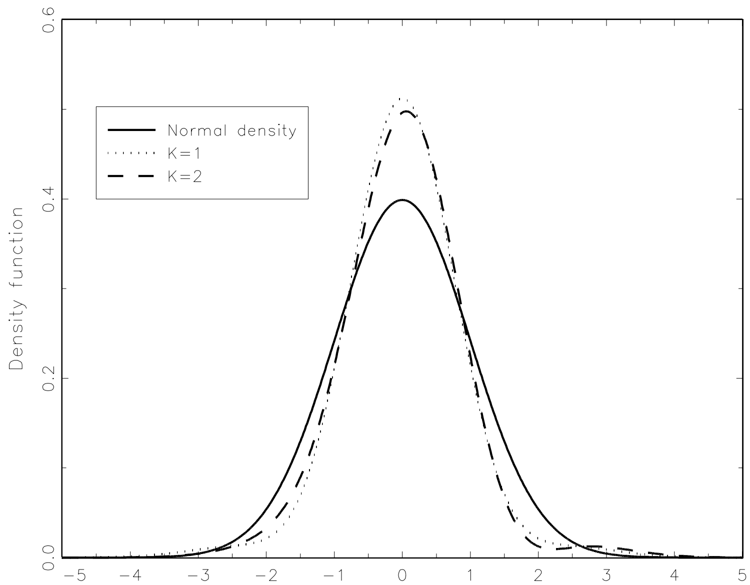


Figure 5: The estimated marginal density function of ε_2.

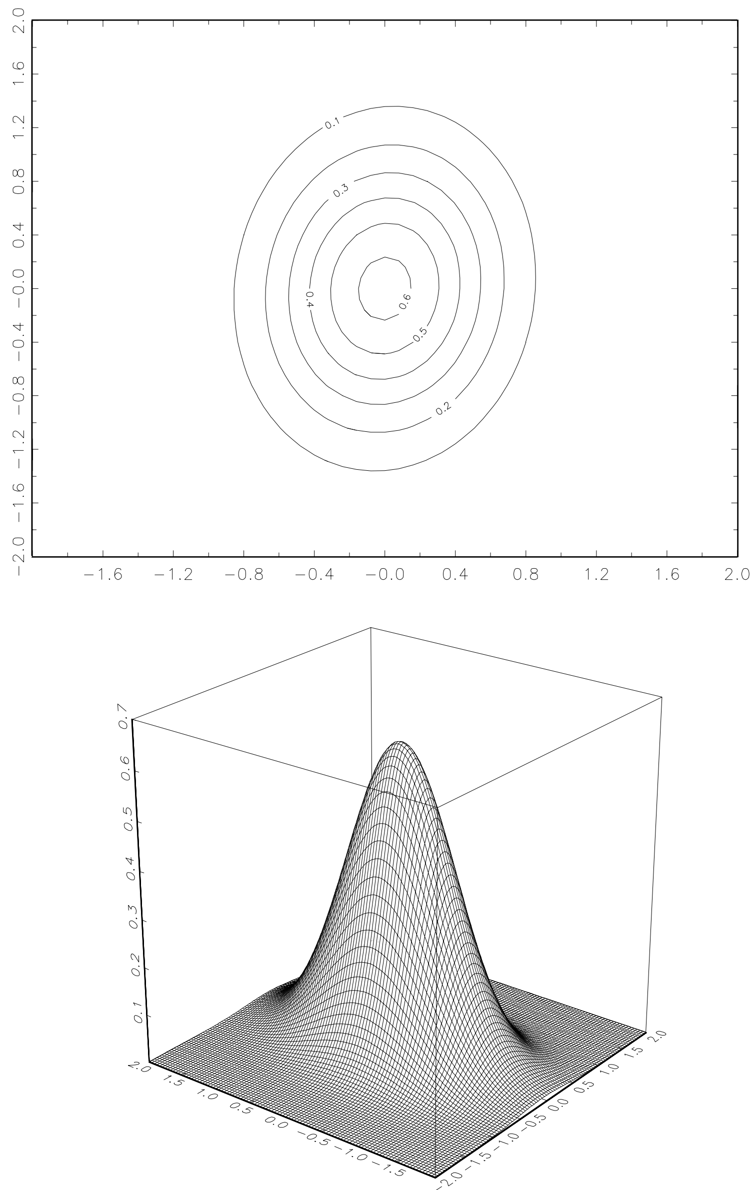


Figure 6: Estimated density under normality assumption (Table 8).

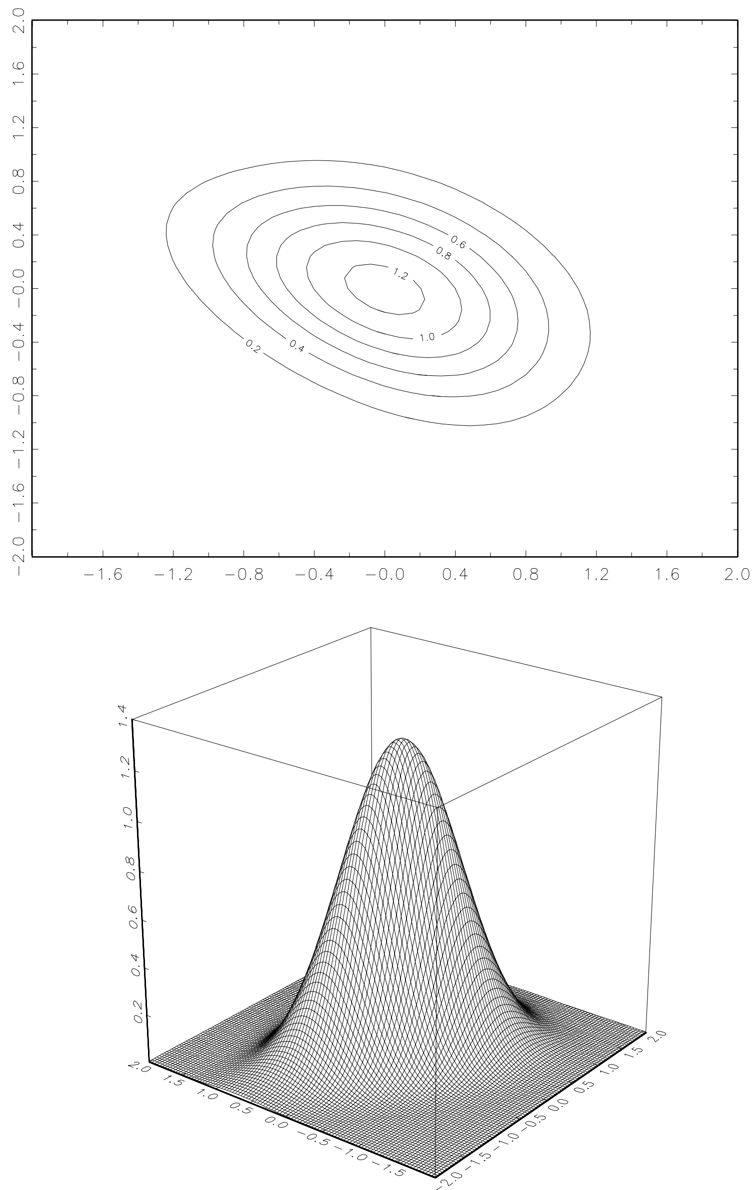


Figure 7: Estimated density $K = 2$ (Table 10).