Some Applications of Semi-Nonparametric Maximum Likelihood Estimation[1] (Tinbergen Institute discussion paper 96-161/7)

Bas van der Klaauw[2] Ruud H. Koning[3]

November 12, 1996

[1]This paper was written while the second author enjoyed the hospitality of the Department of Economics of Brown University. The authors thank Geert Ridder, Bo Honoré, and participants at the 1996 ESRC Econometric Study Group Conference in Bristol and participants at the ESEM 1996 in Istanbul for comments.

[2]Tinbergen Institute, Keizersgracht 482, 1017 EG Amsterdam, The Netherlands, email `klaauw@tinbinst.nl`.

[3]Department of Econometrics, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, and Tinbergen Institute, Keizersgracht 482, 1017 EG Amsterdam, The Netherlands, email `rkoning@econ.vu.nl`.

Abstract

An alternative to estimation of microeconometric models under the assumption of normality of the distribution of the disturbances is semi-nonparametric maximum likelihood estimation. In a particular class of this kind of models, the density function of the disturbances is approximated by a Hermite series. In this paper we will discuss this approach in the context of a popular microeconometric model (the sample selection model) and we apply the model to a truncated switching regression model with endogenous regimes. A new choice of base functions of the Hermite series is presented and the semi-nonparametric approach is used to examine sensitivity to the assumption of normality of estimation results of a model for rent assistance and housing demand in Koning (1995).

1 Introduction

Maximum likelihood is the most popular estimation method in microeconometrics. 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 misspecification of the functional form) 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 both 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.

Starting with Manski (1975), semi-parametric methods have been proposed for specific microeconometric models. These models do not require complete distributional assumptions or less restrictive distributional assumptions than the assumption of normality (for example, the conditional mean of the distribution of the disturbances is 0). These methods yield consistent estimates of the parameters of interest without a complete specification of the distribution of the stochastic variables in the model. A recent survey of methods available is Powell (1994). Recently, Gallant and Nychka (1987) have introduced a semi-nonparametric estimation method that estimates the density function of the disturbances along with the other parameters of the model. An advantage of this method is that it is of general applicability, it is not specific to one particular model. The basic idea is to approximate the unknown density function by a Hermite series. A requirement for consistency of the method is that the number of terms in the series increases with the sample size, and hence, the approximation becomes better as the sample size increases. To our knowledge, this method has been applied only a few times so experience with the method is limited. The aim of this paper is to document some simulation experience with the method. The method is also applied to a specific application. In Koning and Ridder (1993) a model for rent assistance and housing demand is estimated under the assumption of normality of the stochastic variables in the model. In this paper we will examine the sensitivity of these results to the normality assumption

The setup of this paper is as follows. In section 2 we give an introduction to the semi-nonparametric maximum likelihood estimation method of Gallant and Nychka (1987). The Hermite form density is compared to the (multivariate) normal density function. In section 3 we deal with the sample selection model and some simulation results are presented there. In section 4 we estimate the housing demand model of Koning and Ridder (1993) and Koning (1995) semi-nonparametrically. This model is a truncated switching regression model with endogenous regimes and the method has not been applied to such a model before. We end with some concluding remarks concerning our experience with the Gallant-Nychka method. Technical matters and derivations are relegated to Appendices.

2 Semi-nonparametric Maximum Likelihood Estimation

In this section we discuss the semi-nonparametric maximum likelihood method introduced by Gallant and Nychka (1987). The estimation method is based on the approximation of the (unknown) density function by a Hermite series. In the first part of this section we recapitulate the estimation approach of Gallant and Nychka (1987) and in the second part we compare Hermite series with bivariate normal distributions.

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

$$h(\varepsilon) = \frac{P^2(\varepsilon)}{Q^2(\varepsilon)}\phi^2(\varepsilon \mid \tau, \Sigma) \tag{1}$$

can approximate any density function arbitrarily well. In equation (1), $P(\varepsilon)$ and $Q(\varepsilon)$ are polynomials and $\phi(\varepsilon \mid \tau, \Sigma)$ is the multivariate normal density function with mean τ and covariance matrix Σ. Of course, equation (1) is not a proper density function if the polynomials P and Q are not restricted such that h integrates to 1. Gallant and Nychka restrict the density $h(\varepsilon)$ to a subclass $\mathcal{H}_K$ which consists of densities of the Hermite form

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

with Δ a diagonal matrix. $P_K(\cdot)$ is a polynomial of degree K. Gallant and Nychka 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. The true density function is assumed to be a member of the class $\mathcal{H}$. Conditions defining $\mathcal{H}$ precisely are given in Gallant and Nychka. 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 $\mathcal{H}$. Of course, 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 equation (2), 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.

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

$$\begin{aligned} h^*(\varepsilon) &= \left(\sum_{i_1,\dots,i_n=0}^{K} \alpha_{i_1\cdots i_n}(\varepsilon_1-\tau_1)^{i_1}\cdots(\varepsilon_n-\tau_n)^{i_n} \right)^2 \\ &\qquad \times \exp\left(-\left[(\varepsilon_1-\tau_1)^2/\delta_1^2+\cdots+(\varepsilon_n-\tau_n)^2/\delta_n^2\right]\right) \\ &= \sum_{i_1,\dots,i_n,j_1,\dots,j_n=0}^{K} \alpha_{i_1\cdots i_n}\alpha_{j_1\cdots j_n}(\varepsilon_1-\tau_1)^{i_1+j_1}\cdots(\varepsilon_n-\tau_n)^{i_n+j_n} \\ &\qquad \times \exp\left(-\left[(\varepsilon_1-\tau_1)^2/\delta_1^2+\cdots+(\varepsilon_n-\tau_n)^2/\delta_n^2\right]\right) \end{aligned} \tag{3}$$

Because of the squaring in equation (3), 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, some restrictions 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) by scaling the density. Define S by

$$\begin{aligned} S = \int_{\mathbb{R}^n} &\sum_{i_1,\dots,i_n,j_1,\dots,j_n=0}^{K} \alpha_{i_1\cdots i_n}\alpha_{j_1\cdots j_n}(\varepsilon_1-\tau_1)^{i_1+j_1}\cdots(\varepsilon_n-\tau_n)^{i_n+j_n} \\ &\times \exp\left(-\left[(\varepsilon_1-\tau_1)^2/\delta_1^2+\cdots+(\varepsilon_n-\tau_n)^2/\delta_n^2\right]\right) d\varepsilon_1\cdots d\varepsilon_n \end{aligned} \tag{4}$$

Now the following scaled density integrates to 1 by the definition of S:

$$h(\varepsilon) = h^*(\varepsilon)/S. \tag{5}$$

We will refer to densities of the type (5) as snp-densities. It is clear that α in equation (5) is identified up to a scale only, so a normalization is necessary. In particular applications, additional restrictions will be needed to achieve identification, see below. For most applications it will be convenient to set τ to 0 which we will do from now on (unless stated otherwise).

Because the normal density serves as a benchmark in most microeconometric applications, we will compare the Hermite form equation (5) with the normal density. First, we consider the univariate case ($n = 1$). We normalize the vector with α-coefficients by setting $\alpha_0 = (2\pi)^{-1/4}$ it is clear that the $h(\varepsilon)$ reduces to the normal density function if the other α's are 0. If we want to calculate the moments of the Hermite form density it is convenient to introduce some additional notation. Let the $(K+1) \times (K+1)$ matrix $Q^{(l)}$ be defined by its typical element[1]:

$$Q_{ij}^{(l)} = \int_{-\infty}^{\infty} \varepsilon^{i+j+l-2} \exp(-\varepsilon^2/\delta^2) d\varepsilon$$

In this notation, the scalar S in equation (4) is equal to $S = \alpha' Q^{(0)} \alpha$. The elements of $Q^{(l)}$ can be determined explicitly using the recursion formulae in Appendix A. In this notation, characteristics of the Hermite form density are:

$$\mathcal{E}\varepsilon = \alpha' Q^{(1)} \alpha / S \tag{6}$$

$$\mathcal{E}\varepsilon^2 = \alpha' Q^{(2)} \alpha / S \tag{7}$$

$$\mathcal{E}\varepsilon^3 = \alpha' Q^{(3)} \alpha / S \tag{8}$$

$$\mathcal{E}\varepsilon^4 = \alpha' Q^{(4)} \alpha / S \tag{9}$$

The main diagonal of $Q^{(l)}$ consists of zeros if l is odd and hence it is possible to impose restrictions on the density such that the mean is equal to 0, etc. Using the recursion formula given in Appendix A it is easy to derive explicit expressions for the moments of the snp-density. Even though they are not particularly insightful, one sees that if $K \geq 2$ the coefficients of skewness and kurtosis[2] are no longer restricted to 0 and 3 as is the case of the normal distribution. In figure 1 and figure 2 we plot the snp-density with $K = 2$ and $K = 3$. α_0 is set to $(2\pi)^{-1/4}$ and α_2 is set to $-\alpha_0/3$ ($K = 2$) and $-\frac{6\alpha_0\alpha_3 + 30\alpha_2\alpha_3}{2\alpha_0 + 6\alpha_2}$ ($K = 3$) so that the mean of ε is 0 in all cases.

The snp-density is difficult to characterize for the bivariate ($n = 2$) case. From equation (4) it is clear that the snp-density reduces to a bivariate normal density with mean 0 and covariance matrix $\begin{pmatrix} 2\delta_1^2 & 0 \\ 0 & 2\delta_2^2 \end{pmatrix}$ if $\alpha_{00} = 1$ and all other α-parameters are 0. We graph the density surface and some contour lines for the case $K = 1$ for some different values of α_{10}, α_{01}, and α_{11}[3] in figures 3 to 5. From these graphs it is clear that a wide variety of densities can be generated by varying the α-parameters even if K is as low as 1. In interesting question is whether a bivariate normal density with unrestricted covariance is a special case of a bivariate snp-density with $K = 1$. It turns out that given the covariance matrix of a normal density one is able to choose the α-parameters such that the bivariate snp-density has identical first and second moments. However, the form of the marginal distributions of this snp-density differs markedly from the marginal normal distribution[4]. Hence, the bivariate snp-density (5) is not particularly suitable for testing the normality assumption: either the covariance or the marginal distributions are misspecified if the true density is bivariate normal and $K = 1$.

[1] A similar matrix is defined in Gabler, Laisney, and Lechner (1993), p. 64, but their definition contains an error. The exponent of ε in their paper reads $i + j$ instead of $i + j - 2$.

[2] Let μ_i be the i-th central moment. Then these coefficients are defined as $\left(\frac{\mu_3}{\sigma^2}\right)^2$ and $\frac{\mu_4}{\sigma^4}$ respectively.

[3] Because the scale of the α's is not determined, we set $\alpha_{00} = 1$.

[4] We had to rely on numerical comparisons as analytical solutions to the equations equating the moments of both distributions were impossible to find.

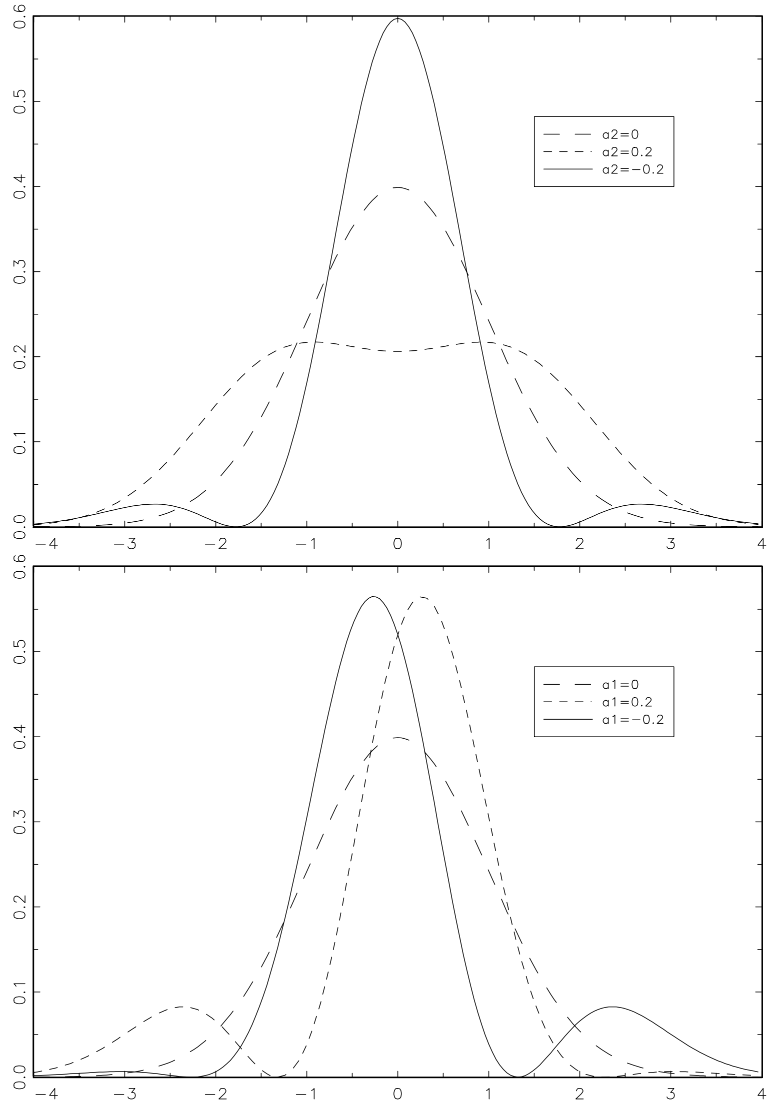

Figure 1: Univariate snp-density, $K = 2$

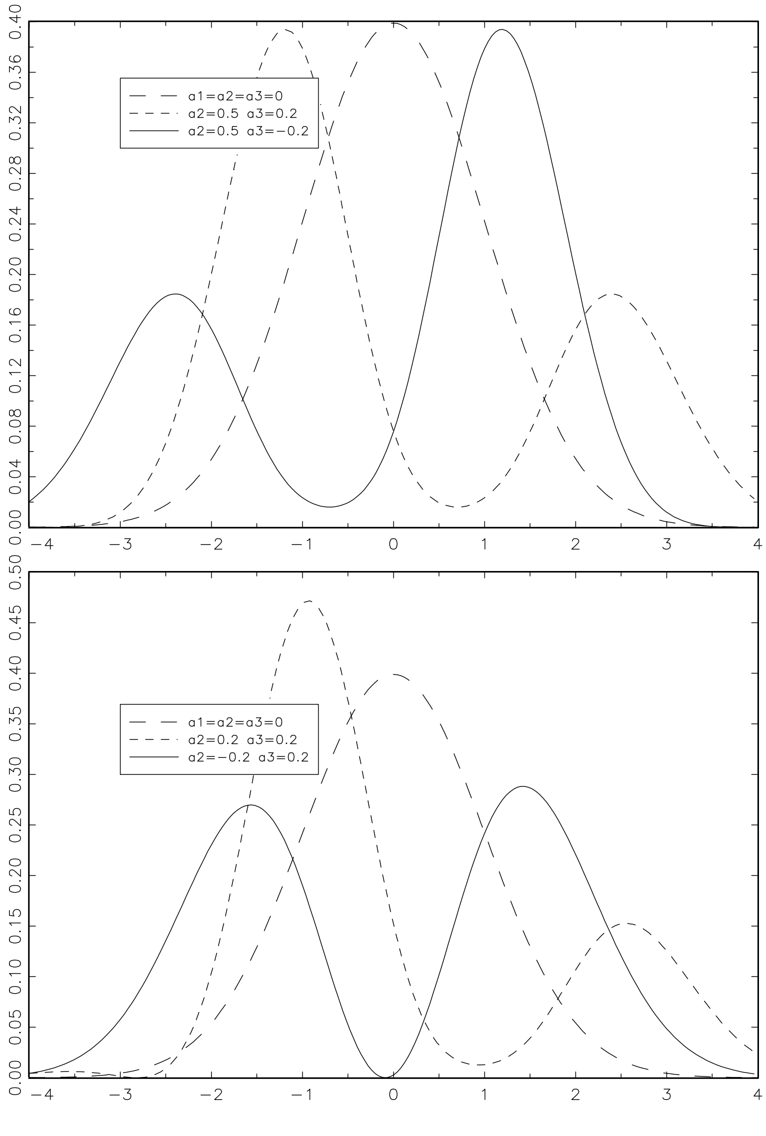

Figure 2: Univariate snp-density, $K = 3$

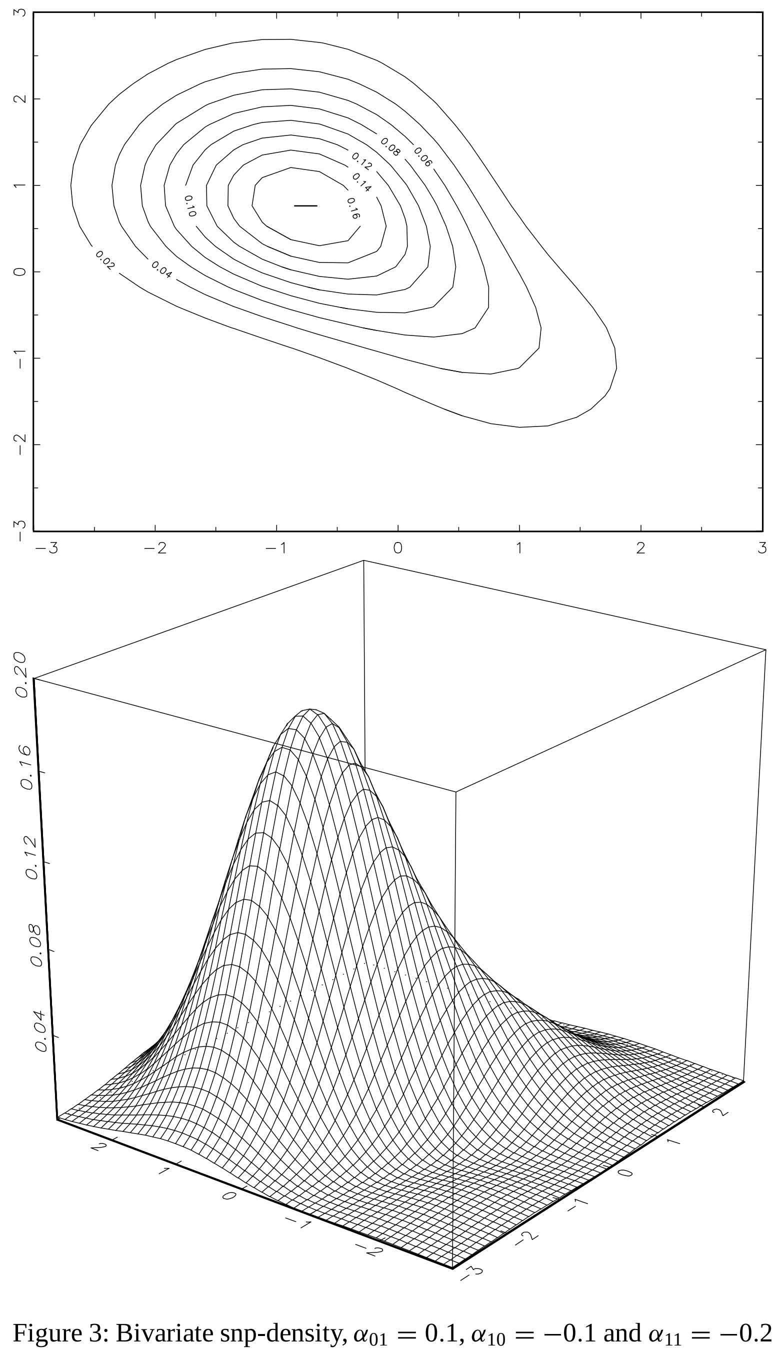

Figure 3: Bivariate snp-density, $\alpha_{01} = 0.1$, $\alpha_{10} = -0.1$ and $\alpha_{11} = -0.2$

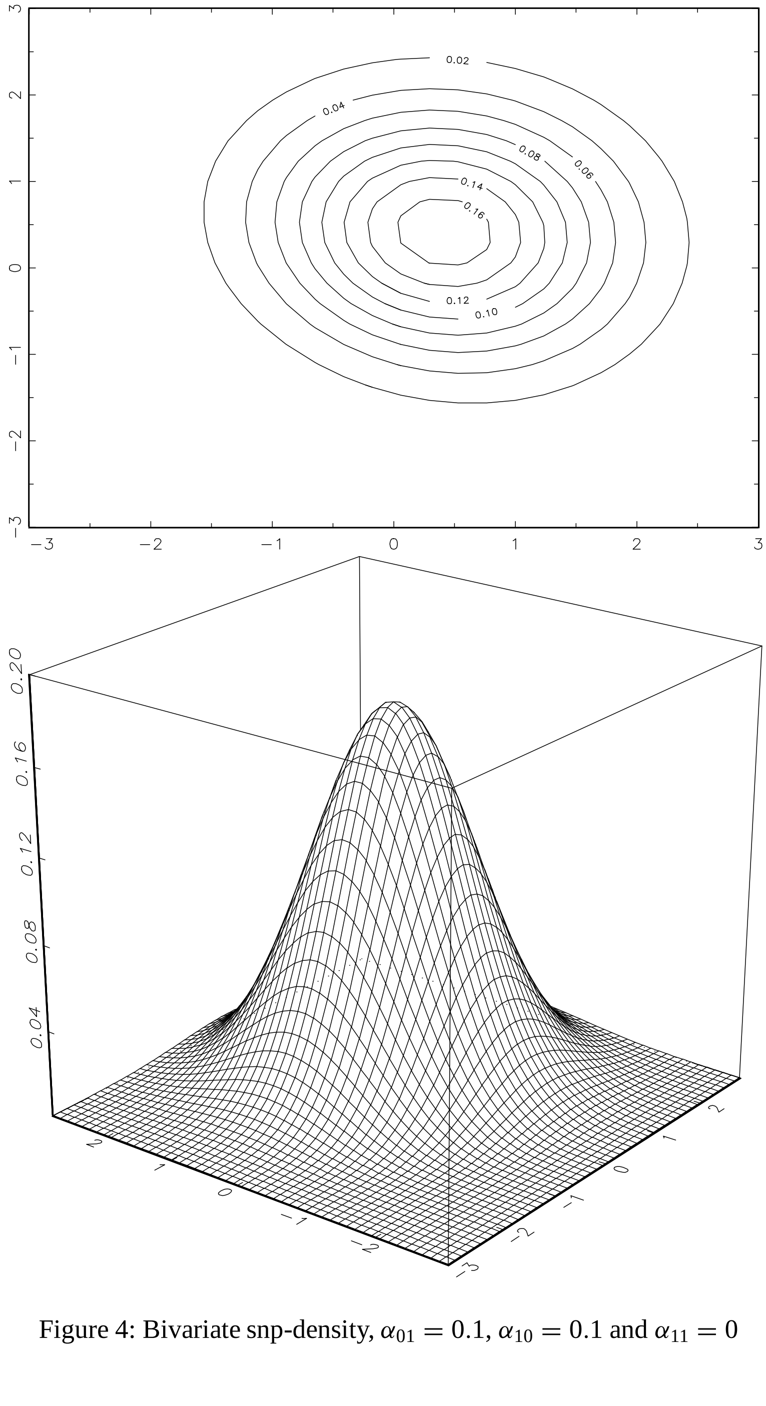

Figure 4: Bivariate snp-density, $\alpha_{01} = 0.1$, $\alpha_{10} = 0.1$ and $\alpha_{11} = 0$

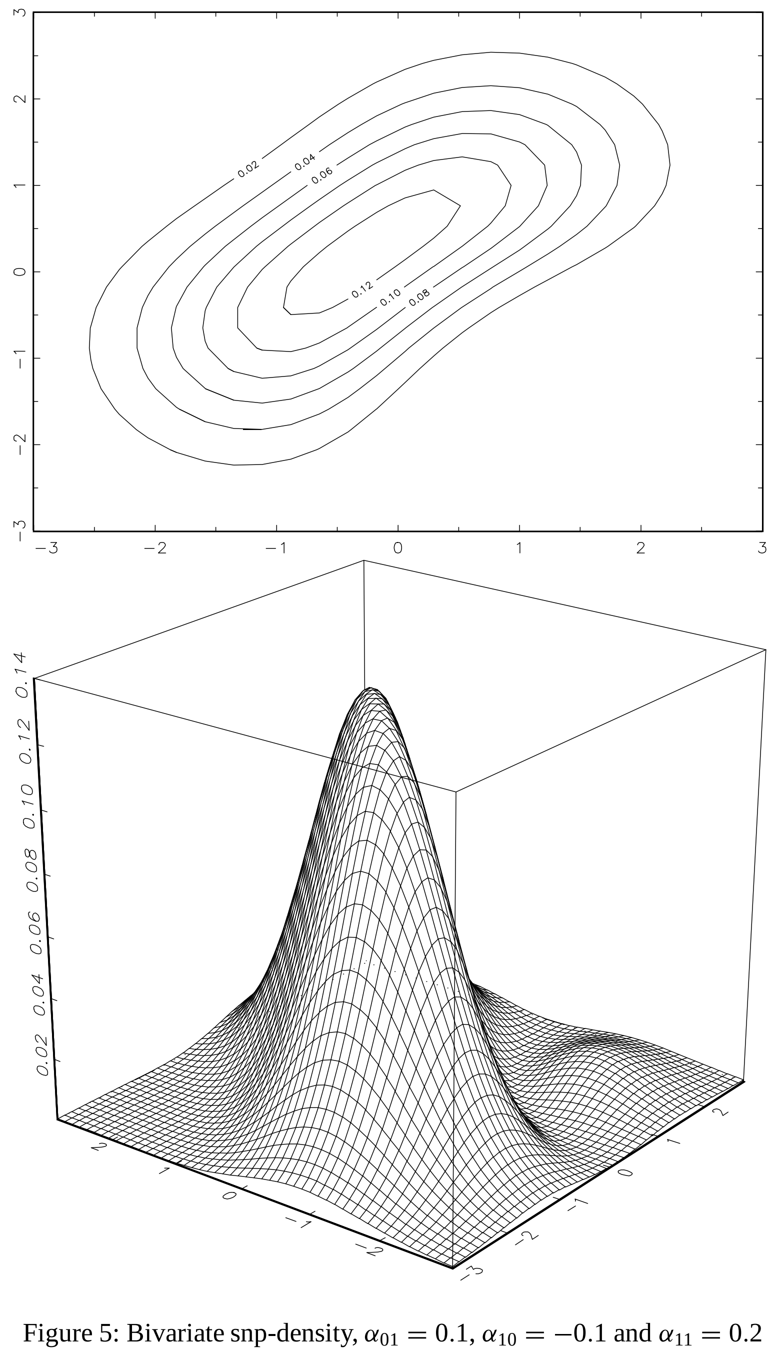

Figure 5: Bivariate snp-density, $\alpha_{01} = 0.1$, $\alpha_{10} = -0.1$ and $\alpha_{11} = 0.2$

3 Semi-nonparametric Estimation of the Sample Selection Model

In this section we consider snp-estimation of the sample selection model, introduced by Heckman (1979). It is also referred to as a Type II Tobit model (Amemiya (1985)). The sample selection model is a two equations model. The first equation is a regression equation

$$y_t = \beta_1' x_{1t} + \varepsilon_{1t}. \tag{10}$$

However, we observe this equation only for a selected sample. The selection rule is given by

$$\begin{aligned} I_t^* &= \beta_2' x_{2t} + \varepsilon_{2t} \\ I_t &= \begin{cases} 1 & I_t^* > 0 \\ 0 & I_t^* \le 0 \end{cases}. \end{aligned} \tag{11}$$

The observations in equation (10) are observed for those with $I_t = 1$ only. If the conditional expectation of ε_1 given $I_t = 1$ is not equal to 0, OLS-estimation of equation (10) will not yield unbiased estimates for β_1. For each observation in the sample we observe the exogenous variables $x_t = (x_{1t}', x_{2t}')'$, and I_t. The outcome of the regression equation (10) is observed only if $I_t = 1$.

If one is willing to assume that $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ follows a bivariate normal distribution, one can estimate the parameters of the sample selection model either using Heckman's two-stage procedure (Heckman (1979)) or by full information maximum likelihood. However, according to Greene (1993) these estimates are rather sensitive to the distributional assumption so one would either like to test this distributional assumption or compare the results obtained under the assumption of normality with other, semi-nonparametric results.

The loglikelihoodfunction for the sample selection model is

$$\begin{aligned} \ell(\theta) &= \sum_{I_t=1} \ln \left(\int_{-\beta_2' x_{2t}}^{\infty} f(y_t - \beta_1' x_{1t}, \varepsilon_2) d\varepsilon_2 \right) \\ &\quad + \sum_{I_t=0} \ln \left(\int_{-\infty}^{-\beta_2' x_{2t}} \int_{-\infty}^{\infty} f(\varepsilon_1, \varepsilon_2) d\varepsilon_1 d\varepsilon_2 \right) \end{aligned} \tag{12}$$

where f is the bivariate density of $(\varepsilon_1, \varepsilon_2)'$. An alternative to estimation under the assumption of normality is, of course, snp-estimation as discussed in the previous section. A clear advantage of this approach is that it estimates the density of the disturbances consistently if the number of terms K in the approximation increases with the number of observations. One choice for the snp-density is $h(\varepsilon) = h^*(\varepsilon)/S$ with

$$h^*(\varepsilon) = \sum_{i,j,k,l=0}^{K} \alpha_{ij} \alpha_{kl} \varepsilon_1^{i+k} \varepsilon_2^{j+l} \exp(-[\varepsilon_1^2/\delta_1^2 + \varepsilon_2^2/\delta_2^2]) \tag{13}$$

and S is a constant ensuring integration to 1. Substituting this density function in the loglikelihoodfunction, one obtains

$$\begin{aligned} \ell(\theta) &= \sum_{I_t=1} \ln \left(\sum_{i,j,k,l=0}^{K} \alpha_{ij} \alpha_{kl} (y_t - \beta_1' x_{1t})^{i+k} \exp(-(y_t - \beta_1' x_{1t})^2/\delta_1^2) \right. \\ &\qquad \left. \int_{-\beta_2' x_{2t}}^{\infty} \varepsilon_2^{j+l} \exp(-\varepsilon_2^2/\delta_2^2) d\varepsilon_2 \right) \\ &\quad + \sum_{I_t=0} \ln \left(\sum_{i,j,k,l=0}^{K} \alpha_{ij} \alpha_{kl} \int_{-\infty}^{\infty} \varepsilon_1^{i+k} \exp(-\varepsilon_1^2/\delta_1^2) d\varepsilon_1 \right. \end{aligned}$$

$$\int_{-\infty}^{-\beta_2' x_{2t}} \varepsilon_2^{j+l} \exp(-\varepsilon_2^2/\delta_2^2) d\varepsilon_2 \Bigg)$$

$$-T \ln \left(\sum_{i,j,k,l=0}^{K} \alpha_{ij}\alpha_{kl} \int_{-\infty}^{\infty} \varepsilon_1^{i+k} \exp(-\varepsilon_1^2/\delta_1^2) d\varepsilon_1 \right.$$

$$\left. \int_{-\infty}^{\infty} \varepsilon_2^{j+l} \exp(-\varepsilon_2^2/\delta_2^2) d\varepsilon_2 \right) \tag{14}$$

We have to impose restrictions on the parameters to achive identification. First, we set $\delta_2 = \sqrt{2}$ to ensure identification of the scale of equation (11). Second, we set $\alpha_{00} = 1$ 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, we could impose restrictions to ensure that the means of ε_1 and ε_2 are 0. For $K = 1$, one obtains the restrictions $\alpha_{01} = 0$ and $\alpha_{10} = 0$. For $K \geq 2$ the restrictions needed to ensure zero means become very cumbersome. Hence, as suggested by Melenberg and Van Soest (1993) we do not impose restrictions on the parameters to the density function of ε to impose a zero mean, but we restrict the intercepts of equation (10) and equation (11) instead.

It is not possible to test for normality using this particular class of snp-densities. Only the bivariate normal distribution with no correlation between ε_1 and ε_2 is a special case. However, by choosing another base class of density functions in the ERA approximation in equation (1) we can test for normality, even if the error terms are correlated. Because any density function with a finite moment generating function can be used as the basis in approximation (1), we can consider the following family of functions:

$$\bar{h}^*(\varepsilon) = \sum_{i,j,k,l=0}^{K} \alpha_{ij}\alpha_{kl}\varepsilon_1^{i+k}\varepsilon_2^{j+l} \exp(-\varepsilon'\Sigma^{-1}\varepsilon) \tag{15}$$

and define a generalized snp-density by $\bar{h}(\varepsilon) = \bar{h}^*(\varepsilon)/S$ (again, S is the constant that ensures integration to 1). A disadvantage of this generalized snp-density is that it does not have the same computationally attractive properties. A clear advantage is that bivariate normality (with unrestricted correlation) is a special case of this family ($\alpha_{ij} = 0$ for all $i + j \geq 1$). Evaluation of integrals will involve evaluation of bivariate normal probabilities, in general. These problems disappear however in the sample selection model where all relevant integrals are of the form

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

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

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

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

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

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

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

$$\int_{-\infty}^{b} \varepsilon_2^j \phi(\varepsilon_2) \int_{-\infty}^{\infty} \varepsilon_1^i \phi(\varepsilon_1|\varepsilon_2) d\varepsilon_1 d\varepsilon_2 = \int_{-\infty}^{b} \varepsilon_2^j \phi(\varepsilon_2) \mathcal{E}(\varepsilon_1^i|\varepsilon_2) d\varepsilon_2. \tag{16}$$

The last integral can be solved easily because

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

The coefficients a depened on the other parameters of the density function only and they are independent of ε_2. Note that both integrals in equation (16) can be calculated using the recursion formulas in appendix A.

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), it is possible to test for normality in this model. 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.

We conducted a limited simulation exercise to examine whether this extension has any promise. We consider the following simulation experiment:

$$y_t = \beta_{10} + \beta_{11}x_t + \beta_{12}w_t + \varepsilon_{1t} \tag{17}$$

$$I_t^* = \beta_{20} + \beta_{21}z_t + \beta_{22}w_t + \varepsilon_{2t} \tag{18}$$

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_t and z_t are independetly $\mathcal{N}(0, 3)$ distributed and w_t is distributed uniformly on $[-3, 3]$. We perform four 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 and variance matrix $\begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix}$, a bivariate t-distribution with $\operatorname{var}\varepsilon_1 = 4$, $\operatorname{var}\varepsilon_2 = 1$ and $\operatorname{cov}(\varepsilon_1, \varepsilon_2) = 1$, or ε is drawn from a centered χ^2-distribution[5]. All simulations were performed on Pentium workstations using the MAXLIK-library of GAUSS.

For each sample, we estimated the model based on a normal density, an snp-density of the form (13) with 1, 2 and 3 terms and an snp-density of the form (15) with 1 and 2 terms. The results are presented in detail in Appendix B. The case of normal disturbances is presented in tables 5 and 6, the case of t-disturbances in tables 7 and 8, and the case of the χ^2-disturbances is presented in tables 9 and 10.

It is remarkable how well standard ML under the assumption of normally distributed disturbances performs. Even if the true disturbances follow a t-distribution one sees that all estimated parameters are within two standard deviations of their true values. This is even the case when the disturbances follow transformed χ^2-distributions which are non-symmetric. In this setup it is not straightforward to test for normality using the LR-statistic as we have normalized the intercepts to 0 when we estimated the model semi-nonparametrically. Because the true intercepts are not 0 we expect some α's to differ from 0 in order to allow for a nonzero mean of the distribution of $(\varepsilon_1, \varepsilon_2)$. It is our impression that the generalized snp-density has some advantages over the univariate snp-density in this case. First, we encountered hardly any convergence problems using the generalized snp-density, while we did have convergence problems in the case of the univariate snp-density. Second, it seems that the generalized snp-density estimates the covariance between ε_1 and ε_2 better. However, in the case of χ^2-distributed disturbances one sees that the generalized snp-density with two terms performs badly: the covariance between ε_1 and ε_2 is vastly overestimated. In general, it is our impression that the price to be paid for a more flexible approximation to the unknown density (*i.e.*, a higher K) is lack of precision of the estimates. Moreover, computer time required for optimization of the loglikelihoodfunction increases quickly with K.

---

[5]To be precise, $\varepsilon_2 \sim \frac{1}{2}(v_1^2 + v_2^2) - 1$ and $\varepsilon_1 = \varepsilon_{2t} + (v_3^2 + v_4^2 + v_5^2)/\sqrt{2} - 3/\sqrt{2}$ with v_1^2 to v_5^2 independent $\chi^2(1)$ random variates. Hence, $\mathcal{E}\varepsilon_1 = \mathcal{E}\varepsilon_2 = 0$ and $\operatorname{var}\varepsilon_1 = 4$ and $\operatorname{var}\varepsilon_2 = \operatorname{cov}(\varepsilon_1, \varepsilon_2) = 1$.

4 Semi-nonparametric Estimation of a Housing Demand Model

We applied the snp-technique to a model of housing demand under Rent Assistance, see Koning and Ridder (1993) and Koning (1995). A comprehensive discussion of the theoretical model underlying the reduced form we estimate here can be found in these references, as well a detailed discussion of the data used. Our purpose here is to examine the sensitivity of the estimation results reported in these references to the assumption of normality.

4.1 A Reduced Form Model of Rental Housing Demand

In Koning and Ridder (1993) a structural model for rental housing demand in The Netherlands is developed and it is estimated in two steps. The structural model allows explicitly for a nonconvexity in the budget set faced by the households introduced by a Rent Assistance Program. The model is estimated in two steps. First, they derive a reduced form model and this model is estimated assuming normality of the stochastic terms in the model. In a second step they impose the restrictions implied by the structural model on the reduced form parameters and the structural parameters are estimated using minimum distance estimation. Here, we focus on the reduced form model only and on the sensitivity of these reduced form estimates to the normality assumption in particular. The reduced form model is:

$$I_t^* = \gamma_0 + \gamma_{Y_v} Y_{vt} + \gamma_Y Y_t + \eta_t \tag{19}$$

$$I_t = \begin{cases} 0 & I_t^* \leq 0 \\ 1 & I_t^* > 0 \end{cases}$$

$$R_t = \begin{cases} \beta_0 + \beta_Y Y_t + \varepsilon_{1t} & I_t = 0 \\ \beta_1 + \beta_{Y_v} Y_{vt} + \varepsilon_{2t} > R_{nt} & I_t = 1 \end{cases} \tag{20}$$

where R_t denotes housing demand that can occur in one of two regimes (labelled by $I_t = 0$ and $I_t = 1$). Households in the second regime receive Rent Assistance. In this regime, the appropriate income measure is virtual income Y_{vt} instead of income Y_t. Moreover, housing demand in the second regime is restricted: it must exceed a minimum rent R_{nt}. This minimum rent depends both on the household composition as well on pre-tax family income. The choice between both regimes is governed by the choice equation (19). The demand system is a switching regime truncated regression model with endogenous regimes. The parameters to be estimated are γ_0, γ_{Y_v}, γ_Y, β_0, β_1, β_Y and β_{Y_v}, as well as the parameters of the distribution of $(\varepsilon_1 \quad \varepsilon_2 \quad \eta)'$. We will assume in the sequel that all observations are mutually independent.

4.2 Estimation Results

In this section we present the estimation results of model (19)–(20) as given in Koning and Ridder (1993) and we compare these results with the ones obtained by snp-estimation.

The loglikelihoodfunction of the complete model in equations (19) and (20) is

$$\begin{aligned} \ell(\theta) &= \sum_{I_t=0} \ln f(R_t, I_t) + \sum_{I_t=1} \ln f(R_t | I_i, R_t \geq R_{nt}) f(I_t) \\ &= \sum_{I_t=0} \ln \int_{-\infty}^{-\bar{I}_t} f_{\varepsilon_1\eta}(R_t - \beta_0 - \beta_Y Y_t, \eta) d\eta + \\ &\quad \sum_{I_t=1} \ln \frac{\int_{-\bar{I}_t}^{\infty} f_{\varepsilon_2\eta}(R_t - \beta_1 - \beta_{Y_v} Y_{vt}, \eta) d\eta}{\Pr(\beta_1 + \beta_{Y_v} Y_{vt} + \varepsilon_{2t} > R_{nt}, I_t^* > 0)} \int_{-\bar{I}_t}^{\infty} f_\eta(\eta) d\eta \end{aligned} \tag{21}$$

where $f_{\varepsilon_1\eta}$ denotes the bivariate density of (ε_1, η), $f_{\varepsilon_2\eta}$ is the joint density of (ε_2, η), f_η is the marginal density of η, and $\bar{I}_t = \gamma_0 + \gamma_{Y_v} Y_{vt} + \gamma_Y Y$.

β_0	2.26	γ_0	1.12	$\sigma_{\varepsilon_1\eta}$	0.15
	(0.27)		(0.13)		(0.14)
β_Y	0.089	γ_{Y_v}	0.75	$\sigma_{\varepsilon_2\eta}$	0.37
	(0.0088)		(0.054)		(0.25)
β_1	4.11	γ_Y	−0.71	σ_{ε_1}	1.37
	(0.26)		(0.047)		(0.030)
β_{Y_v}	0.058			σ_{ε_2}	1.28
	(0.011)				(0.051)
ℓ	−3973.52				

Table 1: Estimation results, normal distribution (standard errors in parenthesis)

First, we assume that the disturbances are normally distributed:

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \eta \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \sigma^2_{\varepsilon_1} & \sigma_{\varepsilon_1\varepsilon_2} & \sigma_{\varepsilon_1\eta} \\ & \sigma^2_{\varepsilon_2} & \sigma_{\varepsilon_2\eta} \\ & & 1 \end{pmatrix}\right)$$

The loglikelikelihood (21) is maximized over the parameters of interest and the parameters of the density function of the disturbances. Note that the parameter $\sigma_{\varepsilon_1\varepsilon_2}$ is not identified because housing demand is observed in one regime only (ie, I_t is either equal to 0 or 1). According to the theoretical model, β_Y and β_{Y_v} should be equal, β_0 must be smaller than β_1, and γ_{Y_v} and γ_Y should have opposite sign and with the γ_Y being slightly smaller than γ_{Y_v} in absolute value. The estimation results are given in table 1.

The signs of the income variables in the demand equations are as expected and it turns out that the restrictions imposed by the theoretical model are not rejected (see Koning and Ridder (1993) and Koning (1995)). Here, however, we are less interested in the parameters estimates per se and more in the sensitivity if the estimation results in table 1 to the distributional assumption made.

To the knowledge of the authors, no simple tests for multivariate normality are available for limited dependent variable models consisting of more than one equation. Tests for distributional assumptions are available for single equation models like the probit and tobit model, see Bera, Jarque, and Lee (1984). A disadvantage of their testing procedure is that it is not clear what to do if the normality assumption is rejected. Using the snp-estimation method discussed in the previous sections, we can examine the sensitivity to the normality assumption. A direct test for normality as is feasible in snp-models with only one random variable is not possible because a trivariate normal distribution *with no restrictions on the covariance structure* is not a special member of the class of density functions we will use. Using the notation of the previous sections, we assume the following density for the disturbances

$$h(\varepsilon)^* = \sum_{i,j,k,l,m,n=0}^{K} \alpha_{ijk}\alpha_{lmn}\varepsilon_1^{i+l}\varepsilon_2^{j+m}\varepsilon_3^{k+n} \exp\left(-\left[\varepsilon_1^2/\delta_1^2 + \varepsilon_2^2/\delta_2^2 + \varepsilon_3^2/\delta_3^2\right]\right)$$
$$h(\varepsilon) = h^*(\varepsilon)/S \tag{22}$$

where S is a constant (depending on the parameters) that ensures integration to 1. Using equation (21), we obtain the following formula for the loglikelihood function[6]:

$$\ell(\theta) = \sum_{I_t=0}\left\{\ln\left(\int_{-\infty}^{-\gamma_0-\gamma_{Y_v}Y_{vt}-\gamma_Y Y_t}\int_{-\infty}^{\infty} h^*\left(R_t-\beta_0-\beta_Y Y_t, \varepsilon_2, \eta\right) d\varepsilon_2 d\eta\right)\right\}$$
$$+\sum_{I_t=1}\left\{\ln\left(\int_{-\gamma_0-\gamma_{Y_v}Y_{vt}-\gamma_Y Y_t}^{\infty}\int_{-\infty}^{\infty} h^*\left(\varepsilon_1, R_t-\beta_1-\beta_{Y_v}Y_{vt}, \eta\right) d\varepsilon_1 d\eta\right)\right.$$

[6] Details of this loglikelihoodfunction and the derivatives with respect to the parameters are available on request from the authors.

β_0	2.26	γ_Y	-0.90	α_{010}	-0.023
			(0.087)		(0.18)
β_Y	0.096	δ_1	1.46	α_{011}	0.48
	(0.0022)		(0.038)		(0.17)
β_1	4.11	δ_2	1.51	α_{100}	-0.20
			(0.078)		(0.16)
β_{Y_v}	0.055	δ_3	1.41	α_{101}	0.62
	(0.0069)				(0.14)
γ_0	1.12	α_{000}	1	α_{110}	-0.48
					(0.13)
γ_{Y_v}	0.96	α_{001}	0.24	α_{111}	-0.13
	(0.10)		(0.19)		(0.18)
ℓ	-3931.17				

Table 2: Estimation results, snp-distribution $K = 1$ (standard errors in parenthesis)

$$
\begin{aligned}
&- \ln\left(\int_{-\gamma_0-\gamma_{Y_v}Y_{vt}-\gamma_Y Y_t}^{\infty}\int_{R_{nt}-\beta_1-\beta_{Y_v}Y_{vt}}^{\infty}\int_{-\infty}^{\infty} h^*(\varepsilon_1,\varepsilon_2,\eta)\, d\varepsilon_1 d\varepsilon_2 d\eta\right)\\
&+ \ln\left(\int_{-\gamma_0-\gamma_{Y_v}Y_{vt}-\gamma_Y Y_t}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h^*(\varepsilon_1,\varepsilon_2,\eta)\, d\varepsilon_1 d\varepsilon_2 d\eta\right)\Bigg\}\\
&- N \ln\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h^*(\varepsilon_1,\varepsilon_2,\eta)\, d\varepsilon_1 d\varepsilon_2 d\eta\right) \qquad (23)
\end{aligned}
$$

Before we estimate the model using this loglikelihoodfunction, we must ensure identification of the parameters first. The identifying restrictions are completely analogous to those made in the sample selection model previously. First, we normalize the scale of the selection equation (19) by setting $\delta_3 = \sqrt{2}$. Second, we must normalize the α's, we do this by setting $\alpha_{000} = 1$. Third, we do not impose any parametric restrictions so that $(\varepsilon_1 \quad \varepsilon_2 \quad \eta)'$ has mean 0. Instead, we fix the intercepts β_0, β_1 and γ_0 to their estimated values in table 1.

The only problem left is the choice of K, the number of terms in the density (22). We follow here the suggestions of Gabler, Laisney, and Lechner: the model with K terms is nested in the model with $K + 1$ terms. One can use the likelihoodratio test to test whether the nulhypothesis that the additional terms are 0 is rejected[7]. This approach assumes that the true density is of the type (22) with $K + 1$ terms. Considering the flexibility of the snp-density, even for small K, we do not think that this is a strong assumption. Moreover, in the application discussed here, we are primarily interested in examining the sensitivity of the estimation results in table 1 to the assumption of normality.

The estimation results of the snp-model are given in tables 2 and 3. The point estimates for the income coefficients of both demand equations in the snp-model with $K = 1$ do not differ much from those in table 1, considering the standard errors in the latter table. The estimates for the intercepts are equal because of the normalization chosen. The estimates for the income coefficients in the choice equation are higher, but the variance η in the snp-model is higher than 1 as well. If one divides $\hat{\gamma}_{Y_v}$ and $\hat{\gamma}_Y$ in table 2 by the estimated standard deviation of η, one obtains 0.74 and -0.69 which numbers compare favourably with their counterparts in table 1.

We compare the first two moments of the snp-density with those of the normal distribution in the second and third column of table 4. The estimated mean of ε_1 is slightly negative and the estimated mean of ε_2 is slightly positive. However, both means are less than a half standard deviation of the estimated intercepts in the normal model. The fact that the mean of η is greater than 0 is in accordance with the estimated standard deviation of η being larger than 1. The variances of ε_1 and ε_2 do not differ by much between the normal model and the

[7] If this procedure is followed for more than one step, the significance level of the likelihoodratio test is no longer known because consecutive tests are not independent.

β_0	2.26	α_{010}	0.071	α_{120}	−0.18
			(0.96)		(0.50)
β_Y	0.10	α_{011}	0.12	α_{121}	−0.41
	(0.0035)		(0.69)		(0.32)
β_1	4.11	α_{012}	0.021	α_{122}	0.30
			(0.45)		(0.36)
β_{Y_v}	0.040	α_{020}	−0.64	α_{200}	−0.54
	(0.0097)		(0.52)		(0.32)
γ_0	1.12	α_{021}	0.40	α_{201}	−0.43
			(0.40)		(0.46)
γ_{Y_v}	0.95	α_{022}	0.24	α_{202}	−0.025
			(0.25)		(0.17)
γ_Y	−0.91	α_{100}	0.77	α_{210}	0.081
	(0.076)		(1.05)		(0.25)
δ_1	1.34	α_{101}	0.81	α_{211}	0.10
	(0.044)		(0.92)		(0.28)
δ_2	1.35	α_{102}	−0.15	α_{212}	−0.026
	(0.080)		(0.49)		(0.16)
δ_3	1.41	α_{110}	−1.09	α_{220}	0.16
			(0.66)		(0.22)
α_{000}	1	α_{111}	0.078	α_{221}	−0.0067
			(0.77)		(0.13)
α_{001}	0.64	α_{112}	0.45	α_{222}	0.025
	(1.04)		(0.47)		(0.11)
α_{002}	0.0050				
	(0.43)				
ℓ	−3908.85				

Table 3: Estimation results, snp-distribution $K = 2$ (standard errors in parenthesis)

	normal	snp ($K = 1$)	snp ($K = 2$)
$\mathcal{E}\varepsilon_1$	0	-0.11	-0.33
$\mathcal{E}\varepsilon_2$	0	0.12	0.030
$\mathcal{E}\eta$	0	0.17	0.48
$\text{var}\,\varepsilon_1$	1.88	1.83	1.85
$\text{var}\,\varepsilon_2$	1.64	1.74	2.56
$\text{var}\,\eta$	1	1.70	2.21
$\text{cov}\,(\varepsilon_1, \eta)$	0.16	0.34	0.14
$\text{cov}\,(\varepsilon_2, \eta)$	0.37	0.18	0.57

Table 4: Estimated means and (co)variances

snp-model. The covariances between ε_1 and η on the one hand and ε_2 and η on the other appear to have been changed by a lot, but one should realize that both values are within two times the standard deviation of the estimates of the normal model.

In a second step, we estimated the demand model with $K = 2$. The number of parameters is enormous: 7 parameters to model the means of the observed variables and 26 parameters to characterize the distribution of ε. The estimation results are given in table 3. Even though we reject the nulhypothesis that all extra α-terms are jointly zero (the likelihoodratio test statistic is 44.64 which must be compared with $\chi^2_{0.95}(19) = 30.14$), we do not think that this specification is an improvement over the one reported in table 2 because the loglikelihoodfunction was rather ill-determined near the optimum (analysis of the eigenvectors and eigenvalues of the Hessian showed that especially the estimates for the α's were ill-determined) and we were unable to verify whether optimum found was a local or a global optimum. For comparison with the other two specifications, we give the first two moments of the estimated distribution for ε in the fourth column of table 4. The variances of both ε_2 and η have increased and the expectation of ε_1 has decreased markedly. However, for the reasons given above we do not attach too much value to these results.

Summarizing, we have found that the estimation results for the truncated switching regression model of Koning and Ridder (1993) are not very sensitive to the assumed normality of the disturbances. The first two moments of the snp-density with $K = 1$ compared very well to those obtained by assuming normality. The snp-density with $K = 2$ contained too many parameters for reliable estimation.

5 Conclusions

In this paper we have used the semi-nonparametric maximum likelihood method developed by Gallant and Nychka (1987) to estimate to microeconometric models. First, we examined some properties of this method in the context of the sample selection model. An simple generalization of the base class of density functions has been used to test for normality in this sample selection model, for which no simple test of normality of the disturbances is available. Even though the number of simulations has been limited, we think that our generalization holds some promise. Computationally the generalization is not much more demanding than the snp-method itself due to the special structure of the sample selection model.

In the second part of the paper we used the snp-approach to examine whether the estimation results for a reduced form model of rental housing demand in The Netherlands as presented in Koning and Ridder (1993) and Koning (1995) are very sensitive to the assumed normality of the disturbances. We implemented the snp-method in this rather complex truncated switching regression model with endogenous regimes and the results indicated that the estimation results in the papers cited are not very sensitive to the assumed normality, even though we were not able to carry out a formal test of normality in this context.

References

Amemiya, T. (1985). *Advanced Econometrics*. Oxford: Basil Blackwell.

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

Gabler, S., F. Laisney, and M. Lechner (1993). Seminonparametric estimation of binary-choice models with an application to labor-force participation. *Journal of Business & Economics Statistics 11*, 61–80.

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

Greene, W.H. (1993). *Econometric Analysis* (Second ed.). New York: MacMillan.

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

Koning, R.H. (1995). *Essays on Applied Microeconometrics*. Capelle aan de IJssel: Labyrint Publication.

Koning, R.H. and G. Ridder (1993). Rent assistance and housing demand. Research Memorandum 544, Department of Economics, University of Groningen.

Manski, C.F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics 3*, 205–228.

Melenberg, B. and A. Van Soest (1993). Semi-parametric estimation of the sample selection model. Working Paper 9334, 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, Volume IV*, Amsterdam, pp. 2444–2512. North-Holland.

A Recursion Formulae of the Hermite Form Density

$I(k)$ is defined as the univariate integral

$$I(k) = \int_{-\infty}^{\infty} u^k \exp(-u^2/\delta^2) du$$

Using partial integration one obtains the recursion formulae

$$I(k) = \begin{cases} \delta\sqrt{\pi} & k = 0 \\ 0 & k = 1, 3, 5, \ldots \\ \frac{1}{2}(k-1)\delta^2 I(k-2) & k = 2, 4, 6, \ldots \end{cases} \tag{24}$$

B Simulation Results

	normal	snp-density			generalized snp	
$T = 500$		$K = 1$	$K = 2$	$K = 3$	$K = 1$	$K = 2$
β_{10}	1.02	0	0	0	0	0
	(0.076)					
β_{11}	0.51	0.50	0.51	0.51	0.51	0.51
	(0.069)	(0.075)	(0.070)	(0.069)	(0.069)	(0.068)
β_{12}	−0.52	−0.58	−0.58	−0.53	−0.52	−0.51
	(0.067)	(0.069)	(0.070)	(0.073)	(0.063)	(0.072)
β_{20}	1.03	0	0	0	0	0
	(0.080)					
β_{21}	−1.01	−0.95	−0.97	−1.06	−1.01	−1.04
	(0.062)	(0.056)	(0.067)	(0.15)	(0.061)	(0.076)
β_{22}	1.01	0.95	0.97	1.07	1.01	1.05
	(0.067)	(0.055)	(0.064)	(0.15)	(0.064)	(0.079)
σ_1	1.97				1.97	1.74
	(0.071)				(0.068)	(0.14)
σ_{12}	0.96				0.95	0.87
	(0.066)				(0.064)	(0.088)
δ_1		2.95	2.60	2.51		
		(0.096)	(0.15)	(0.28)		
α_{00}		1	1	1	1	1
α_{01}		0.55	1.37	0.60	0.56	0.62
		(0.056)	(0.12)	(0.25)	(0.071)	(0.14)
α_{02}			0.33	0.035		0.18
			(0.12)	(0.12)		(0.078)
α_{03}				−0.014		
				(0.061)		
α_{10}		0.11	−0.54	−0.015	0.011	−0.018
		(0.023)	(0.15)	(0.18)	(0.032)	(0.093)
α_{11}		0.17	0.44	0.15	0.047	−0.094
		(0.030)	(0.086)	(0.23)	(0.029)	(0.082)
α_{12}			0.30	0.089		0.016
			(0.064)	(0.069)		(0.053)
α_{13}				0.0086		
				(0.036)		
α_{20}			−0.12	−0.019		0.036
			(0.040)	(0.043)		(0.046)
α_{21}			0.011	−0.021		0.011
			(0.029)	(0.044)		(0.034)
α_{22}			0.046	0.024		0.0069
			(0.018)	(0.017)		(0.010)
α_{23}				0.011		
				(0.0088)		
α_{30}				−0.0011		
				(0.015)		
α_{31}				−0.0025		
				(0.018)		
α_{32}				0.0031		
				(0.0066)		
α_{33}				0.0017		
				(0.0029)		
$\mathcal{E}\varepsilon_1$	0	1.19	1.67	1.15	1.03	1.03
$\mathcal{E}\varepsilon_2$	0	0.92	1.02	1.32	1.03	1.10
$\text{var}\,\varepsilon_1$	3.89	3.97	5.29	3.97	3.89	3.93
$\text{var}\,\varepsilon_2$	1	0.85	1.52	0.85	0.78	0.96
$\text{cov}\,(\varepsilon_1, \varepsilon_2)$	0.96	0.34	0.22	1.00	0.85	0.94
ℓ	−730.49	−733.79	−732.85	−726.32	−730.17	−728.52
success	100	100	100	100	100	100

Table 5: Estimation results of the experiment with bivariate normal distributed disturbances, 100 replications, $T = 500$.

	normal	snp-density			generalized snp	
$T = 500$		$K = 1$	$K = 2$	$K = 3$	$K = 1$	$K = 2$
β_{10}	0.99	0	0	0	0	0
	(0.037)					
β_{11}	0.50	0.50	0.50	0.50	0.50	0.50
	(0.048)	(0.046)	(0.047)	(0.045)	(0.045)	(0.046)
β_{12}	−0.50	−0.56	−0.51	−0.50	−0.50	−0.49
	(0.048)	(0.052)	(0.059)	(0.055)	(0.046)	(0.055)
β_{20}	1.00	0	0	0	0	0
	(0.044)					
β_{21}	−1.00	−0.89	−1.09	−1.09	−0.88	−0.99
	(0.044)	(0.042)	(0.087)	(0.12)	(0.040)	(0.056)
β_{22}	1.00	0.89	1.09	1.09	0.88	0.98
	(0.041)	(0.039)	(0.080)	(0.12)	(0.033)	(0.050)
σ_1	1.99				1.99	1.91
	(0.050)				(0.043)	(0.076)
σ_{12}	1.00				0.97	0.86
	(0.021)				(0.019)	(0.043)
δ_1		2.96	2.67	2.58		
		(0.063)	(0.11)	(0.22)		
α_{00}		1	1	1	1	1
α_{01}		0.55	0.64	0.56	0.53	0.53
		(0.049)	(0.10)	(0.15)	(0.043)	(0.073)
α_{02}			0.047	0.022		0.079
			(0.087)	(0.072)		(0.047)
α_{03}				−0.011		
				(0.035)		
α_{10}		0.10	−0.14	−0.030	0.013	−0.017
		(0.017)	(0.10)	(0.090)	(0.017)	(0.055)
α_{11}		0.16	0.27	0.16	0.040	0.031
		(0.024)	(0.064)	(0.12)	(0.019)	(0.049)
α_{12}			0.12	0.093		0.024
			(0.044)	(0.051)		(0.028)
α_{13}				0.0070		
				(0.022)		
α_{20}			−0.048	−0.020		−0.0038
			(0.022)	(0.025)		(0.018)
α_{21}			0.015	−0.026		0.00034
			(0.014)	(0.033)		(0.015)
α_{22}			0.023	0.022		0.0017
			(0.0097)	(0.012)		(0.0043)
α_{23}				0.010		
				(0.0064)		
α_{30}				−0.00030		
				(0.0068)		
α_{31}				−0.0055		
				(0.0094)		
α_{32}				0.0015		
				(0.0033)		
α_{33}				0.0016		
				(0.0018)		
$\mathcal{E}\varepsilon_1$	0	1.16	1.32	1.05	1.00	0.99
$\mathcal{E}\varepsilon_2$	0	0.84	1.18	1.15	0.88	1.01
$\operatorname{var}\varepsilon_1$	3.95	4.03	5.05	3.94	3.89	3.89
$\operatorname{var}\varepsilon_2$	1	0.85	1.11	0.85	0.79	0.93
$\operatorname{cov}(\varepsilon_1, \varepsilon_2)$	1.00	0.34	0.99	1.05	0.85	0.98
ℓ	−1597.38	−1603.79	−1597.64	−1593.36	−1597.24	−1595.49
success	99	100	100	100	100	100

Table 6: Estimation results of the experiment with bivariate normal distributed disturbances, 100 replications, $T = 1000$.

	normal	snp-density			generalized snp	
$T = 500$		$K = 1$	$K = 2$	$K = 3$	$K = 1$	$K = 2$
β_{10}	1.04	0	0	0	0	0
	(0.10)					
β_{11}	0.49	0.47	0.49	0.48	0.49	0.49
	(0.069)	(0.068)	(0.064)	(0.058)	(0.061)	(0.061)
β_{12}	−0.52	−0.59	−0.52	−0.55	−0.53	−0.53
	(0.074)	(0.076)	(0.071)	(0.062)	(0.070)	(0.068)
β_{20}	1.15	0	0	0	0	0
	(0.17)					
β_{21}	−1.14	−1.05	−1.22	−1.26	−1.13	−1.48
	(0.15)	(0.089)	(0.16)	(0.21)	(0.099)	(0.19)
β_{22}	1.14	1.03	1.22	1.26	1.12	1.48
	(0.15)	(0.086)	(0.16)	(0.21)	(0.094)	(0.18)
σ_1	1.94				1.83	2.19
	(0.23)				(0.12)	(0.20)
σ_{12}	0.84				0.86	1.08
	(0.19)				(0.17)	(0.33)
δ_1		2.71	2.88	3.00		
		(0.16)	(0.33)	(0.37)		
α_{00}		1	1	1	1	1
α_{01}		0.71	0.84	1.46	0.81	1.73
		(0.095)	(0.30)	(0.75)	(0.15)	(1.08)
α_{02}			0.034	0.092		0.50
			(0.017)	(0.26)		(0.44)
α_{03}				−0.13		
				(0.076)		
α_{10}		0.16	0.085	0.20	0.054	0.14
		(0.041)	(0.094)	(0.26)	(0.072)	(0.27)
α_{11}		0.24	0.17	0.28	0.023	0.10
		(0.048)	(0.063)	(0.19)	(0.055)	(0.30)
α_{12}			0.083	0.10		−0.019
			(0.052)	(0.14)		(0.18)
α_{13}				0.0073		
				(0.047)		
α_{20}			−0.070	−0.057		−0.063
			(0.026)	(0.030)		(0.044)
α_{21}			−0.015	−0.080		−0.062
			(0.017)	(0.055)		(0.053)
α_{22}			0.024	0.012		−0.018
			(0.016)	(0.017)		(0.041)
α_{23}				0.015		
				(0.011)		
α_{30}				−0.0096		
				(0.012)		
α_{31}				−0.016		
				(0.016)		
α_{32}				−0.0011		
				(0.0058)		
α_{33}				0.0016		
				(0.0026)		
$\mathcal{E}\varepsilon_1$	0	1.35	1.07	1.34	1.10	1.19
$\mathcal{E}\varepsilon_2$	0	0.98	1.44	1.31	1.12	1.78
$\operatorname{var}\varepsilon_1$	3.75	3.12	3.02	3.12	3.17	2.81
$\operatorname{var}\varepsilon_2$	1	0.91	0.86	0.91	0.83	0.62
$\operatorname{cov}(\varepsilon_1, \varepsilon_2)$	0.84	0.18	0.44	0.32	0.70	0.45
ℓ	−709.70	−693.19	−685.58	−672.11	−693.97	−679.73
success	100	74	82	86	100	100

Table 7: Estimation results of the experiment with bivariate t distributed disturbances, 100 replications, $T = 500$.

	normal		snp-density		generalized snp	
$T = 500$		$K = 1$	$K = 2$	$K = 3$	$K = 1$	$K = 2$
β_{10}	0.99	0	0	0	0	0
	(0.068)					
β_{11}	0.50	0.50	0.51	0.50	0.50	0.50
	(0.045)	(0.044)	(0.037)	(0.033)	(0.039)	(0.034)
β_{12}	−0.49	−0.58	−0.51	−0.54	−0.52	−0.50
	(0.049)	(0.049)	(0.049)	(0.044)	(0.047)	(0.043)
β_{20}	1.09	0	0	0	0	0
	(0.068)					
β_{21}	−1.09	−0.99	−1.12	−1.07	−1.00	−1.16
	(0.078)	(0.054)	(0.11)	(0.17)	(0.061)	(0.098)
β_{22}	1.11	1.00	1.13	1.08	1.01	1.17
	(0.071)	(0.055)	(0.11)	(0.16)	(0.059)	(0.093)
σ_1	1.94				1.85	2.14
	(0.22)				(0.086)	(0.15)
σ_{12}	0.83				0.82	0.96
	(0.13)				(0.12)	(0.19)
δ_1		2.67	2.83	2.98		
		(0.11)	(0.21)	(0.24)		
α_{00}		1	1	1	1	1
α_{01}		0.66	0.68	1.04	0.73	0.84
		(0.036)	(0.11)	(0.26)	(0.098)	(0.17)
α_{02}			0.017	0.019		0.086
			(0.13)	(0.14)		(0.12)
α_{03}				−0.095		
				(0.057)		
α_{10}		0.15	0.065	0.11	0.090	−0.035
		(0.027)	(0.065)	(0.10)	(0.054)	(0.074)
α_{11}		0.26	0.17	0.27	0.0015	0.11
		(0.039)	(0.052)	(0.18)	(0.037)	(0.11)
α_{12}			0.077	0.11		0.12
			(0.042)	(0.080)		(0.042)
α_{13}				−0.0011		
				(0.049)		
α_{20}			−0.062	−0.043		−0.073
			(0.019)	(0.019)		(0.023)
α_{21}			−0.014	−0.051		−0.070
			(0.012)	(0.034)		(0.018)
α_{22}			0.020	0.0042		−0.0010
			(0.014)	(0.012)		(0.0072)
α_{23}				0.0084		
				(0.0087)		
α_{30}				−0.0055		
				(0.0061)		
α_{31}				−0.011		
				(0.011)		
α_{32}				−0.0024		
				(0.0035)		
α_{33}				0.00088		
				(0.0029)		
$\mathcal{E}\varepsilon_1$	0	1.31	1.01	1.13	1.10	1.05
$\mathcal{E}\varepsilon_2$	0	0.91	1.15	1.12	0.97	1.22
$\operatorname{var}\varepsilon_1$	3.75	3.14	2.84	3.14	3.22	3.32
$\operatorname{var}\varepsilon_2$	1	0.94	0.84	0.94	0.84	0.97
$\operatorname{cov}(\varepsilon_1, \varepsilon_2)$	0.83	0.26	0.46	0.32	0.64	0.90
ℓ	−1559.85	−1516.91	−1501.60	−1477.31	−1526.57	−1487.14
success	99	67	71	78	77	100

Table 8: Estimation results of the experiment with bivariate t distributed disturbances, 100 replications, $T = 1000$.

	normal	snp-density			generalized snp	
$T = 500$		$K = 1$	$K = 2$	$K = 3$	$K = 1$	$K = 2$
β_{10}	0.92	0	0	0	0	0
	(0.10)					
β_{11}	0.50	0.50	0.50	0.50	0.50	0.50
	(0.062)	(0.064)	(0.052)	(0.054)	(0.065)	(0.051)
β_{12}	−0.48	−0.63	−0.49	−0.52	−0.52	−0.44
	(0.081)	(0.085)	(0.071)	(0.073)	(0.082)	(0.071)
β_{20}	1.17	0	0	0	0	0
	(0.10)					
β_{21}	−1.15	−1.10	−1.02	−1.14	−1.09	−0.90
	(0.11)	(0.091)	(0.11)	(0.13)	(0.081)	(0.11)
β_{22}	1.14	1.09	1.01	1.13	1.08	0.91
	(0.12)	(0.10)	(0.12)	(0.14)	(0.090)	(0.11)
σ_1	2.06				2.12	1.91
	(0.12)				(0.11)	(0.16)
σ_{12}	1.28				1.18	1.56
	(0.15)				(0.15)	(0.18)
δ_1		3.07	2.98	2.82		
		(0.16)	(0.19)	(0.25)		
α_{00}		1	1	1	1	1
α_{01}		0.64	0.50	0.83	0.65	1.56
		(0.047)	(0.077)	(0.11)	(0.057)	(0.38)
α_{02}			−0.062	−0.065		0.42
			(0.039)	(0.055)		(0.19)
α_{03}				−0.085		
				(0.022)		
α_{10}		0.15	−0.046	0.030	−0.014	−0.74
		(0.016)	(0.034)	(0.058)	(0.039)	(0.26)
α_{11}		0.12	0.11	0.10	−0.00036	−0.22
		(0.012)	(0.015)	(0.037)	(0.010)	(0.15)
α_{12}			0.078	0.065		−0.0046
			(0.015)	(0.025)		(0.042)
α_{13}				0.0097		
				(0.011)		
α_{20}			−0.080	−0.090		−0.15
			(0.013)	(0.025)		(0.038)
α_{21}			−0.0038	−0.065		0.011
			(0.0047)	(0.018)		(0.027)
α_{22}			0.025	0.026		0.021
			(0.0048)	(0.010)		(0.0072)
α_{23}				0.014		
				(0.0047)		
α_{30}				−0.0075		
				(0.0031)		
α_{31}				−0.0082		
				(0.0031)		
α_{32}				0.0013		
				(0.0014)		
α_{33}				0.0015		
				(0.00082)		
$\mathcal{E}\varepsilon_1$	0	1.37	0.95	1.11	1.00	0.72
$\mathcal{E}\varepsilon_2$	0	0.91	0.96	1.07	0.90	0.92
$\operatorname{var}\varepsilon_1$	4.24	3.91	5.31	3.91	4.17	4.29
$\operatorname{var}\varepsilon_2$	1	0.77	1.10	0.77	0.76	1.05
$\operatorname{cov}(\varepsilon_1, \varepsilon_2)$	1.28	0.057	0.67	0.62	0.91	1.60
ℓ	−800.56	−803.39	−780.71	−765.70	−799.86	−764.49
success	100	99	99	97	100	100

Table 9: Estimation results of the experiment with bivariate χ^2 distributed disturbances, 100 replications, $T = 500$.

	normal	snp-density			generalized snp	
$T = 1000$		$K = 1$	$K = 2$	$K = 3$	$K = 1$	$K = 2$
β_{10}	0.94	0	0	0	0	0
	(0.087)					
β_{11}	0.50	0.50	0.50	0.50	0.50	0.50
	(0.048)	(0.050)	(0.040)	(0.039)	(0.050)	(0.036)
β_{12}	−0.49	−0.63	−0.49	−0.52	−0.52	−0.45
	(0.054)	(0.061)	(0.047)	(0.046)	(0.058)	(0.043)
β_{20}	1.16	0	0	0	0	0
	(0.082)					
β_{21}	−1.11	−0.94	−1.04	−1.06	−0.92	−0.91
	(0.089)	(0.067)	(0.088)	(0.10)	(0.059)	(0.092)
β_{22}	1.12	0.94	1.05	1.05	0.92	0.93
	(0.082)	(0.062)	(0.083)	(0.096)	(0.057)	(0.084)
σ_1	2.05				2.10	1.94
	(0.081)				(0.079)	(0.097)
σ_{12}	1.26				1.19	1.55
	(0.19)				(0.16)	(0.14)
δ_1		3.06	3.00	2.84		
		(0.11)	(0.13)	(0.20)		
α_{00}		1	1	1	1	1
α_{01}		0.65	0.50	0.84	0.64	1.32
		(0.035)	(0.053)	(0.066)	(0.048)	(0.25)
α_{02}			−0.067	−0.067		0.27
			(0.027)	(0.037)		(0.091)
α_{03}				−0.087		
				(0.014)		
α_{10}		0.15	−0.040	0.029	−0.015	−0.58
		(0.012)	(0.027)	(0.043)	(0.036)	(0.16)
α_{11}		0.12	0.11	0.10	−0.0015	−0.092
		(0.0087)	(0.0097)	(0.025)	(0.0094)	(0.069)
α_{12}			0.075	0.065		0.010
			(0.011)	(0.021)		(0.028)
α_{13}				0.0095		
				(0.0085)		
α_{20}			−0.077	−0.087		−0.15
			(0.0099)	(0.017)		(0.022)
α_{21}			−0.0042	−0.063		−0.00015
			(0.0034)	(0.014)		(0.018)
α_{22}			0.024	0.025		0.018
			(0.0032)	(0.0054)		(0.0039)
α_{23}				0.014		
				(0.0029)		
α_{30}				−0.0071		
				(0.0017)		
α_{31}				−0.0077		
				(0.0020)		
α_{32}				0.0012		
				(0.00079)		
α_{33}				0.0014		
				(0.00048)		
$\mathcal{E}\varepsilon_1$	0	1.36	0.91	1.10	0.99	0.75
$\mathcal{E}\varepsilon_2$	0	0.92	0.95	1.06	0.90	0.92
$\operatorname{var}\varepsilon_1$	4.19	5.05	2.84	3.89	4.11	4.13
$\operatorname{var}\varepsilon_2$	1	0.77	1.07	0.77	0.76	1.02
$\operatorname{cov}(\varepsilon_1, \varepsilon_2)$	1.26	0.05	0.67	0.62	0.91	1.52
ℓ	−1586.21	−1591.26	−1547.47	−1517.29	−1583.71	−1516.29
success	99	98	99	87	100	100

Table 10: Estimation results of the experiment with bivariate χ^2 distributed disturbances, 100 replications, $T = 1000$.