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

Bas van der Klaauw
Department of Economics, Free University Amsterdam, De Boelelaan 1105, NL-1081 HV Amsterdam, The Netherlands and Scholar (*klaauw@tinbergen.nl*)

Ruud H. Koning
Department of Econometrics, University of Groningen, P.O. Box 800, NL-9700 AV Groningen, The Netherlands (*r.h.koning@eco.rug.nl*)

In this article we introduce a test for the normality assumption in the sample selection model. The test is based on a flexible parametric specification of the density function of the error terms in the model. This specification follows a Hermite series with bivariate normality as a special case. All parameters of the model are estimated both under normality and under the more general flexible parametric specification, which enables testing for normality using a standard likelihood ratio test. If normality is rejected, then the flexible parametric specification provides consistent parameter estimates. The test has reasonable power, as is shown by a simulation study. The test also detects some types of ignored heteroscedasticity. Finally, we apply the flexible specification of the density to a travel demand model and test for normality in this model.

KEY WORDS: Flexible parametric density estimation; Hermite series; Heteroscedasticity; Sample selection.

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 (e.g., omitting exogenous variables or inappropriately assuming linearity) and misspecification of the stochastic structure of the model (e.g., neglecting heteroscedasticity or misspecification of the distribution of the random variables). The maximum likelihood estimator is generally inconsistent in these cases. This is particularly true in the case of limited dependent variable models (e.g., Hurd 1979; Smith 1989). In this article we focus on one particular form of misspecification: misspecification of the distribution of the error terms. We retain the assumption of correct specification of the functional form of the relationship.

The model that we study is the sample selection model introduced by Heckman (1979). This type of model accounts for problems arising 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 only the wages of employed workers are observed (e.g., Buchinsky 1998; Melenberg and Van Soest 1993). This model is not only very useful when the outcome variable is only observed for a (nonrandom) part of the population, but it can easily be extended to the switching regression model. The switching regression model is the point of departure for many interesting problems in economics. For example, it is used to evaluate active labor market programs (Heckman, LaLonde, and Smith 1999), to investigate the effect of migration on earnings (Tunalı 2000), to study the returns to eduction (Heckman, Tobias, and Vytlacil 2000), and to estimate the effect of union membership on earnings (Lee 1978). In the switching regression model, many of the interesting policy parameters, such as the treatment effect on the treated, depend not only on the functional form of the relationship, but also on the joint distribution of the error terms. Therefore, for policy evaluation and also for simulations, both the parameters of the functional form of the relationship and the joint distribution of the error terms must be known. Of course, this does not hold for either the switching regression model, or also for other models, like the sample selection model.

Usually, the sample selection model is estimated either by maximum likelihood or by Heckman's two-step procedure (Heckman 1979), under the assumption that the error term in the regression equation and the error term in the selection equation follow a bivariate normal distribution. However, it is often argued that the parameter estimates are sensitive to the distributional assumptions (Manski 1989). In fact, Newey (1999a) proved that only under very restrictive assumptions, the parameter estimates are consistent against misspecification of the distribution of the error terms. In many cases, the normal distribution might be overly restrictive. As mentioned earlier, the sample selection model is often used to model earnings, and it is well known from the economic literature that the distribution of (the logarithm of) earnings in the population has fatter tails than the normal distribution function (e.g., Heckman et al. 2000).

Journal of Business & Economic Statistics
January 2003, Vol. 21, No. 1
DOI 10.1198/073500102288618739

We introduce a formal test for the normality assumption within the sample selection model. The test is motivated by the seminonparametric maximum likelihood estimation method of Gallant and Nychka (1987) in which the true distribution of the error terms is approximated by a Hermite series. The density of the error terms follows a very flexible parametric specification, which is estimated using maximum likelihood together with all other parameters of the model. The test thus provides an alternative distribution of the error terms in the event that normality is rejected. Because the parameters are estimated using maximum likelihood, the estimators are efficient, $\sqrt{N}$ consistent, and asymptotically normal. To examine the power of the normality test, we perform a simulation study, which will also provide practical experience about the actual flexibility of the parametric specification of the density. For the binary choice model and duration models, Smith (1989) proposed a similar type of polynomial specification to test for normality. Gabler, Laisney, and Lechner (1993) provided Monte Carlo evidence on this test and applied it to a model of female labor force participation.

So far we did not discuss heteroscedasticity of the error terms. Ignored heteroscedasticity causes the usual estimators in limited dependent-variable models to be inconsistent. For example, Hurd (1979) showed that the bias caused by ignored heteroscedasticity is severe in the truncated regression model. Not many estimation procedures are robust against heteroscedasticity. An exception is the quantile regression approach of Buchinsky (1998). The disadvantage of quantile regression methods is that the complete distribution of the error terms is not estimated. This complicates performing simulations, and because the mean is not estimated, it is difficult to define relevant policy parameters (see Abadie, Angrist, and Imbens 1998, who applied the quantile regression estimator to the switching regression model). In our simulation study we also focus on the robustness of the flexible parametric density against heteroscedastic error terms.

An advantage of the flexible parametric density used in this article is its general applicability; it is not specific to one particular econometric model. The approach taken herein can be used to examine the sensitivity of estimation results to the assumption of normality in other microeconometric models where no formal test of normality is available. We illustrate both the flexible parametric density and the normality test in a model of travel demand and car ownership.

The article is organized as follows. Section 2 discusses the sample selection model and the flexible parametric density that naturally leads to a normality test for the sample selection model. Section 3 reports some simulations to investigate the performance of the flexible parametric density and the power of the normality test. Section 4 discusses an application of the estimation method and the normality test to the model of travel demand. Finally, Section 5 concludes.

2. IDENTIFICATION AND ESTIMATION OF THE SAMPLE SELECTION MODEL

In this section we discuss some identification and estimation issues associated with the sample selection model. In particular, we focus on the flexible parametric density that leads to a test for bivariate normality.

Within a population, we are interested in the distribution of an outcome variable y conditional on a vector of exogenous variables x_1. Therefore, we specify for individual i ($i = 1, \ldots, N$) the regression equation

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

The dependent variable y_i is observed for a subsample only. A binary variable z_i is introduced that equals 1 if the realization of y_i is observed and otherwise equals 0. The selection rule depends on a set of exogenous variables x_2 and is given by

$$z_i^* = \beta_2' x_{2i} + \varepsilon_{2i}$$

and

$$z_i = \begin{cases} 1, & z_i^* > 0 \\ 0, & z_i^* \leq 0. \end{cases} \tag{2}$$

As mentioned in Section 1, we focus only on the joint distribution of the error terms $(\varepsilon_{1i}, \varepsilon_{2i})$, and thus we simply assume linear specifications in both the regression and the selection equation. If the conditional expectation of ε_{1i} given $z_i = 1$ does not equal 0, ordinary least squares (OLS) estimation of (1) will not yield consistent estimates for the parameters of interest, β_1.

If one is willing to assume that the error terms $(\varepsilon_{1i}, \varepsilon_{2i})$ follow a bivariate normal distribution, then the model can be estimated using Heckman's two-stage procedure (see Heckman 1979). In the first step, a probit model is estimated for the selection equation to obtain $\hat{\beta}_2$, in the second step, the regression equation

$$y_i = \beta_1' x_{1i} + \rho\sigma_1 \frac{\phi(\hat{\beta}_2{}' x_{2i})}{\Phi(\hat{\beta}_2{}' x_{2i})} + \eta_i \tag{3}$$

is estimated (by OLS) for the subsample for which a realization of y_i is actually observed ($z_i = 1$). The parameter ρ denotes the correlation between ε_1 and ε_2, and σ_1^2 is the variance of ε_1 (the variance of ε_2 is normalized to 1). If the vector x_{1i} equals the vector x_{2i}, then identification of the model hinges entirely on the functional form and distributional assumptions, which makes the parameter estimates very sensitive to the distributional assumptions (see Manski 1989 for an extensive overview). Therefore, in most economic applications, an exclusion restriction is imposed—at least one variable enters only the selection equation and not the regression equation.

The bivariate normality assumption may be overly restrictive. Therefore, starting with Lee (1982), who modified Heckman's two-step procedure to allow in a general way for departures from normality, a number of estimation methods have been introduced that relax the distributional assumptions. These (semiparametric) estimation methods generalize the probit estimation in the first step to, for example, kernel-based estimation and use more flexible error correction terms in the regression equation (e.g., Ahn and Powell 1993; Cosslett 1991; Ichimura and Lee 1991; Newey 1999b). This provides consistent estimators for the parameters of interest β_1, but computation of the corrected standard errors of these parameters is quite cumbersome (e.g., Ahn and Powell

1993; Cosslet 1991). A main disadvantage of most semiparametric estimation methods is that the intercept in the outcome equation remains unidentified (and thus cannot be estimated). As stressed by Heckman (1990), in many economic applications the intercept is an important parameter of interest.

The alternative to two-step estimation is maximum likelihood estimation. This requires full specification of the joint density function of the error terms, which is denoted by $f(\varepsilon_1, \varepsilon_2)$. Under the assumption of correct specification, maximum likelihood estimation yields consistent estimates of all parameters of interest and straightforward computation of the asymptotic distribution. The log-likelihood function of the sample selection model is

$$\log \mathcal{L} = \sum_{i=1}^{N} z_i \log\Big(\int_{-\beta_2' x_{2i}}^{\infty} f(y_i - \beta_1' x_{1i}, \varepsilon_2)\, d\varepsilon_2\Big) + (1 - z_i) \log\Big(\int_{-\infty}^{-\beta_2' x_{2i}} \int_{-\infty}^{\infty} f(\varepsilon_1, \varepsilon_2)\, d\varepsilon_1 d\varepsilon_2\Big). \quad (4)$$

In maximum likelihood estimation, consistency depends crucially on the correct specification of the joint density of the error terms. Therefore, it is convenient to specify this density so that it is flexible. We use flexible parametric specifications that are inspired by seminonparametric estimation as introduced by Gallant and Nychka (1987). The unknown density function $f(\varepsilon_1, \varepsilon_2)$ is approximated by a Hermite series.

We consider densities $h(\varepsilon)$ in a subclass $\mathcal{H}_K$ that are of the Hermite form

$$h^*(\varepsilon) = P_K^2(\varepsilon)\phi^2(\varepsilon|\Sigma), \quad (5)$$

where $P_K(\varepsilon)$ is a polynomial of degree K and $\phi(\varepsilon|\Sigma)$ is the normal density function with mean 0 and covariance matrix Σ. In fact, Gallant and Nychka (1987) restricted Σ to being a diagonal matrix. Because of the squaring, no restrictions are necessary to ensure that $h^*(\varepsilon)$ is nonnegative. Some restrictions are required to ensure that the density integrates to 1 and for identification. We return to these restrictions later.

Gallant and Nychka (1987) showed that a large class $\mathcal{H}$ of density functions can be approximated arbitrarily well by increasing the number of terms K of the polynomial $P_K(\varepsilon)$. They gave the precise conditions defining $\mathcal{H}$. 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, and only very violently oscillatory densities are excluded from $\mathcal{H}$. Gallant and Nychka (1987) proved that the densities in $\mathcal{H}$ can be estimated consistently by increasing K with the number of observations. 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 (see, e.g., Cameron and Johansson 1997).

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. This is the interpretation that we follow in this article. This 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. Because we want to test for bivariate normality (with unrestricted correlation) in the sample selection model, we do not restrict the covariance matrix, unlike Gallant and Nychka (1987). This provides a formal test for normality in the sample selection model; that is, $P_K(\varepsilon)$ is constant for all values of ε. Such a test for normality has not yet been derived in the literature. An alternative could be to perform Hausman tests on the slope coefficients of the regressions equation. One could either use Hausman tests to test normality against a semiparametric alternative or also against the flexible parametric alternative discussed in this article. The first alternative may be somewhat cumbersome, because it requires computing standard errors in the semiparametric estimation, which is not always straightforward. The second alternative is easy to implement, but these tests are not expected to be as powerful as the tests derived in this article. The simulations in Section 3 indicate that the slope coefficients are not very sensitive to departures from normality. Hausman tests leave out any information on the distribution of the error terms.

In line with Gallant and Nychka (1987), we parameterize $h^*(\varepsilon)$ as

$$h^*(\varepsilon) = \Big(\sum_{i,j=0}^{K} \alpha_{ij}\varepsilon_1^i \varepsilon_2^j\Big)^2 \exp(-\varepsilon'\Sigma^{-1}\varepsilon) = \sum_{i,j,k,l=0}^{K} \alpha_{ij}\alpha_{kl}\varepsilon_1^{i+k}\varepsilon_2^{j+l} \exp(-\varepsilon'\Sigma^{-1}\varepsilon).$$

This bivariate density can be easily generalized to a density function for more variables, which is important when, for example, studying switching regression models containing three error terms. To ensure that this density integrates to 1, a restriction on the parameters is necessary. This restriction can take the form of an explicit restriction on the parameters of the density. However, for computational convenience, we follow Gabler et al. (1993), who ensured integration to 1 by scaling the density. Define S by

$$S = \int_{R^2} \sum_{i,j,k,l=0}^{K} \alpha_{ij}\alpha_{kl}\varepsilon_1^{i+k}\varepsilon_2^{j+l} \exp(-\varepsilon'\Sigma^{-1}\varepsilon)\, d\varepsilon = \sum_{i,j,k,l=0}^{K} \alpha_{ij}\alpha_{kl} \int_{-\infty}^{\infty} \varepsilon_2^{j+l} \int_{-\infty}^{\infty} \varepsilon_1^{i+k} \exp(-\varepsilon'\Sigma^{-1}\varepsilon)\, d\varepsilon_1 d\varepsilon_2.$$

Because of the definition of S, the following density integrates to 1:

$$h(\varepsilon) = h^*(\varepsilon)/S.$$

It is clear that the α parameters are identified up to a scale only, so normalization is necessary. Therefore, we impose $\alpha_{00} = 1$. For $K = 0$, $h(\varepsilon)$ now reduces to the bivariate normal density. The flexibility of this family of 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.

In the sample selection model, identification is achieved by setting $\Sigma_{22} = \sqrt{2}$, to ensure identification of the scale of (2). To fix the location of the distribution function, we optimize the log-likelihood function (4) under the restrictions that the

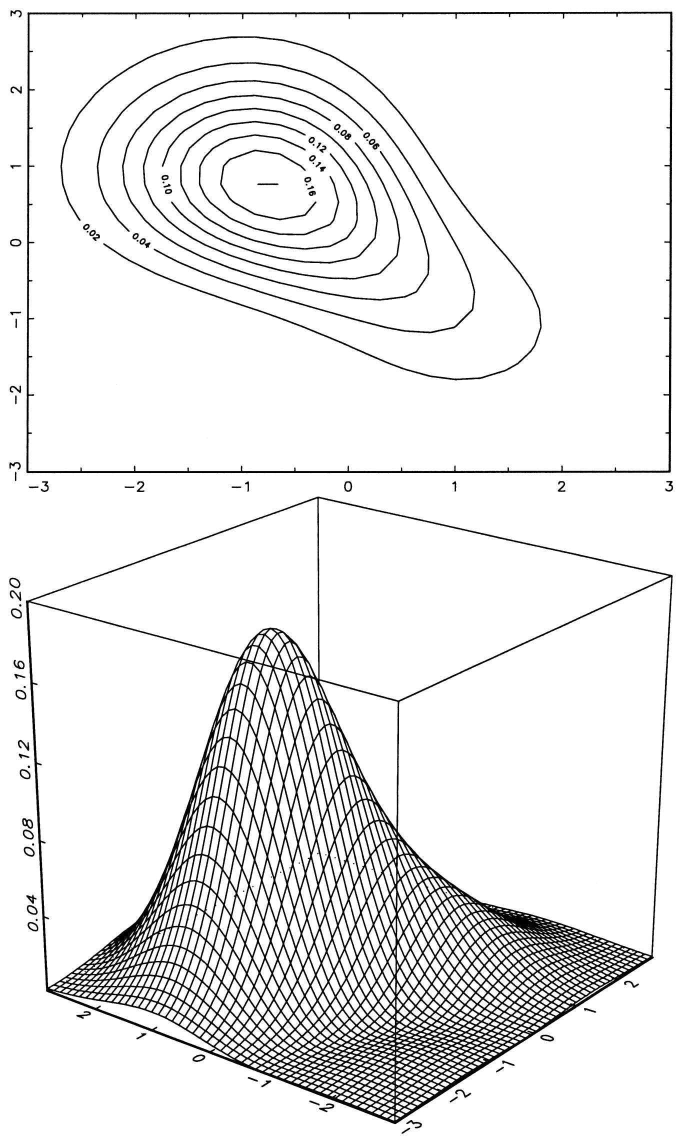

Figure 1. Bivariate Flexible Parametric Density, $\alpha_{01} = .1$, $\alpha_{10} = -.1$, and $\alpha_{11} = -.2$.

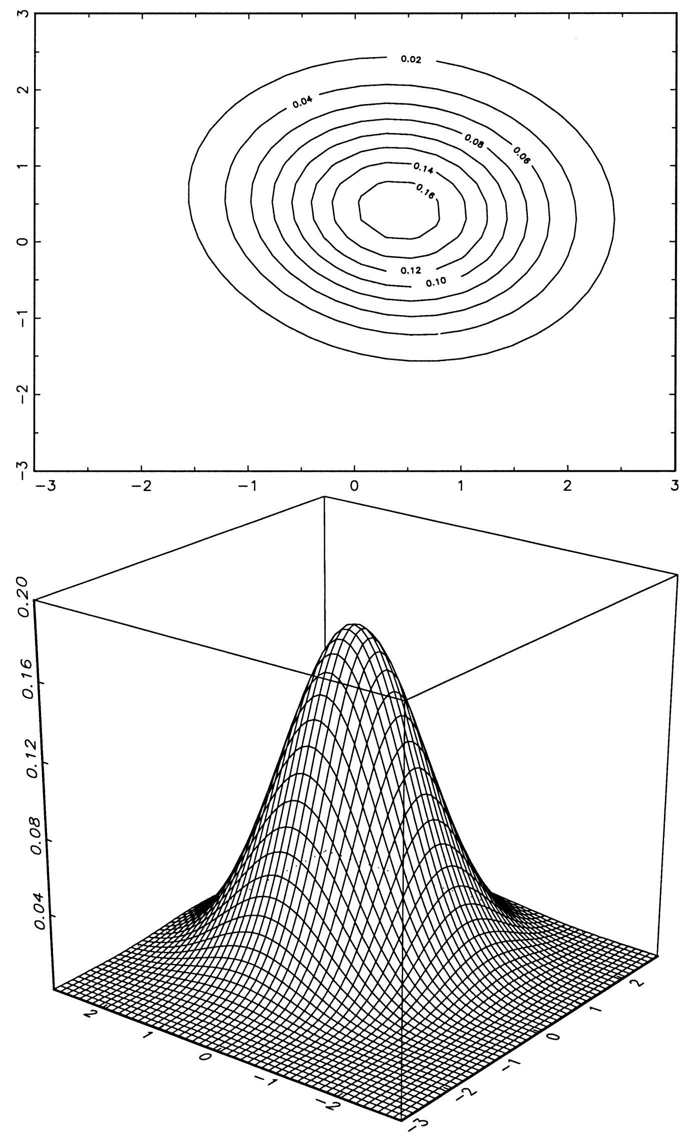

Figure 2. Bivariate Flexible Parametric Density, $\alpha_{01} = .1$, $\alpha_{10} = .1$, and $\alpha_{11} = 0$.

means of ε_1 and ε_2 equal 0. Alternatively, one could impose explicit restrictions on the α parameters to ensure that the means of ε_1 and ε_2 are 0. However, for $K \geq 2$, these are rather complex nonlinear restrictions. Another alternative was proposed by Melenberg and Van Soest (1993), who suggested not imposing restrictions on the parameters of the density function of ε to ensure a zero mean, but rather restricting the intercepts of (1) and (2). For the purpose of this article, such a strategy is not very useful. Our approach allows testing 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 mean zero 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 chi-squared distribution with $(K+1)^2 - 3$ degrees of freedom.

Optimizing the log-likelihood function subject to the means of ε_1 and ε_2 being equal to 0 requires evaluating the relevant integrals in the log-likelihood function and the expressions for the means. Because of the general covariance matrix in the base function $\phi(\varepsilon|\Sigma)$, these integrals do not have the same computationally attractive properties as they do when using the base function proposed by Gallant and Nychka (1987). However, 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. Appendix A gives the recursion formulas that can be used to determine the relevant integrals explicitly.

Finally, the flexible parametric density discussed earlier assumes homoscedasticity of the error terms. As is clear from (3), the estimators are not robust against ignored heteroscedasticity (i.e., $\rho\sigma_1$ is not the same for all observations). In the

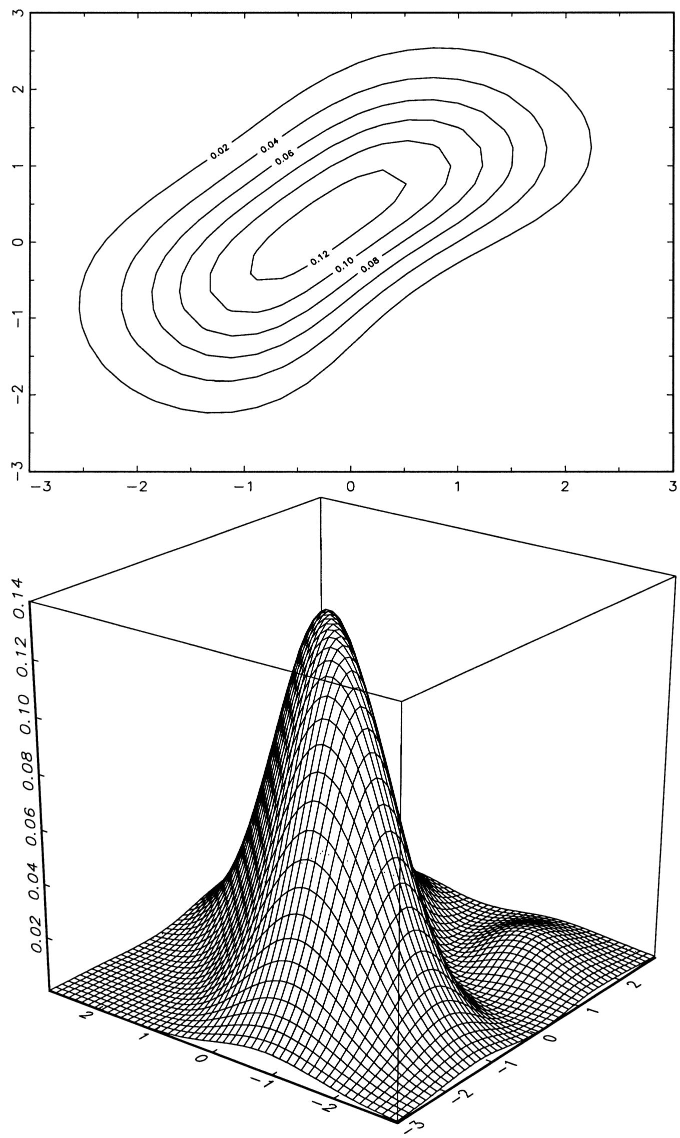

Figure 3. Bivariate Flexible Parametric Density, $\alpha_{01} = .1$, $\alpha_{10} = -.1$, and $\alpha_{11} = .2$.

next section we also discuss the performance of the flexible parametric density in case the error terms are heteroscedastic. In particular, we focus on heteroscedasticity in ε_1 when the correlation ρ between ε_1 and ε_2 is constant. Note that the variance of ε_2 is generally unidentified. Thus many types of heteroscedasticity in the distribution of ε_2 could be solved by generalizing the single index in (2).

3. SIMULATION STUDY

In this section we perform a limited simulation exercise to investigate the performance of the flexible parametric density and the power of the normality test discussed in Section 2. We focus two main sources of inconsistency, departures from normality and heteroscedasticity in the error terms.

We consider the simulation experiment

$$y_i = \beta_{10} + \beta_{11} x_i + \beta_{12} w_i + \varepsilon_{1i}$$

and

$$z_i^* = \beta_{20} + \beta_{21} v_i + \beta_{22} w_i + \varepsilon_{2i}, \qquad i = 1, \ldots, N,$$

with true parameters $\beta_{10} = 1$, $\beta_{11} = .5$, $\beta_{12} = -.5$, $\beta_{20} = 1$, $\beta_{21} = -1$, 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 five experiments in which we vary the distribution of ε. The sample size N is 1000. (Similar simulation experiments with a sample size N equal to 500 gave the same results and are available on request.) Within each experiment, we draw 100 samples. In the first three experiments, we draw ε from either a bivariate normal distribution with mean 0, a bivariate t distribution, or a centered chi-squared distribution. The t distribution has fatter tails than the normal distribution and the chi-squared distribution is asymmetric, so both these cases are a deviation from normality. For all experiments, we set $\text{var}(\varepsilon_1) = 4$, $\text{var}(\varepsilon_2) = 1$, and $\text{cov}(\varepsilon_1, \varepsilon_2) = 1$. In the last two experiments, the error terms are bivariate normal distributed but heteroscedastic. First, the distribution of ε depends on the explanatory variable x. Second, we distinguish three groups of individuals who have different covariance matrices. (This simulation experiment has some similarities with experiments in Hurd 1979.) The details of the simulation experiments are given in Appendix B. The simulations were performed on Pentium workstations using the constrained maximum likelihood (CML) library of GAUSS.

Because the computer time required for optimization of the log-likelihood function increases quickly with K, we estimate only 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 flexible parametric densities

Table 1. Results of the Simulation Experiment With Bivariate Normal Distributed Disturbances

			Flexible parametric			
$N = 1000$	Normal		$K = 1$		$K = 2$	
β_{10}	1.00	(.072)	1.00	(.073)	1.00	(.074)
β_{11}	.50	(.025)	.50	(.024)	.50	(.025)
β_{12}	−.51	(.051)	−.51	(.049)	−.51	(.049)
β_{20}	1.00	(.060)	1.00	(.062)	1.00	(.061)
β_{21}	−.99	(.045)	−.99	(.043)	−.99	(.043)
β_{22}	.99	(.054)	.99	(.054)	.99	(.054)
σ_1	2.00	(.060)	2.00	(.059)	2.00	(.059)
σ_{12}	1.00	(.034)	1.00	(.036)	1.00	(.036)
α_{00}			1		1	
α_{01}			0		.0001	(.019)
α_{02}					.0026	(.012)
α_{10}			0		.0003	(.010)
α_{11}			−.0002	(.0064)	.0034	(.017)
α_{12}					.0006	(.017)
α_{20}					−.0004	(.0051)
α_{21}					−.0005	(.0067)
α_{22}					−.0014	(.0050)
sdev(ε_1)			2.00	(.061)	2.00	(.070)
sdev(ε_2)			1.00	(.013)	1.00	(.020)
cov($\varepsilon_1, \varepsilon_2$)			1.00	(.087)	1.01	(.10)
Log-likelihood	−1412.71		−1412.35		−1411.18	
Rejections			0		1	

NOTE: Standard errors are given in parentheses.

Table 2. Results of the Simulation Experiment With Bivariate t Distributed Disturbances

			Flexible parametric			
N = 1000	Normal		K = 1		K = 2	
β_{10}	.99	(.076)	1.03	(.064)	.98	(.040)
β_{11}	.50	(.028)	.50	(.029)	.50	(.019)
β_{12}	−.51	(.055)	−.51	(.044)	−.51	(.026)
β_{20}	1.06	(.11)	1.05	(.071)	1.08	(.055)
β_{21}	−1.05	(.096)	−1.02	(.065)	−1.08	(.059)
β_{22}	1.07	(.11)	1.04	(.076)	1.09	(.065)
σ_1	1.92	(.22)	2.03	(.19)	2.05	(.096)
σ_{12}	.95	(.22)	1.18	(.22)	.97	(.053)
α_{00}			1		1	
α_{01}			0		.012	(.019)
α_{02}					−.078	(.062)
α_{10}			0		.0025	(.018)
α_{11}			−.12	(.082)	−.014	(.037)
α_{12}					.0099	(.028)
α_{20}					−.047	(.012)
α_{21}					−.0068	(.013)
α_{22}					.013	(.0049)
sdev(ε_1)			1.88	(.18)	2.00	(.098)
sdev(ε_2)			.93	(.059)	1.00	(.030)
cov($\varepsilon_1, \varepsilon_2$)			.72	(.32)	.94	(.16)
Log-likelihood	−1371.30		−1348.14		−1331.18	
Rejections			69		97	

NOTE: Standard errors are given in parentheses.

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 Table 1; the case of the t disturbances, in Table 2; and the case of the chi-squared disturbances, in Table 3. The simulation results from the heteroscedastic error terms are presented in Table 4 for the covariance matrix dependent on the observed variable x and in Table 5 for the covariance matrix dependent on an unobserved variable.

It is remarkable how well standard maximum likelihood under the assumption of normally distributed disturbances performs. Departures from normality do not cause serious bias

Table 3. Results of the Simulation Experiment With Bivariate Chi-Squared Distributed Disturbances

			Flexible parametric			
N = 1000	Normal		K = 1		K = 2	
β_{10}	.94	(.089)	.93	(.047)	.95	(.061)
β_{11}	.50	(.029)	.50	(.027)	.50	(.018)
β_{12}	−.50	(.054)	−.50	(.036)	−.49	(.021)
β_{20}	1.09	(.075)	1.05	(.039)	1.12	(.041)
β_{21}	−1.04	(.076)	−1.04	(.037)	−1.06	(.028)
β_{22}	1.03	(.083)	1.04	(.037)	1.06	(.024)
σ_1	2.03	(.084)	2.01	(.053)	1.89	(.045)
σ_{12}	1.28	(.15)	1.27	(.053)	1.19	(.020)
α_{00}			1		1	
α_{01}			0		−.11	(.021)
α_{02}					−.039	(.030)
α_{10}			0		−.088	(.014)
α_{11}			−.018	(.032)	.066	(.031)
α_{12}					−.0067	(.032)
α_{20}					−.046	(.019)
α_{21}					.037	(.017)
α_{22}					.0034	(.0037)
sdev(ε_1)			1.94	(.074)	1.61	(.042)
sdev(ε_2)			.97	(.038)	.85	(.012)
cov($\varepsilon_1, \varepsilon_2$)			1.12	(.17)	.56	(.066)
Log-likelihood	−1403.24		−1391.81		−1345.13	
Rejections			89		100	

NOTE: Standard errors are given in parentheses.

Table 4. Results of the Simulation Experiment, Where the Error Terms are Distributed According to a Bivariate Normal Distribution With Heteroscedasticity Depending on the Explanatory Variable x

			Flexible parametric			
N = 1000	Normal		K = 1		K = 2	
β_{10}	1.04	(.066)	1.02	(.068)	1.00	(.058)
β_{11}	.55	(.068)	.50	(.051)	.51	(.051)
β_{12}	−.51	(.063)	−.51	(.047)	−.50	(.044)
β_{20}	.99	(.059)	1.01	(.086)	1.28	(.10)
β_{21}	−1.00	(.047)	−.92	(.063)	−1.26	(.097)
β_{22}	1.00	(.050)	.93	(.072)	1.26	(.094)
σ_1	2.10	(.39)	2.21	(.38)	2.17	(.44)
σ_{12}	.58	(.071)	1.51	(.31)	.95	(.40)
α_{00}			1		1	
α_{01}			0		.012	(.021)
α_{02}					.14	(.15)
α_{10}			0		.016	(.033)
α_{11}			−.17	(.032)	−.054	(.12)
α_{12}					−.0027	(.013)
α_{20}					−.086	(.035)
α_{21}					−.0023	(.0077)
α_{22}					.0096	(.013)
sdev(ε_1)			2.08	(.33)	2.25	(.64)
sdev(ε_2)			.94	(.044)	1.29	(.18)
corr($\varepsilon_1, \varepsilon_2$)	.29	(.064)	.69	(.084)	.44	(.14)
cov($\varepsilon_1, \varepsilon_2$)			1.06	(.39)	1.40	(1.24)
Log-likelihood	−1436.58		−1362.86		−1315.80	
Rejections			100		100	

NOTE: Standard errors are given in parentheses.

Table 5. Results of the Simulation Experiment, Where the Error Terms are Distributed According to a Bivariate Normal Distribution With Heteroscedasticity Depending on an Unobserved Variable r Taking Three Possible Values, $\sqrt{.5}$, 1, and $\sqrt{1.5}$

			Flexible parametric			
N = 1000	Normal		K = 1		K = 2	
β_{10}	1.01	(.060)	1.01	(.035)	1.00	(.031)
β_{11}	.50	(.027)	.50	(.026)	.50	(.027)
β_{12}	−.50	(.047)	−.50	(.037)	−.50	(.036)
β_{20}	.99	(.054)	1.00	(.032)	1.00	(.027)
β_{21}	−1.00	(.047)	−1.00	(.033)	−1.00	(.029)
β_{22}	1.00	(.046)	1.00	(.031)	1.00	(.026)
σ_1	1.99	(.057)	2.00	(.036)	2.00	(.012)
σ_{12}	1.00	(.098)	1.00	(.023)	1.00	(.013)
α_{00}			1		1	
α_{01}			0		.0013	(.020)
α_{02}					−.015	(.032)
α_{10}			0		−.0004	(.011)
α_{11}			−.0013	(.014)	−.015	(.031)
α_{12}					.0063	(.026)
α_{20}					−.0074	(.0096)
α_{21}					−.0030	(.012)
α_{22}					.0027	(.0067)
sdev(ε_1)			1.99	(.052)	2.01	(.25)
sdev(ε_1)			1.00	(.028)	1.02	(.13)
corr($\varepsilon_1, \varepsilon_2$)	.50	(.050)	.50	(.020)	.50	(.0091)
cov($\varepsilon_1, \varepsilon_2$)			.99	(.15)	1.09	(.25)
Log-likelihood	−1405.10		−1405.09		−1402.10	
Rejections			0		12	

NOTE: Standard errors are given in parentheses.

in the parameter estimates. In case the true disturbances follow a t distribution, all estimated parameters are within two standard deviations of their true values. This is the case even when the disturbances follow a transformed chi-squared distribution that is nonsymmetric. This suggests that if obtaining estimates for the parameters β is the only interest, than it is sufficient to assume normality and obtain estimates. However, in the presence of heteroscedasticity in the error terms that is dependent on one of the observed exogenous variables, the parameter estimates under normality are somewhat biased (see Table 4). In particular, the parameter estimate of β_{11} differs from the true value. This bias disappears when estimating the model using the flexible parametric specification. When the heteroscedasticity does not depend on exogenous variables but the covariance matrix differs between groups, the parameter estimates obtained under normality are again close to their true value.

Tables 1–5 also report the number of rejections of normality for each simulation experiment. The simulation results indicate that the 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. The number of correct rejections is very high except in the simulation experiment, where the distribution of the error terms differs between groups. The test is very powerful when the error terms are not bivariate normal distributed, and it is successful in detecting heteroscedasticity depending on observed exogenous variables. The simulation results indicate that the test has more power when the fitted distribution belongs to $\mathcal{H}_2^*$ than when it belongs to $\mathcal{H}_1^*$. Because departures from normality do not cause serious bias in the parameter estimates of the regression equation, Hausman tests for normality against a semiparametric alternative probably will not be as powerful as the likelihood-based tests discussed in this article.

The results from the simulation study suggest that the estimates for the β_1 parameters (derived under the assumption of normality) are reasonably robust to misspecification of the error term. This is not true only if the covariance matrix of the error terms depends on observed explanatory variables. In this case, using the flexible parametric density can help improve the parameter estimates. Furthermore, if one is interested not only in the parameter estimates of β_1, but also in the complete distribution function of outcome variable y conditional on explanatory variables x, then using the flexible parametric specification is worthwhile.

4. AN APPLICATION TO A MODEL OF TRAVEL DEMAND

In this section we apply the flexible parametric estimation procedure and the normality tests to a model describing travel demand. 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 means. The latter is most likely public transportation; other possibilities may be bikes, car pooling, and so forth. 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 taxes for car ownership. Variable costs are the costs of fuel. The costs of other transportation do not contain any fixed costs. The variable costs, denoted by v_o, are associated with the costs of public transportation. However, opportunity costs may also be important, because traveling with public transportation usually 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

$$\mathrm{E}[c_c + v_c y] \le \mathrm{E}[v_o y]. \tag{6}$$

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

$$\mathrm{E}[y] \ge \frac{c_c}{v_o - v_c}. \tag{7}$$

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

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

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

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

where v_1 and v_2 denote unobserved individual specific effects. We normalize $\mu = 1$.

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

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

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

$$x'\beta - z'\gamma + \varepsilon_2 \ge 0,$$

with the household-specific term ε_2 equal to $v_1 - v_2$, which is known to the household but 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 impose an exclusion restriction, that is, find a variable that 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, whereas the fixed and variable costs of using a car and the variable costs of other transportation affect only the decision of whether to use a car. Thus the exclusion restriction should be a variable that affects only the costs functions. Because the Dutch government provides free public transportation passes to some students and some civil servants, we use the dummy variable that indicates whether the household has such a free public transportation 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 34,454 households. To avoid complications of households owning more than one car, we focus on single-person households. Furthermore, we exclude individuals younger than age 18 years, the legal age for obtaining a drivers license in The Netherlands. This restricts the database to 7404 individuals. To construct our final dataset, we also exclude 130 individuals who own a car but for whom the number of kilometers driven in the past year is unknown and also 756 individuals for whom one or more explanatory variables are missing. Our final dataset consists of 6518 observations.

Table 6 provides some characteristics of the dataset. 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. Although 58% of the men are car owners, only 41% of the women have a car. Until people reach age 65, the car ownership rate increases with age; after that, there is a large drop. Furthermore, car ownership rates increase with income and 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 transportation 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 to either $\mathcal{H}_1^*$ or $\mathcal{H}_2^*$. For these two cases, we test whether the restriction of normality of the disturbances can be imposed.

All parameter estimates are presented in Tables 7 and 8. The parameter estimates of the distance function (Table 7) and of the costs function (Table 8) are not very sensitive to the specification of the distribution function of the error term, which is in line with the results from the simulation study. The most important covariates in the distance function are gender, age, income, level of education, and individual labor market status; those in the costs function are age, degree of urbanization, and income. Although the availability of a free public transportation 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 important in both the distance and costs functions are age and income. Age affects both the distance function and the costs function negatively. Young people travel more kilometers than older people, whereas the costs of car use relative to other transportation modes decrease with age. By comparing the coefficients, we can see that persons age 50–64 are most likely to own a car. Income has an opposite effect on the distance and costs functions. Persons with greater income travel more kilometers, whereas the costs of car use relative to other transportation decrease. Considering that using public transportation takes more time, this latter effect can be explained by differences in opportunity costs. Time is more costly for individuals with high incomes. The degree of urbanization affects only the costs function. Because less public transportation is available in areas with low degrees of urbanization, the costs of using public transportation are higher in these areas (e.g., waiting times are longer). In contrast, car use in areas with a high degree of urbanization is more costly, because individuals often incur additional costs, such as parking costs. Employed individuals, both full-time and part-time, travel more kilometers. Because we do not distinguish between traveling for private purposes and traveling for professional purposes, this may be caused by commuting or because their work requires them to travel. Finally, persons with a higher level of education travel more than those with less education.

Table 6. Some Characteristics of the Dataset

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
Net income (guilders)		
0–15,000	158	999
15,000–23,000	411	1046
23,000–30,000	539	574
30,000–38,000	695	402
38,000–52,000	773	278
52,000+	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	15,093	
Observations	3110	3408

Table 7. Estimation Results of the Travel Demand Model: Distance Function

	Distance function (β)					
			Flexible parametric			
	Normal		$K=1$		$K=2$	
Intercept	9.41	(.32)	9.77	(.11)	9.87	(.15)
Gender						
Female	−.29	(.038)	−.24	(.022)	−.23	(.026)
Age						
25–29	−.13	(.088)	−.14	(.061)	−.15	(.066)
30–39	−.24	(.084)	−.23	(.059)	−.23	(.064)
40–49	−.28	(.089)	−.30	(.062)	−.30	(.067)
50–64	−.33	(.10)	−.35	(.064)	−.36	(.071)
65+	−.59	(.086)	−.50	(.067)	−.50	(.073)
Region						
North	.0027	(.050)	.0066	(.045)	.0010	(.047)
East	−.0007	(.031)	.013	(.027)	.016	(.028)
South	−.054	(.040)	−.057	(.034)	−.061	(.034)
Degree of urbanization						
High	.025	(.042)	−.0070	(.032)	−.012	(.033)
Average	.064	(.066)	−.0088	(.035)	−.035	(.039)
Low	.077	(.073)	−.0043	(.038)	−.033	(.043)
Very low	.064	(.082)	−.011	(.044)	−.037	(.050)
Net income (guilders)						
15,000–23,000	.086	(.075)	−.12	(.061)	.022	(.067)
23,000–30,000	.23	(.11)	.089	(.065)	.093	(.075)
30,000–38,000	.30	(.14)	.12	(.066)	.11	(.081)
38,000–52,000	.37	(.16)	.17	(.069)	.15	(.086)
52,000+	.50	(.18)	.30	(.072)	.27	(.090)
Level of education						
Lower secondary	.10	(.051)	.043	(.040)	.035	(.044)
Higher secondary	.17	(.062)	.097	(.041)	.085	(.046)
University	.28	(.064)	.21	(.043)	.20	(.047)
Labor market status						
Part-time work	.017	(.058)	.023	(.050)	.035	(.052)
Student	−.28	(.11)	−.19	(.10)	−.16	(.12)
Unemployed	−.29	(.069)	−.18	(.059)	−.17	(.063)
Nonparticipant	−.25	(.046)	−.23	(.037)	−.22	(.039)
σ_1	.63	(.0088)	.71	(.018)	.76	(.23)
σ_{12}	.041	(.17)	−.45	(.034)	−.54	(.31)
α_{00}			1		1	
α_{01}			0		−.0035	(.049)
α_{02}					.038	(.12)
α_{10}			0		−.012	(.045)
α_{11}			.49	(.031)	.38	(.41)
α_{12}					.15	(.12)
α_{20}					−.052	(.14)
α_{21}					.16	(.079)
α_{22}					.0005	(.098)

NOTE: Standard errors are in parentheses.

The likelihood ratio test statistics of normality against the flexible parametric density with $K = 1$ and $K = 2$ equal 283.7 and 299.6, implying that we must reject the hypothesis that the disturbances follow a normal distribution. The main difference with the estimates under normality is that we do not find any correlation between ε_1 and ε_2 under the assumption of normality. This suggests that there is no unobserved selection between car ownership and car use, which implies that both the regression and selection equations can be estimated consistently separately of each other by, for example, OLS and probit. The flexible parametric densities show a correlation between both disturbance terms that could not be captured by a normal density. The estimated covariance between ε_1 and ε_2 is $-.23$ for $K = 1$ and $-.30$ for $K = 2$. In the context of our theoretical model, this implies a positive correlation between v_1 and v_2, because $\text{cov}(\varepsilon_1, \varepsilon_2) = \text{var}(v_1) - \text{cov}(v_1, v_2)$. Obviously, there are some unobserved covariates, which increase both the costs of car use relative to other transportation and the expected number of kilometers traveled. Thus such a covariate has similar effects to, 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.

Table 8. Estimation Results of the Travel Demand Model: Costs Function

	Costs function (γ)					
			Flexible parametric			
	Normal		$K=1$		$K=2$	
Intercept	10.81	(.35)	10.86	(.14)	10.91	(.29)
Gender						
Female	.046	(.053)	.017	(.038)	.047	(.095)
Age						
25–29	−.49	(.13)	−.41	(.10)	−.40	(.13)
30–39	−.56	(.13)	−.46	(.10)	−.46	(.12)
40–49	−.65	(.13)	−.58	(.11)	−.58	(.13)
50–64	−1.00	(.15)	−.85	(.11)	−.86	(.18)
65+	−.76	(.14)	−.61	(.12)	−.59	(.13)
Region						
North	−.11	(.092)	−.075	(.079)	−.077	(.088)
East	.049	(.058)	.060	(.050)	.065	(.055)
South	−.16	(.072)	−.13	(.062)	−.13	(.070)
Degree of urbanization						
High	−.18	(.068)	−.18	(.056)	−.18	(.076)
Average	−.51	(.086)	−.47	(.057)	−.49	(.15)
Low	−.57	(.096)	−.53	(.064)	−.56	(.17)
Very low	−.65	(.11)	−.59	(.078)	−.61	(.19)
Net income (guilders)						
15,000–23,000	−.21	(.10)	−.26	(.085)	−.20	(.11)
23,000–30,000	−.48	(.14)	−.48	(.089)	−.46	(.17)
30,000–38,000	−.72	(.16)	−.68	(.090)	−.69	(.24)
38,000–52,000	−.90	(.18)	−.83	(.093)	−.86	(.30)
52,000+	−1.04	(.20)	−.90	(.10)	−.95	(.37)
Level of education						
Lower secondary	−.16	(.079)	−.16	(.066)	−.16	(.089)
Higher secondary	−.26	(.089)	−.24	(.068)	−.25	(.12)
University	−.15	(.095)	−.12	(.075)	−.13	(.12)
Labor market status						
Part-time work	.15	(.095)	.13	(.080)	.15	(.094)
Student	.14	(.19)	.13	(.16)	.16	(.19)
Unemployed	.058	(.11)	.11	(.095)	.13	(.14)
Nonparticipant	−.023	(.082)	−.054	(.068)	−.040	(.094)
Free public transport pass						
Yes	.21	(.16)	.14	(.12)	.15	(.13)
Log-likelihood	−6386.34		−6244.49		−6236.54	

NOTE: Standard errors are in parentheses.

Figures 4 and 5 show the marginal densities of the disturbances ε_1 and ε_2 estimated under the flexible parametric densities and normality. Both flexible parametric densities have more mass close to the mode and slightly fatter tails than the normal density. The estimated standard deviation of ε_1 is .64 in both the $K = 1$ and $K = 2$ specifications, almost similar to the estimate under normality. For the standard deviation of ε_2, we find .89 and .90 for ε_2 for the $K = 1$ and the $K = 2$ density. Figures 6 and 7 graph the estimated densities of the normal model and the model estimated using an flexible parametric density ($K = 2$). We see that the estimated flexible parametric density is a bit more spread out than the normal density, and that the estimated contour lines in Figure 7 are not the ellipsoids of Figure 6.

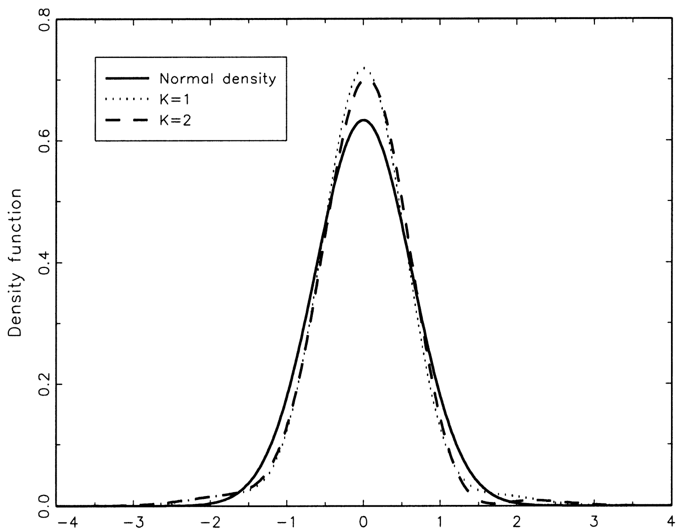

Figure 4. The Estimated Marginal Density Function of ε_1.

5. CONCLUSION

In this article we have derived a test for normality in sample selection models. To our knowledge, no such test has been derived previously. The test is motivated on the seminonparametric maximum likelihood method introduced by Gallant and Nychka (1987). A flexible parametric density function is used to approximate the density function of the error terms. The bivariate normal density is a special case of this flexible parametric density function that allows us to test for normality. The test also provides an alternative density function in the event that normality is rejected.

Although the simulation study provided in this article is limited, we believe that this test is promising. The test performed well in almost all simulation experiments, the percentage of incorrect rejections is below the significance level, whereas the power is high. This is not true only when the distribution of the error terms is heteroscedastic depending on an unobserved variable. The test is powerful against departures from normality and against heteroscedasticity depending on an observed exogenous variable. However, it should be noted that in most cases the parameter estimates do not seem very sensitive to the distributional assumptions of the disturbances. Even in the event that the disturbances do not follow a normal distribution, maximum likelihood under the assumption of normality provides estimates close to the true values. Only if the covariance matrix of the error terms depends on an observed exogenous variable are the parameter estimates biased. The simulation study indicates that the bias disappears when using the flexible parametric density. The flexible parametric density is thus useful in cases where not only parameter estimates, but also the complete distributional structure are of interest, and in the case of heteroscedasticity.

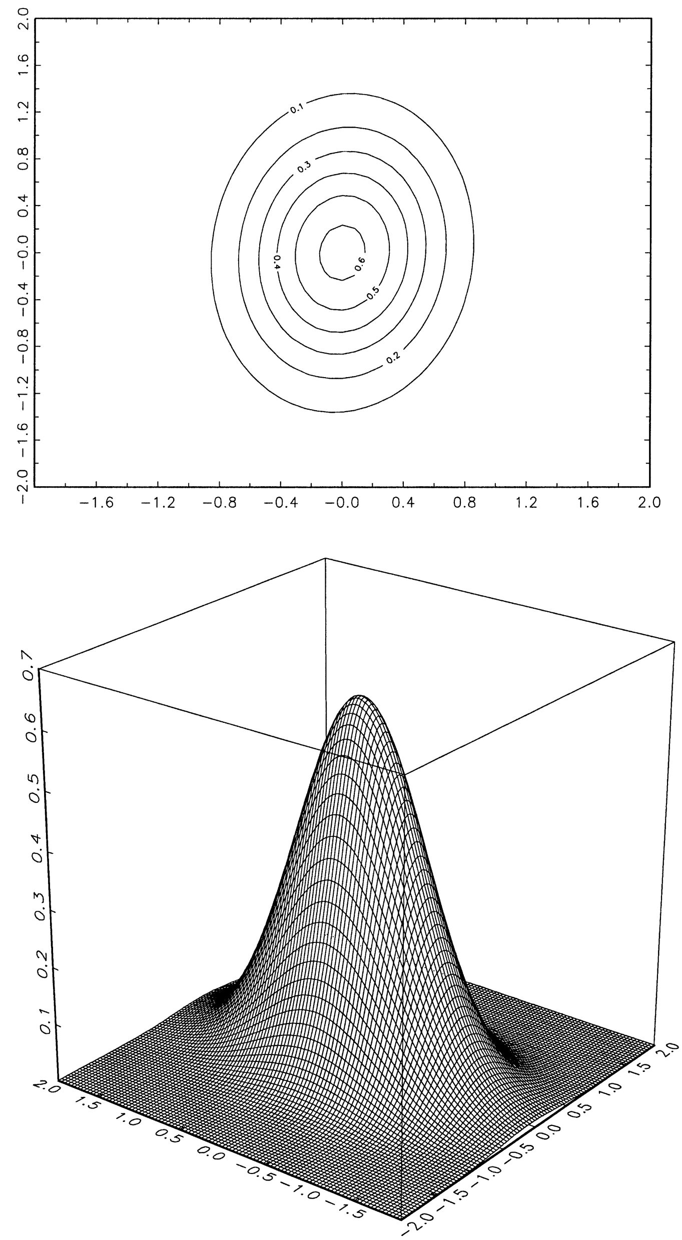

Figure 6. Estimated Density Under the Normality Assumption.

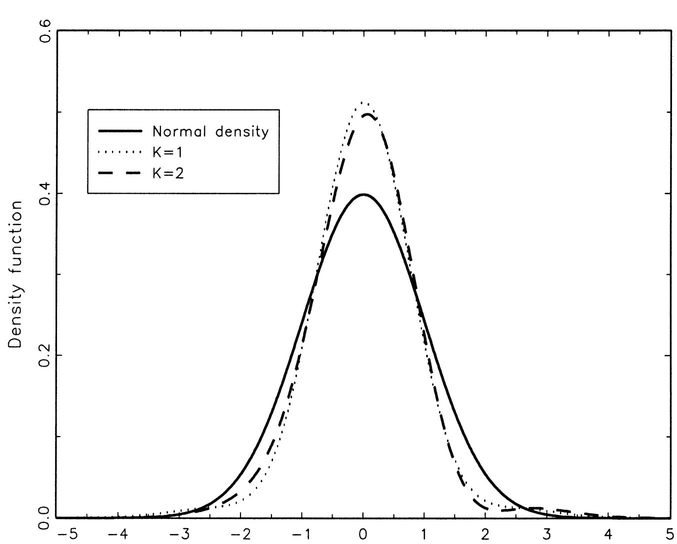

Figure 5. The Estimated Marginal Density Function of ε_2.

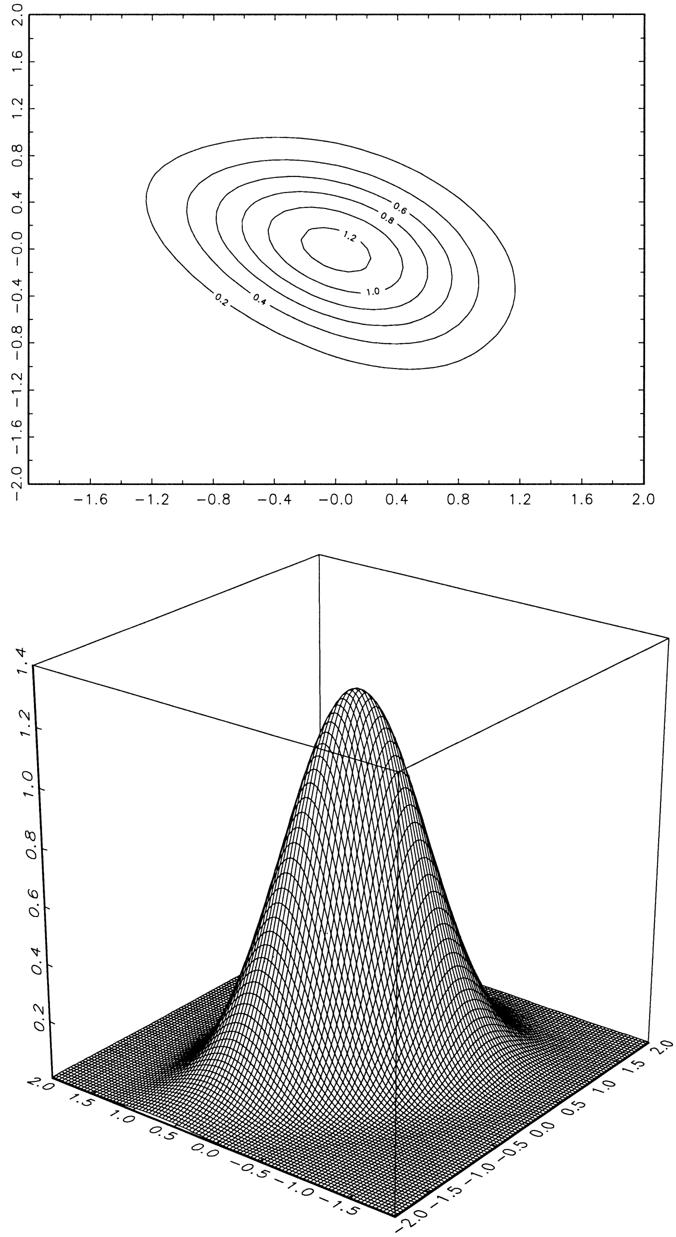

Figure 7. Estimated Density $K = 2$.

Finally, we have applied the normality test to a model of travel demand. 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 flexible parametric density is capable of correcting for the unobserved selectivity, which is not captured by the normal density.

ACKNOWLEDGMENTS

The authors thank Christopher Ferrall, Jean-Francois Lamarche, Jeffrey Wooldridge, an anonymous associate editor, and an anonymous referee for useful comments and suggestions. The Gauss code for the estimation procedure discussed in this article can be downloaded from *www.tinbergen.nl/∼klaauw/normal.htm.*

APPENDIX A: RELEVANT INTEGRALS OF THE FLEXIBLE PARAMETRIC DENSITY IN THE SAMPLE SELECTION MODEL

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

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

and

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

This latter integral also appears in S when $b = \infty$ and in the mean restrictions where also $b = \infty$ and $h^*(\varepsilon)$ is replaced by $\varepsilon_i h^*(\varepsilon)$ for $i = 1, 2$. (Because of the structure of the Hermite series, this solves similarly.)

Substituting for $h^*(\varepsilon)$, 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

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

The last integral can be solved easily because

$$\mathrm{E}(\varepsilon_1^i|\varepsilon_2) = \gamma_0 + \gamma_1\varepsilon_2 + \cdots + \gamma_i\varepsilon_2^i.$$

The coefficients γ depend only on the other parameters of the density function, and they are independent of ε_2.

Both integrals can be calculated using the recursion formulas that follow. Therefore, we define $I_k(a, b)$ as the univariate integral

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

Using partial integration, we obtain the recursion formulas

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

if $k \geq 2$; otherwise,

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

and

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

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

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

APPENDIX B: THE SIMULATION EXPERIMENTS

The simulation experiments in Section 3 were performed as follows. For all experiments, we impose $\mathrm{var}(\varepsilon_1) = 4$, $\mathrm{var}(\varepsilon_2) = 1$, and $\mathrm{cov}(\varepsilon_1, \varepsilon_2) = 1$.

In the first simulation experiments, we draw ε from either a bivariate normal distribution with mean 0, a bivariate t distribution, or a centered chi-squared distribution. Drawing from the bivariate normal distribution is straightforward. In the bivariate t distribution, $\varepsilon_2 = u_1/\sqrt{3}$ and $\varepsilon_1 = \varepsilon_2 + u_2$, with u_1 and u_2 independent draws from a t_3 distribution. For the bivariate chi-squared distribution, $\varepsilon_2 = \frac{1}{2}(u_1^2 + u_2^2) - 1$ and $\varepsilon_1 = \varepsilon_2 + (u_3^2 + u_4^2 + u_5^2)/\sqrt{2} - 3/\sqrt{2}$, with u_1^2–u_5^2 independent $\chi^2(1)$ random variates.

In the last two experiments, the error terms are bivariate normal distributed but heteroscedastic. First, the distribution of ε depends on the explanatory variable x. We define the weight $r = \frac{1}{2}\exp(\sqrt{\log(4)/2}\cdot x/\sqrt{3})$. The error term ε_2 is drawn from a standard normal distribution and $\varepsilon_1 = r(\varepsilon_2 + \sqrt{3}u)$, where u is a realization from a standard normal distribution. For each individual, the correlation between ε_1 and ε_2 is equal to .5 and thus is independent of the value of x. Furthermore, $\mathrm{E}_x[\mathrm{var}(\varepsilon_1|x)] = 4$ and $\mathrm{E}_x[\mathrm{var}(\varepsilon_2|x)] = \mathrm{E}_x[\mathrm{cov}(\varepsilon_1, \varepsilon_2|x)] = 1$.

Second, we distinguish three groups of individuals who have different covariance matrices. The unobserved weight r can take three values, $\sqrt{.5}$, 1, and $\sqrt{1.5}$. The population is over these three values. Here ε_2 is again drawn from a standard normal distribution, and $\varepsilon_1 = r(\varepsilon_2 + \sqrt{3}u)$, with u a realization of a standard normal distribution that is independent of ε_2.

[Received September 2000. Revised October 2001.]

REFERENCES

Abadie, A., Angrist, J. D., and Imbens, G. W. (2002), "Instrumental Variables Estimation of Quantile Treatment Effects," *Econometrica*, 70, 91–117.

Ahn, H., and Powell, J. L. (1993), "Semiparametric Estimation of Censored Selection Models With a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58, 3–29.

Buchinsky, M. (1998), "The Dynamics of Changes in the Female Wage Distribution in the U.S.A.: A Quantile Regression Approach, *Journal of Applied Econometrics*, 13, 1–30.

Cameron, A. C., and Johansson, P. (1997), "Count Data Regression Using Series Expansions With Applications," *Journal of Applied Econometrics*, 12, 203–223.

Cosslett, S. R. (1991), "Semiparametric Estimation of a Regression Model With Sample Selectivity," in *Nonparametric and Semiparametric methods in Econometrics and Statistics*, eds. W. A. Barnett, J. L. Powell, and G. Tauchen, Cambridge, U.K.: Cambridge University Press.

Gabler, S., Laisney, F., and Lechner, M. (1993), "Semi-Nonparametric Maximum Likelihood Estimation of Binary Choice Models With an Application to Labour Force Participation," *Journal of Business & Economic Statistics*, 11, 61–80.

Gallant, A. R., and Nychka, D. W. (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.

——— (1990), "Varieties of Selection Bias," *American Economic Review Papers and Proceedings*, 80, 313–318.

Heckman, J. J., LaLonde, R. J., and Smith, J. A. (1999), "The Economics and Econometrics of Active Labor Market Programs," in *Handbook of Labor Economics*, Vol. 3, eds. O. Ashenfelter and D. Card, Amsterdam: North-Holland.

Heckman, J. J., Tobias, J. L., and Vytlacil, E. (2000), "Simple Estimators for Treatment Parameters in a Latent Variable Framework With an Application to Estimating the Returns to Education," Working Paper 7950, NBER.

Hurd, M. (1979), "Estimation in Truncated Samples When There is Heteroskedasticity," *Journal of Econometrics*, 11, 247–258.

Ichimura, H., and Lee, L.-F. (1991), "Semi-Parametric Least Squares Estimation of Multiple Index Models: Single Equation Estimation," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, eds. W. A. Barnett, J. L. Powell, and G. Tauchen, Cambridge, U.K.: Cambridge University Press.

Lee, L.-F. (1978), "Unionism and Wage Rates: A Simultaneous Equations Model With Qualitative and Limited Dependent Variables," *International Economic Review*, 19, 415–433.

——— (1982), "Some Approaches to the Correction of Selectivity Bias," *Review of Economic Studies*, 49, 355–372.

Manski, C. F. (1989), "Anatomy of the Selection Problem," *Journal of Human Resources*, 24, 343–360.

Melenberg, B., and van Soest, A. (1993), "Semi-Parametric Estimation of the Sample Selection Model," working paper, CentER, Tilburg.

Newey, W. K., (1999a), "Consistency of the Two-Step Sample Selection Estimators Despite Misspecification of Distribution," *Economics Letters*, 63, 129–132.

——— (1999b), "Two-Step Series Estimation of Sample Selection Models," working paper, Massachusetts Institute of Technology.

Smith, R. J. (1989), "On the Use of Distributional Misspecification Checks in Limited Dependent Variable Models," *Economic Journal Supplement*, 99, 178–192.

Tunalı, İ. (2000), "Rationality of Migration," *International Economic Review*, 41, 893–920.