A practical test for the choice of mixing distribution in discrete choice models

Fosgerau, Mogens and Bierlaire, Michel

Technical University of Denmark

2007

A practical test for the choice of mixing distribution in discrete choice models

Mogens Fosgerau

Danish Transport Research Institute, Denmark.

Michel Bierlaire

Ecole Polytechnique Fédérale de Lausanne, Transport and Mobility Laboratory, Switzerland.

Abstract

The choice of a specific distribution for random parameters of discrete choice models is a critical issue in transportation analysis. Indeed, various pieces of research have demonstrated that an inappropriate choice of the distribution may lead to serious bias in model forecast and in the estimated means of random parameters. In this paper, we propose a practical test, based on seminonparametric techniques. The test is analyzed both on synthetic and real data, and is shown to be simple and powerful.

Key words: mixed logit, random parameters, nonparametric, seminonparametric, hypothesis testing

1 Introduction

An important advantage of discrete choice models to analyze transportation demand is their disaggregate nature, allowing them to capture heterogeneity in the population under interest. The analyst identifies segments, typically characterized by socio-economic characteristics such as income, age, or gender, or by the choice context, defined for example by the trip purpose. Each segment must be sampled in order to have a sufficient amount of data to estimate statistically significant models. However, even after controlling for observable

Email addresses: mf@dtf.dk (Mogens Fosgerau), michel.bierlaire@epfl.ch (Michel Bierlaire).

characteristics, there is as a rule lots of heterogeneity left. This remaining heterogeneity can be accounted for with random parameters.

Except for some specific models, the error structure of the resulting model becomes very complicated, and cannot be represented by a closed form probability model. We obtain mixtures of models, where the underlying choice probability, conditional on the value of the parameters β is written as

$$P_n(i|\mathcal{C}_n, \beta) \tag{1}$$

where $\mathcal{C}_n$ is the choice set of decision-maker n, $i \in \mathcal{C}_n$ is the alternative under consideration and $\beta \in \mathbb{R}^K$ is a vector of parameters. Assuming that β is randomly distributed with PDF $f(\cdot)$, the mixture of models is defined by

$$P_n(i|\mathcal{C}_n) = \int_\beta P_n(i|\mathcal{C}_n, \beta) f(\beta) d\beta. \tag{2}$$

In practice, the kernel choice model (1) is often a Multinomial Logit (MNL) model, but any closed form probability model (such as Generalized Extreme Value models) is adequate. Although proposed about 30 years ago (see for instance Electric Power Research Institute, 1977), the use of mixtures of MNL models (MMNL) has only become popular more recently thanks to the increasing power of computers allowing for the systematic use of Monte-Carlo simulation to approximate the complex error structures of these models (see, among many others, discussions by Revelt and Train, 1998; Train, 1998; McFadden and Train, 2000; Ben-Akiva et al., 2001; Hensher and Greene, 2003; Train, 2003; Viton, 2004).

This modeling approach has been found to be very useful to capture many complex transportation phenomena, such as the analysis of the value of travel time (e.g. Algers et al., 1998; Hess, Bierlaire, and Polak, 2005; Greene et al., 2006) and reliability (e.g. Brownstone and Small, 2005; Small et al., 2005), route choice (see Han et al., 2001; Bekhor et al., 2002; Frejinger and Bierlaire, 2007), airport choice (Hess and Polak, 2005), airline choice (Carrier, 2003), vehicle choice (Brownstone et al., 2000; Hess et al., 2006), and congestion pricing (Bhat and Castelar, 2002).

An important issue is the choice of a specific distribution for the random parameters. Actually, various pieces of research have demonstrated that an inappropriate choice of the distribution may lead to serious bias in model forecast and in the estimated mean of random parameters. A noticeable example is the Normal distribution, used as a default for many applications. Hess et al. (2005) discuss wrong interpretations of willingness-to-pay indicators when normal distributions are considered. Fosgerau (2006) looks at various distributions and concludes that a bad choice may lead to extreme bias. Hess and Axhausen (2005) have examined how well a wide range of parametric distributions can

reproduce given target distributions, which are constructed to reflect common assumptions about taste variation in transport demand models.

We note immediately that using only the goodness-of-fit to compare models does not allow one to reach valid conclusions about the validity of the random parameters distribution. Therefore, we propose a test based on seminonparametric (SNP) techniques to decide if a given distribution is appropriate or not.

The term seminonparametric distinguishes a certain class of models from parametric, nonparametric and semiparametric models. Parametric models are the standard classical models and include, e.g., the MNL or the MMNL models. One specifies a model structure and estimates a number of parameters or deep parameters such as the mean and standard deviation of a model parameter. A nonparametric model has very little structure and is based on local approximations of some kind to the relationship of interest rather than the estimation of parameters. A recent general reference to nonparametric methods is Pagan and Ullah (1999). Examples of nonparametric techniques in a transport context are nonparametric regression (Fosgerau, 2006) and local logit (Fosgerau, forthcoming). Semiparametric models are a hybrid between parametric and nonparametric models. They introduce parametric assumptions like the specification of some relationship to be a linear combination of independent variables while perhaps the errors remain nonparametric. A notable semiparametric model for discrete choice data is the Klein and Spady (1993) estimator, which has been applied in the transport context by Horowitz (1993), Fosgerau (2005) and Fosgerau (2006). Seminonparametric models are not based on local approximations but use instead series approximations to approximate functions such as densities. SNP methods were introduced by Gallant and Nychka (1987). In this paper, we employ a series approximation to approximate an unknown density and hence our approach is seminonparametric in nature.

In the next section, we describe the general methodology. In Section 3, we illustrate the power of the test on synthetic data, where the "true" distribution is specified and known in advance. As an illustration, we also apply the test on real data in Section 4. After concluding in Section 5, we provide some technical details for SNP methods based on Legendre polynomials in the Appendix.

2 Methodology

We want to test if a random parameter ω of a discrete choice model follows an a priori postulated distribution. We label this our base distribution with CDF F and density f, and not that this embodies the assumption that F is absolutely continuous, and thus has no mass points.

The true distribution may be different from F. We denote the true CDF by G and its density by g. We can rewrite the distribution G in terms of F as

$$G(\omega) = Q(F(\omega)),$$

where Q is a monotone function from $[0,1]$ to $[0,1]$. As such, Q is a CDF for a stochastic variable on the unit interval. We can differentiate this to express the density g as

$$g(\omega) = q(F(\omega))f(\omega).$$

The next step is to approximate q in a seminonparametric fashion. For this, we need an orthonormal basis for functions on the unit intervals. Among the many possibilities, we follow Bierens (forthcoming), and let L_k be transformed Legendre polynomials (see appendix). Defining

$$q_N(x) = 1 + \sum_{k=1}^{N} \delta_k L_k(x), \tag{3}$$

we approximate q by

$$q(x) \approx \frac{1}{K} q_N^2(x),$$

where

$$K = \int_{-\infty}^{+\infty} q_N^2(F(\omega))f(\omega)d\omega \tag{4}$$

is a normalizing constant such that the density g integrates to 1. Squaring q_N guarantees positivity, so that g is a density. We call the $\delta_k L_k(x)$ SNP terms and N is the number of such terms. The coefficients δ_k are unknown and must be estimated.

Bierens (forthcoming) shows that any density on the unit interval can be approximated in this way. This approximation is convenient for several reasons. Legendre polynomials have a recursive definition which is easy to implement in software. Orthonormality of the transformed polynomials is likely to reduce problems with correlation in estimation, and makes it easy to compute the normalizing constant. Indeed, defining $z = F(\omega)$ so that $\omega = F^{-1}(z)$ and $dz = f(\omega)d\omega$, we write (4) as

$$K = \int_0^1 q_N^2(z)dz = \int_0^1 (1 + \sum_{k=1}^{N} \delta_k L_k(x))^2 dx = 1 + \sum_{k=1}^{N} \delta_k^2,$$

the last equality being obtained from the orthonormality of the polynomials. A great deal of flexibility is obtained to approximate g, already with a small number of terms, and flexibility can be gradually increased by adding more terms, if necessary.

Assume now that β is a parameter of a discrete choice model. The probability

for alternative i to be chosen in choice set $\mathcal{C}$ is given by

$$P_n(i|\mathcal{C}_n) = \int_{-\infty}^{+\infty} P_n(i|\beta, \mathcal{C}_n) g(\beta) d\beta,$$

where $P_n(i|\beta, \mathcal{C}_n)$ is a closed form model, such as the Generalized Extreme Value model (McFadden, 1978). Then,

$$P_n(i|\mathcal{C}_n) \approx \frac{1}{K} \int_{-\infty}^{+\infty} P_n(i|\beta, \mathcal{C}_n) q_N^2(F(\beta)) f(\beta) d\beta$$

$$= \frac{1}{K} \int_0^1 P_n(i|F^{-1}(z), \mathcal{C}_n) q_N^2(z) dz,$$

where, again, $z = F(\beta)$. This integral is approximated by Monte-Carlo simulation, and the term $F^{-1}(z)$ corresponds to the draws of the base distribution.

Now, under the null hypothesis that the base distribution is the true distribution, we have $f = g$, which implies that q is identically 1 and thus that $\delta_k = 0$, for all k in (3). Then the model

$$P_n(i|\mathcal{C}_n) = \int_{-\infty}^{+\infty} P_n(i|\beta, \mathcal{C}_n) g(\beta) d\beta, \tag{5}$$

is equivalent to the model

$$P_n(i|\mathcal{C}_n) = \int_{-\infty}^{+\infty} P_n(i|\beta, \mathcal{C}_n) f(\beta) d\beta. \tag{6}$$

By construction, model (6) is a special case of model (5) where all coefficients (except the constant) of the polynomial approximation of q are set to 0. Consequently, a likelihood ratio test for nested hypotheses is appropriate to test the null hypothesis. If $\mathcal{L}_U$ is the log-likelihood of the sample with model (5), and $\mathcal{L}_R$ is the log-likelihood of the sample with model (6), then, under $H_0 : f = g$, the likelihood ratio statistic

$$-2(\mathcal{L}_R - \mathcal{L}_U)$$

is χ^2 distributed with N degrees of freedom, where N is the number of terms considered in the polynomial approximation.

Note that the number of SNP terms must be chosen in advance. Increasing the number of SNP terms makes the alternative hypothesis more general but also increases the demand on the data. Our experience reported later in this paper suggests that 2 or 3 SNP terms give a large degree of flexibility, which may be sufficient for most purposes, while 1 SNP term is not always sufficient to reject a false null hypothesis.

3 Simulation study

We first illustrate the concept on semi-simulated data, in order to measure the power (that is, the ability to reject false hypotheses) and the size (that is, the rate at which true hypotheses are rejected) of the test. By semi-simulated data, we mean that we have used an existing database, and performed sample enumeration with a prespecified "true" model to generate simulated choices.

The data derive from a stated choice experiment, which is part of the Danish value of time study. Some design considerations for this study are detailed in Burge et al. (2004). We have selected a route choice experiment concerning a recent trip by bus. The experimental design is particularly simple involving only in-vehicle travel time and cost. By design, the ratio of cost to time differences range between 1 and 200 DKK per hour (1 EUR $\approx$ 7.5 DKK).

Each respondent made 9 choices, one of which was always a dominated choice included as a check on respondents. Respondents who failed to choose the dominant alternative were excluded from the analysis. So were all dominant choices [1] . This leaves 1070 respondents who carried out an average of 7.7 non-dominated choices each.

The "true" model is specified as a binary model based on the following utility function:

$$U_{jnt} = \beta_{Tn}\mathrm{TT}_{jnt} + \beta_C\mathrm{TC}_{jnt} + \varepsilon_{jnt} \tag{7}$$

where U_{jnt} is the utility associated with alternative j by individual n for question number t, β_{Tn} is a random coefficient distributed across individuals, β_C is fixed and ε_{jnt} are i.i.d. extreme value distributions, so that

$$P_n(i|\mathcal{C}_n = \{i,j\}, \beta_{Tn}) = \prod_t P_{nt}(i|\mathcal{C}_n, \beta_{Tn})$$

and

$$P_{nt}(i|\mathcal{C}_n, \beta_{Tn}) = \Pr(U_{int} \geq U_{jnt}) = \frac{e^{\beta_{Tn}\mathrm{TT}_{int}+\beta_C\mathrm{TC}_{int}}}{\sum_{k=i,j} e^{\beta_{Tn}\mathrm{TT}_{knt}+\beta_C\mathrm{TC}_{knt}}}.$$

The simulated choices were generated using a cost coefficient of -0.3 and time coefficients following either a normal or a lognormal distribution. These two distributions were chosen to have the same mean and variance and most of their mass within the range of time-cost trade-offs in the data. More specifically, with cost in DKK and time in minutes, the normal distribution had mean 0.5 and standard deviation 0.2, while the lognormal distribution had mean -1 and standard deviation 0.6 in the underlying normal distribution.

[1] Due to rounding, there could be other choices that did not involve a positive price of time. They are similarly treated as dominant and removed from the sample.

A total of 100 data sets have been generated for each "true" model. We have applied the test using one SNP term. We have tested two null hypotheses: (i) the true distribution is normal and (ii) the true distribution is lognormal. The estimations have been performed with simulated maximum likelihood, using a total of 500 Halton draws. All estimations for this paper are carried out in Ox (Doornik, 2001). Consequently, the test has been implemented in Biogeme (Bierlaire, 2003; Bierlaire, 2005), and the results of the two implementations have been successfully verified against each other. Biogeme is freely available from `biogeme.epfl.ch`.

The number of rejected models is reported in Table 1. At the 95% level of confidence, the null hypothesis that the true distribution of β_T is normal is (falsely) rejected 9% of the times with the first model, and (correctly) rejected 100% of the times with the second model. Note that the exact 95% confidence interval for the true size of the test with 100 draws, that is [4.9%–16.4%], contains the 5% nominal size. The null hypothesis that the true distribution of β_T is lognormal is (correctly) rejected 99% of the times with the first model, and (falsely) rejected 5% of the times with the second model.

			H_0	
			Normal	Lognormal
95%	True dist:	Normal	9	99
	True dist:	Lognormal	100	5
99%	True dist:	Normal	1	78
	True dist:	Lognormal	88	0

Table 1
Simulated data: number of rejections with 1 SNP term

We analyze these results in more detail in Figures 1 and 2, where the cumulative distribution of the likelihood ratio statistic is reported for the 4×100 experiments. Figure 1 reports the results for testing the null hypothesis that the true distribution is a normal (corresponding to the first column in Table 1), and Figure 2 reports the results for testing the null hypothesis that the true distribution is a lognormal. The threshold for the 95% test is shown (3.84, from the χ^2 distribution with one degree of freedom), as well as the 99% (6.63). At the 99% level of confidence, the number of false rejections drops, as well as the number of correct rejections, as reported in Table 1.

Although both the power and the size of the test are very good when just one SNP term is used, we have also applied the test with two SNP terms. The results are reported in Table 2.

The test seems to perform very well in these circumstances. The power of the

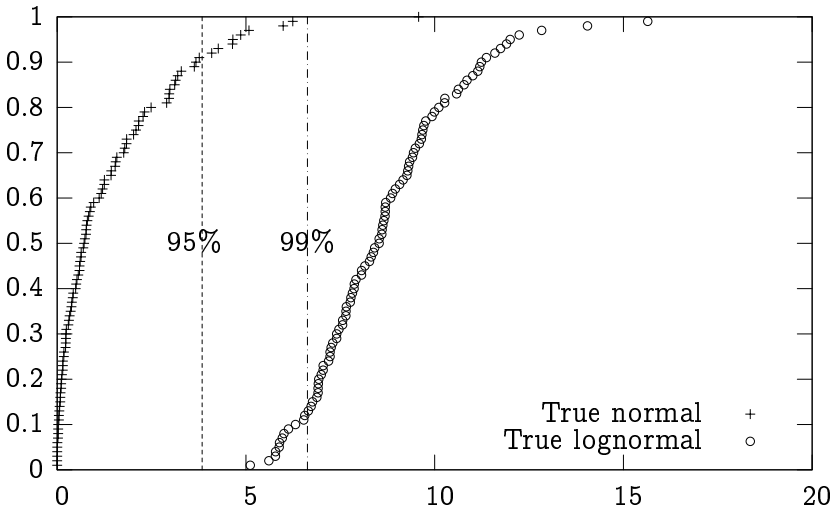

Fig. 1. Distribution of the likelihood ratio for 100 experiments under H_0="true normal"

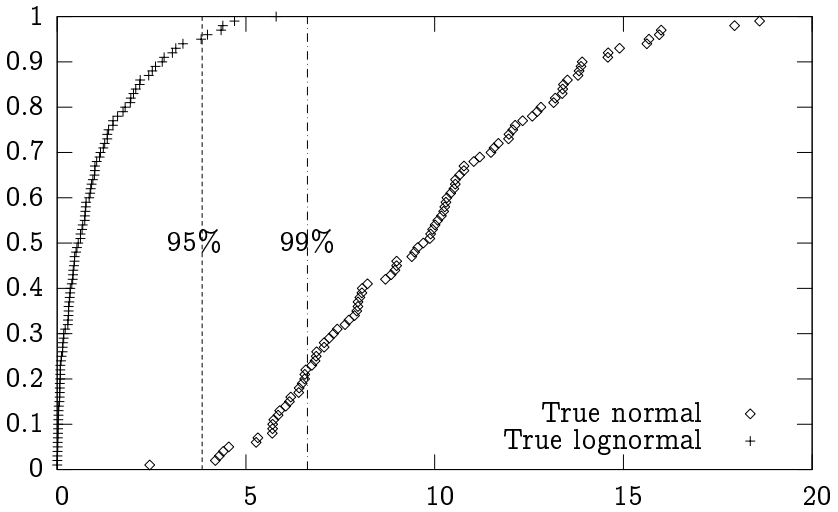

Fig. 2. Distribution of the likelihood ratio for 100 experiments under H_0="true lognormal"

test is very high, allowing us to reject a very large proportion of false nulls, even at the 99% level of confidence.

4 Case study

We now apply the test to the real data set, using again the model specification (7). We test the model with one SNP term, where the base distribution of β_T is a normal. The results of the two estimations are reported in Table 3. The

			H_0	
			Normal	Lognormal
95%	True dist:	Normal	9	100
	True dist:	Lognormal	100	3
99%	True dist:	Normal	4	99
	True dist:	Lognormal	100	0

Table 2
Simulated data: number of rejections with 2 SNP terms

likelihood ratio (LR) test is

$$-2(-4153.57 + 4150.14) = 6.86$$

and the H_0 hypothesis that β_T follows a normal distribution can be rejected at the 99% level of confidence. Note that the coefficient δ_1 of the SNP term is significantly different from 0. A visual comparison of the estimated densities of the normal in the first model and of g in the second shows however that they look quite similar (see Figure 3). The value of travel time (VTT), computed for both distributions with truncation at zero, are quite similar. This shows that the test is strong and able to detect small differences.

$\mathcal{L} =$	$\beta_T \sim f(x) = N(\mu, \sigma^2)$ -4153.57			$\beta_T \sim g(x)$ -4150.14		
	Estim.	Std.err.	t-test	Estim.	Std.err.	t-value
β_C	-0.36	0.01	-25.1	-0.35	0.01	-25.3
$\mu(\beta_T)$	0.03	0.01	1.9	-0.15	0.06	-2.7
$\sigma(\beta_T)$	0.34	0.01	24.4	0.38	0.02	16.2
δ_1				0.25	0.08	3.3
VTT (DKK/h)	25.33			25.49		

Table 3
Testing a normal distribution

We now test the model with one SNP term, where the base distribution of β_T is lognormal. The results of the two estimations are reported in Table 4.

The LR test is

$$-2(-4304.32 + 4302.94) = 2.76$$

and the H_0 hypothesis that β_T follows a lognormal distribution cannot be rejected at the 95% level of confidence and not even at the 85% level. A visual comparison of the densities of the lognormal in the first model and of g in the second show no visible differences (see Figures 4 and 5), and the VTT is about the same with the two models.

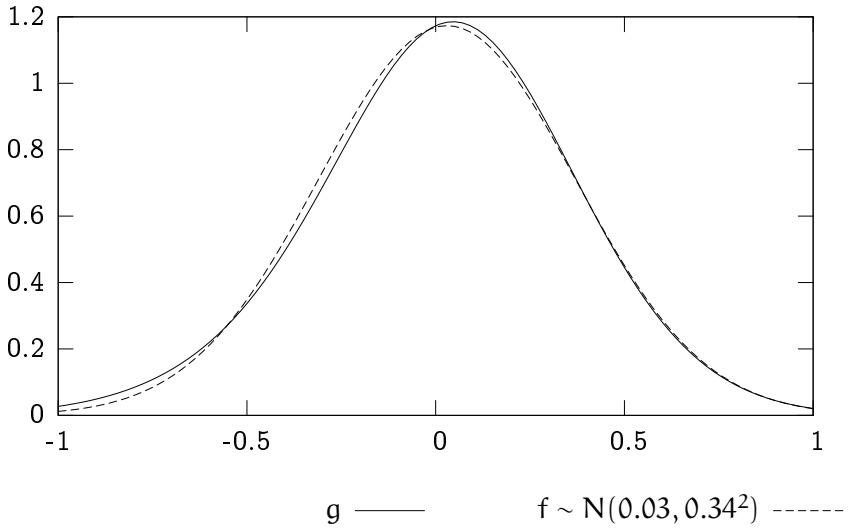

Fig. 3. Comparison of f and g, f normal

	$\ln(\beta_T) \sim f(x) = N(\mu, \sigma^2)$			$\beta_T \sim g(x)$		
$\mathcal{L} =$	-4304.32			-4302.94		
	Estim.	Std.err.	t-test	Estim.	Std.err.	t-value
β_C	-0.45	0.01	-38.3	-0.45	0.02	-28.7
$\mu(\beta_T)$	-2.52	0.05	-46.8	-2.92	0.06	-46.5
$\sigma(\beta_T)$	1.43	0.05	30.6	1.50	0.04	36.8
δ_1				0.14	0.04	3.3
VTT (DKK/h)	30.48			32.13		

Table 4
Testing a lognormal distribution

We have also performed the test with 2 and 3 SNP terms. The results are presented in Table 5, where we denote by g_N the distribution obtained with N SNP terms. We obtain a large improvement in the likelihood when a second SNP term is added. Also, the average VTT significantly changes between the model with one and the model with two terms. The likelihood ratio test for the model with two terms in the polynomial is

$$-2(-4304.32 + 4263.57) = 81.50$$

which is far beyond the 5.99 threshold of the 95% level, and even far beyond the 9.21 threshold of the 99% level. Therefore, we clearly reject the lognormal in this case. A visual comparison (Figures 6 and 7) illustrates well the cause of this rejection.

This completes our illustration of the applicability of the test. In practice, the model specification must be developed and improved. Fosgerau (2006)

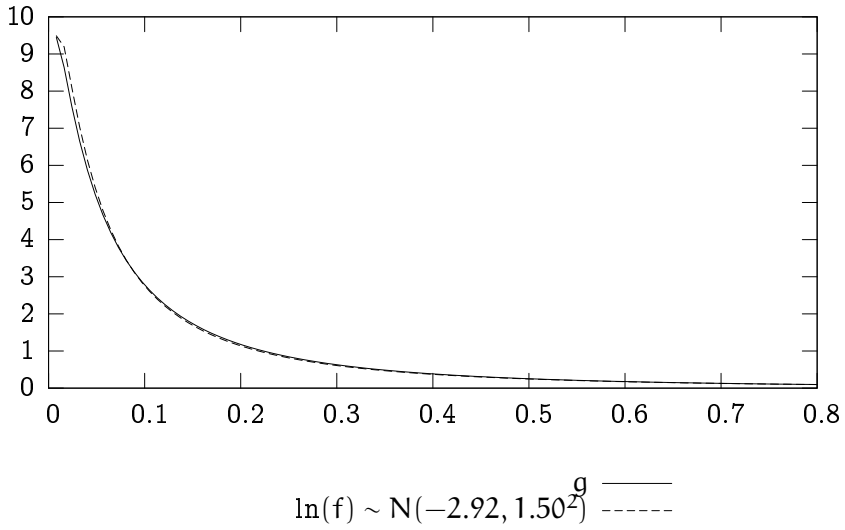

Fig. 4. Comparison of f and g, f lognormal

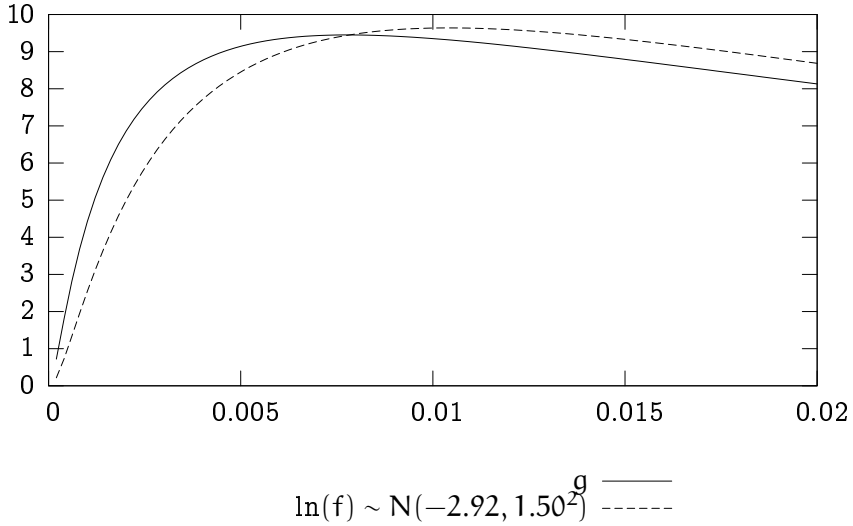

Fig. 5. Comparison of f and g, f lognormal (zoom)

addresses this issue using the same data.

We conclude this section with some important notes.

- The test does not always reject the base distribution. We have used it for complex models in the context of the Danish value-of-time study, and have accepted a lognormal VTT in several cases, using 3 SNP terms.
- In the presence of SNP terms, like for many complex models, the estimation algorithms may be trapped in local maxima. We have experienced this in a few instances. Although it does not provide any guarantee to find the global maximum, it is good practice to use several different starting points, or to use heuristics which are designed to escape from local maxima.
- If the estimates of the base model correspond to the global maximum of the

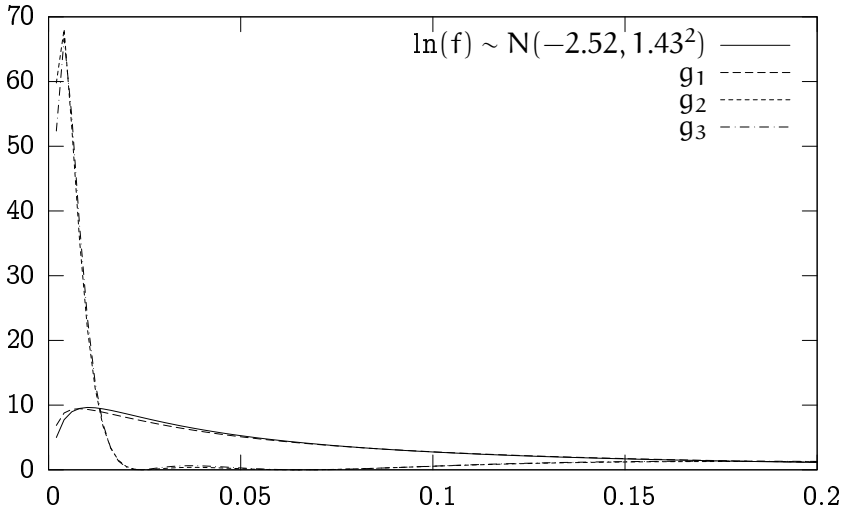

Fig. 6. Comparison of f, g_1, g_2 and g_3, f lognormal

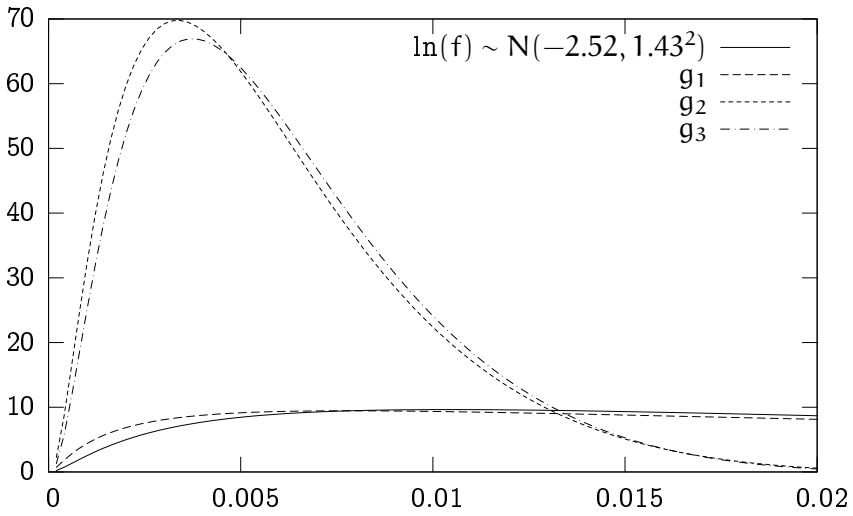

Fig. 7. Comparison of f, g_1, g_2 and g_3, f lognormal (zoom)

log-likelihood function, and those of the SNP model to a local maximum, the test may reject the base distribution less often than it should, but a rejection is still valid.

5 Conclusion

We have proposed a method based on a seminonparametric (SNP) specification to test if a random parameter of a discrete choice model indeed follows a given distribution. The simulation study shows that the test is well able to discriminate between normal and lognormal with 1 SNP term only.

We have demonstrated the application of the test on a case study. We reject the normal with 1 SNP term. The lognormal is not rejected with 1 term but it is clearly rejected with two terms. So we conclude that it is generally desirable to include two or three SNP terms in order to test the base distribution against an alternative that is as general as possible. There are relevant alternatives which are not captured with just 1 SNP term. However, there seems to be little point in using more than 2 or 3 SNP terms.

In summary, the test works well on simulated data, and we have shown a successful application of the test to real data. It can reject bad models and, by including more terms for flexibility, show how a particular distribution may fail. So we believe that it is a very powerful tool for practical applications.

In the future, it would be interesting to adapt the test for more than one random parameter. A simple heuristic could consist in testing each parameter independently. The robustness of this approach must be analyzed both in the case of independent and correlated random parameters. Also, the practical use of multivariate polynomial approximations should be assessed.

Other uses of the model are possible. Namely, one can treat the SNP extension to some base distribution as a way of generating a larger model universe with greater flexibility allowed for the mixing distribution. Note however that the trade off between model flexibility and data overfitting must always be considered by the analyst.

As the sample size and the number of SNP terms tends to infinity, one could hope that the model would be capable of approximating any true mixing distribution. Identification and consistency of this procedure is the subject of ongoing research.

A Appendix: Legendre polynomials

The Legendre polynomials $\widehat{L}_n(x)$ are defined by

$$\widehat{L}_n(x) = \sum_{m=0}^{M} (-1)^m \frac{(2n-2m)!}{2^n m!(n-m)!(n-2m)!} x^{n-2m}$$

where $M = n/2$ or $M = (n-1)/2$, whichever is an integer. They can also be defined recursively,

$$\widehat{L}_n(x) = ((2n-1)x\widehat{L}_{n-1}(x) - (n-1)\widehat{L}_{n-2}(x))/n,$$

where $\widehat{L}_0(x) = 1$ and $\widehat{L}_1(x) = x$ (see (Abramowitz and Stegun, 1972, chap. 8 and 22)). They are orthogonal on $[-1, 1]$ in the sense that

$$\int_{-1}^{1} \widehat{L}_m(x)\widehat{L}_n(x)dx = 0 \quad \text{if } m \neq n.$$

In our context, it is more appropriate for the polynomials to be orthogonal on $[0, 1]$, as the arguments are defined by a CDF. Therefore, Bierens (forthcoming) proposes the following transformation:

$$L_n(x) = \sqrt{2n+1}\widehat{L}_n(2x-1)$$

so that they are orthonormal on $[0, 1]$, that is

$$\int_0^1 L_m(x)L_n(x)dx = \begin{cases} 0 \text{ if } m \neq n \\ 1 \text{ if } m = n. \end{cases}$$

The recursive definition of these transformed polynomials is given by

$$L_n(x) = \frac{\sqrt{4n^2-1}}{n}(2x-1)L_{n-1}(x) - \frac{(n-1)\sqrt{2n+1}}{n\sqrt{2n-3}}L_{n-2}(x).$$

The first polynomials are

$$\begin{aligned} L_0(x) &= 1 \\ L_1(x) &= \sqrt{3}(2x-1) \\ L_2(x) &= \sqrt{5}(6x^2-6x+1) \\ L_3(x) &= \sqrt{7}(20x^3-30x^2+12x-1). \end{aligned}$$

References

Abramowitz, M., Stegun, I. A. (Eds.), 1972. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th Edition. Dover, New York.

Algers, S., Bergstrom, P., Dahlberg, M., Dillen, J., 1998. Mixed logit estimation of the value of travel time. Tech. Rep. 1998:15, Uppsala - working paper series, available at http://ideas.repec.org/p/fth/uppaal/199815.html.

Bekhor, S., Ben-Akiva, M., Ramming, M. S., 2002. Adaptation of logit kernel to route choice situations. Transportation Research Record 1805, 78–85.

Ben-Akiva, M., Bolduc, D., Walker, J., 2001. Specification, identification, and estimation of the logit kernel (or continuous mixed logit) model. Tech. rep., Department of Civil Engineering, MIT, Cambridge, Ma.

Bhat, C. R., Castelar, S., 2002. A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the San Francisco Bay area. Transportation Research Part B: Methodological 36B, 593–616.

Bierens, H. J., forthcoming. Semi-nonparametric modeling of densities on the unit interval, with applications to censored mixed proportional hazard models and ordered probability models: Identification and consistency results. Econometric Theory.

Bierlaire, M., 2003. BIOGEME: a free package for the estimation of discrete choice models. In: Proceedings of the 3rd Swiss Transportation Research Conference. Ascona, Switzerland, www.strc.ch.

Bierlaire, M., 2005. An introduction to BIOGEME version 1.4, biogeme.epfl.ch.

Brownstone, D., Bunch, D., Train, K., 2000. Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. Transportation Research Part B: Methodological 34, 315–338.

Brownstone, D., Small, K., 2005. Valuing time and reliability: assessing the evidence from road pricing demonstrations. Transportation Research Part A 39 (4), 279–293.

Burge, P., Rohr, C., Vuk, G., Bates, J., 2004. Review of international experience in VOT study design. In: Proceedings of the European Transport Conference.

Carrier, E., 2003. Modeling airline passenger choice: Passenger preference for schedule in the passenger origin-destination simulator (PODS). Ph.D. thesis, Massachusetts Institute of Technology.

Doornik, J. A., 2001. Ox: An Object-Oriented Matrix Language. Timberlake Consultants Press, London.

Electric Power Research Institute, 1977. Methodology for predicting the demand for new electricity-using goods. Final report. Project 488-1 EA-593, Electric Power Research Institute, Palo Alto, Ca.

Fosgerau, M., 2005. Unit income elasticity of the value of travel time savings. In: European Transport Conference.

Fosgerau, M., 2006. Investigating the distribution of the value of travel time savings. Transportation Research Part B: Methodological 40 (8), 688–707.

Fosgerau, M., forthcoming. Specification of a model to measure the value of travel time savings. Transportation Research Part A.

Frejinger, E., Bierlaire, M., 2007. Capturing correlation with subnetworks in route choice models. Transportation Research Part B: Methodological 41 (3), 363–378.

Gallant, A. R., Nychka, D. W., 1987. Semi-nonparametric maximum likelihood estimation. Econometrica 55 (2), 363–390.

Greene, W. H., Hensher, D. A., Rose, J., 2006. Accounting for heterogeneity in the variance of unobserved effects in mixed logit models. Transportation Research Part B: Methodological 40 (1), 75–92.

Han, B., Algers, S., Engelson, L., 2001. Accommodating drivers' taste variation

and repeated choice correlation in route choice modeling by using the mixed logit model. In: 80th Annual Meeting of the Transportation Research Board.
Hensher, D., Greene, W., 2003. The mixed logit model: the state of practice. Transportation 30 (2), 133–176.
Hess, S., Axhausen, K. W., 2005. Distributional assumptions in the representation of random taste heterogeneity. In: Proceedings of the 5th Swiss Transportation Research Conference. Www.strc.ch.
Hess, S., Bierlaire, M., Polak, J., 2005. Estimation of value of travel-time savings using mixed logit models. Transportation Research Part A 39 (3), 221–236.
Hess, S., Polak, J., 2005. Mixed logit modelling of airport choice in multi-airport regions. Journal of Air Transport Management 11 (2), 59–68.
Hess, S., Train, K., Polak, J., 2006. On the use of modified latin hypercube sampling (MLHS) method in the estimation of mixed logit model for vehicle choice. Transportation Research Part B: Methodological 40 (2), 147–163.
Horowitz, J., 1993. Semiparametric estimation of a work trip mode choice model. Journal of Econometrics 58, 49–70.
Klein, R., Spady, R., 1993. An efficient semiparametric estimator for binary response models. Econometrica 61 (2), 387–422.
McFadden, D., 1978. Modelling the choice of residential location. In: A. Karlquist *et al.* (Ed.), Spatial interaction theory and residential location. North-Holland, Amsterdam, pp. 75–96.
McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. Journal of Applied Econometrics 15 (5), 447–470.
Pagan, A., Ullah, A., 1999. Nonparametric Econometrics. Cambridge University Press.
Revelt, D., Train, K., 1998. Mixed logit with repeated choices: Households' choices of applicance efficiency level. Review of Economics and Statistics LXXX (4), 647–657.
Small, K. A., Winston, C., Yan, J., 2005. Uncovering the distribution of motorists' preferences for travel time and reliability. Econometrica 73 (4), 1367–1382.
Train, K., 1998. Recreation demand models with taste differences over people. Land Economics 74 (2), 230–239.
Train, K., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, http://emlab.berkeley.edu/books/choice.html.
Viton, P. A., 2004. Will mixed logit change urban transport policies? Journal of Transport Economics and Policy 38 (3), 403–423.

	$\ln(\beta_T) \sim f(x) = N(\mu, \sigma^2)$			$\beta_T \sim g_1(x)$			$\beta_T \sim g_2(x)$			$\beta_T \sim g_3(x)$		
$\mathcal{L}$=	-4304.32			-4302.94			-4263.57			-4263.29		
	Estim.	Std.err.	t-test	Estim.	Std.err.	t-test	Estim.	Std.err.	t-test	Estim.	Std.err.	t-test
β_C	-0.45	0.01	-38.3	-0.45	0.02	-28.7	-0.45	0.01	-38.5	-0.45	0.01	-34.94
$\mu(\beta_T)$	-2.52	0.05	-46.8	-2.92	0.06	-46.5	-3.25	0.04	-86.9	-3.16	0.03	-92.28
$\sigma(\beta_T)$	1.43	0.05	30.6	1.50	0.04	36.8	1.29	0.02	51.7	1.26	0.02	63.80
δ_1				0.14	0.04	3.3	-0.07	0.05	-1.52	-0.02	0.08	-0.19
δ_2							1.20	0.25	4.76	1.29	0.32	4.07
δ_3										-0.12	0.13	-0.92
VTT (DKK/h)	30.48			32.13			52.85			46.56		
LR against 0				2.76			81.50			82.06		
LR against 1							78.74			79.30		
LR against 2										0.56		

Table 5. Testing a lognormal distribution with more than one term