IMPACTOF *A PRIORI* DISTRIBUTIONS ON MIXED LOGIT MODEL ESTIMATION TESTS ON SYNTHETIC DATA

Majken Vildrik Sørensen, Otto Anker Nielsen
Centre for Traffic and Transport, Technical University of Denmark,
Building 115st, DK2800 Lyngby, Denmark,
mvi@ctt.dtu.dk; oan@ctt.dtu.dk

Abstract

This paper focuses on the MNL and the Mixed Logit models' ability to correctly estimate a model. Synthetic data has been formed followed by estimation of MNL and Mixed Logit models conditional of a variety of shapes of distribution for distributed terms. Model selection based solely on a combination of MSL estimation, t-test and likelihood ratio test is shown to be ambiguous. A test for determining the shape of distribution is proposed. Application of the method retrieved the distribution from data.

Keywords: Mixed Logit model; model specification; shape of distribution; Maximum Simulated Likelihood; behaviour.
Topic area: D1 Passenger Transport Demand Modeling

1 Introduction

Models including error components are today at the frontier of application and development in transport modelling. The name Error Component Models are often used indiscriminately with Random Coefficients (Parameters) Logit, Models with Stochastic (Distributed) Coefficients (Preferences), Mixed Logit Models, Logit Kernel Models or lately Hybrid Choice Models, for models where additional stochastic terms are included in the traditional (linear) utility function. The use of Mixed Logit models (or Hybrid Logit) has grown rapidly during the past five years as research has demonstrated the applicability of the model and especially the software/purpose specific code has become available. Three examples of this are Alogit4ec (RAND Europe, commercial), GAUSS code at http://elsa.berkeley.edu/train/software.html (freeware) or the self-contained BioGeme at http://roso.epfl.ch/biogeme (freeware).

The method of simulated likelihood (MSL) is generally applied for estimation of mixed logit models, though alternatives exists, e.g. the Method of Simulated Moments, see e.g. Train (2003). Estimation by means of MSL is optimisation of the model's utility coefficients and distribution related parameters conditional on the a priori assumed shape of distribution for the distributed terms. Only few of the analyses so far, has dealt with the interesting question of correlation between these error components.

The question of which priori shape of distribution to apply for the coefficients, has not been intensively researched. This paper concerns the issue of how to determine which shape of distribution to employ in a model estimation and which of the model coefficients the distributed terms should be related to.

By use of synthetic data it is demonstrated that even distributed terms with wrong shape of distribution assumed for distributed terms may improve model fit. Further, that assuming fewer distributed terms than is the case does provide an improvement in model fit. Hence, few model

runs for different assumed shape of distribution does not provide a sufficiently base for determining the distribution of distributed terms. The paper extents the paper of Sørensen and Nielsen (2002); Sørensen (2002) with a more comprehensive test of how to incorporate distributed terms into traffic models (construction of the utility function) and the use of synthetic data sets in Sørensen (2003a).

2 The mixed model

Estimation of which alternative is preferred over others involves exhaustive evaluation of in principle all (relevant) alternatives' utility. Generally, alternatives utility cannot be completely described, setting out the need for a (distributed) residual. This residual is not in it self interesting, merely the distribution of the maximum of the residuals over alternatives. Specifically, for individual j

$$U_{ji} = V_{ji} + \epsilon_j, \qquad (1)$$

is used to describe the indirect utility for alternative i. Peitz (1995) derived the indirect utility function from a direct utility function consistent with RUM.

The most common of models, the multinomial logit model (MNL), is obtained by assuming the residuals are IID Gumbel (Extreme Value, type I). The attractiveness of the model is easily seen from the formulation of the choice probabilities, $P(i) = \exp(V_i)/\sum_j \exp(V_j)$ where i, j are alternatives. A consequence of the model is the independence between alternatives property (IIA).

Traditionally, for model estimations it has been assumed that individual behaviour is identical across a population. One simple step away from this way of thinking, is segmentation of a population by e.g. trip purpose, into smaller samples each consistent in behaviour and generally different between samples. The idea of segmentation can be extended to one sample per individual, where the difference between the samples is described by a (continuous/discrete) distribution. Such models are labelled 'Mixed Logit', 'Random Coefficients/ Parameters Logit' (RCL/ RPL), 'Error Components Logit', Mixed MultiNomial Logit (MMNL), 'Models with a logit (Probit) Kernel or as a member of the rigorous 'Hybrid Logit' class[1].

Briefly formulated the construction of the utility function is altered from Uji = Vji +νj, where e.g. Vji = ∑ βX [2] respective νj represents the explained respective the unexplained variation in the utility function for choice of alternative i by individual j, to

$$U_{ji} = V_{ji} + \eta_j + \epsilon_j. \qquad (2)$$

This corresponds to partitioning of the unexplained variation (the ν)into a systematic (describable) ηj, and an unsystematic part ϵ_j.

In Random Coefficients Logit (RCL) ηj is formulated as a mean zero random variate ξ multiplied by an attribute, potentially one for each attribute, while the unsystematic part (ϵ) is (again) assumed Gumbel distributed though not identical to ν. Hence, the formulation is as follows (the second equality requires the deterministic part of utility to be linear in at least the 'distributed' coefficients)

---

[1] Consistencies and differences between the model variants including a map of model relations are described in Sørensen (2003a, 2004). To the authors knowledge the first application with transportation modelling (market shares) references (Boyd and Mellman, 1980; Cardell and Dunbar, 1980) or the technical documentation of the work at EPRI, from 1977.

[2] Lineasr in coefficients utility functions is not necessary for the extension to mixed models.

$$U_{ji} = V_{ji} + \xi X + \epsilon_j$$
$$= \sum_k (\beta + \xi_j) X_j + \epsilon_j, \qquad (3)$$

where the distribution of ξ is interpreted as difference in preferences between individuals. Each of the ξj are either assumed to follow a stochastic distribution with mean zero and variance σ_j^2 or are identical 0 (fixed coefficients). Traditionally distributions suggested are the normal (McFadden and Train, 2000) and the lognormal (BenAkiva et al., 1993)[3] (Train, 1999), though other distributions have been suggested which include the χ^2 (Nunes et al., 2001), the uniform (Revelt and Train, 2000), the triangular (Revelt and Train, 2000; Train, 2001) and the Rayleigh (Siikamaki and Layton, 2001). Covariance between the stochastic elements ξ is allowed; though most applications have been limited to the normal distribution. The unsystematic parts are IID.

Typically, has been specified as a vector, that is, the random effects are assumed (statistically) independent, which again implies that a person with a higher than average value of travel time is not any more likely than any other person to have a higher than average value of e.g. waiting time. Few examples of applications including as a matrix (at least one nonzero offdiagonal element) are (Nielsen et al., 2001; Sørensen and Nielsen, 2001; Sørensen, 2002; Hensher and Reyes, 2000).

3 MSL and a priori shape of distribution

The common way to estimate a mixed model is essentially, to simulate weights for 'ordinary' choice probabilities and then maximise the weighted likelihood function. The method is labelled Maximum Simulated Likelihood (MSL) and involves one integral $(-\infty, \infty)$ for each added dimension of stochastic term(s), plus the one inherited from the residual term.

The likelihood is an extended version of the ordinary likelihood for a 'fixed coefficients model', given by

$$\mathcal{L} = \mathcal{L}(\beta|X) = \sum_n \sum_i g_{in} \ln(P_i). \qquad (4)$$

where β are coefficients and gin is 1 if alternative i is chosen, 0 otherwise; Pi is the probability for choice of alternative under some probability model e.g. MNL.

Adding further distributed terms to the utility, corresponds to letting be distributed $f(\beta|\theta)$ where θ is the parametrisation of the distribution. Let β^r refer to the rth of R draws from $f(\beta|\theta)$ and Pi(β^r)the probability for choice of alternative given the drawn value β^r. Averaging over successive draws from $f(\beta|\theta)$, gives

$$\breve{P}_i = \frac{1}{R} \sum_n \sum_j g_{jn} \ln P_j(\beta^r). \qquad (5)$$

The simulated loglikelihood is then given by

$$SL = SL(\theta|X) = \sum_n \sum_j g_{jn} \ln \breve{P}_j, \qquad (6)$$

[3] In this application the distributed term was the ratio of two coefficients (the VOT) rather than the coefficient itself. The idea was to reformulate the utility function such that U = $\alpha c + \beta t + \epsilon_j = \alpha(c + \eta t) + \epsilon_j$, where distribution was allowed for η, hereby circumventing finding the distribution of the ratio of two stochastic distributed variables.

and the estimator for the coefficients is the θ*, for which SL obtains the maximum value[4]. As the maximum likelihood estimation is conditional on the formulation of the probability function Pi, the simulated maximum likelihood estimation is conditional on the formulation of the probability function Pi and the functional form assumed for the distribution.

Given a specification of the utility that includes distribution of stochastic terms, the likelihood function can be formulated and solved by Maximum Simulated Likelihood (MSL).

One criticism of the MSL is that it only optimises parameter values of the a priori specified distribution. In practice it is possible to estimate mixed models based on erroneous shapes of distribution, yet obtain significant parameter estimates!

In figure 1 below, it is illustrated why 'wrong' shapes of distributions perform better in terms of likelihoodvalues than fixed coefficient models. The histogram illustrates the case where individuals 'true' β coefficient is distributed. Ordinary 'fixed coefficient models' are attempting to model this by a degenerate distribution, the dark red line. Further four examples of shapes; to symmetrical (normal and triangular) and two positively skewed (sign specific or not) are included. Common for the distributions are their ability to very well describe the variation around the mode; however, the normal and nonsign specific distributions poorly describes the variation below A, whereas the triangular distribution poorly describes the variation above B.

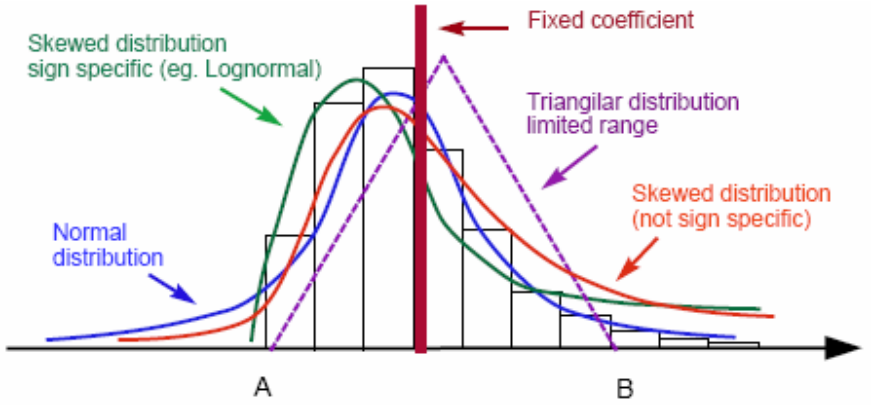

Figure 1: Examples of different shapes of distributions a priori assumed to describe variance in individual preferences.

A natural question follows; should we choose between various shapes or merely rely on choosing between fixed or distributed coefficients? As the following section shows, we cannot us MSL estimation to provide information on whether the applied shape is the best suited for the task.

4 Synthetic data

For illustration of the estimation issue, several different synthetic data sets have been generated and tested. This paper will focus on only one data set described by two attributes each multiplied by a distributed coefficient. The data set consisted of 9,762 records, where all were formed as choice between three alternatives available for all 'individuals'. The alternative with the highest random utility U was chosen; where utility was modelled by two descriptive variables multiplied

[4] Hajivassiliou and Ruud (1994) derive properties of the asymptotic distribution, e.g. that the MSL estimator is consistent if the number of draws for each observation rises with the number of observations.

by a coefficient in linear specification, two alternative specific constants and a Gumbel distributed residual added. The utility of alternative j was described as

$$U_j = ASC_j + (\beta_1 + \xi_1^{shape})X_{j1} + (\beta_2 + \xi_2^{shape})X_{j2} + \epsilon_j, \qquad (7)$$

ξ_k^{shape} is zero for 'fixed coefficient' or follows any other shape of distribution, ϵ is the Gumbel distributed residual and ASC3 is zero for identification. The data Xjk was generated by independent normal distributions with different seeds, means and variances such that $X_1 \sim N(10, 3^2)$ and $X_2 \sim N(7, 2^2)$.

Table 1: Actual coefficients moments of coefficients and ASC's

	Mean	Median	Variance	Skewness	Kurtosis
ASC_1	1.50	1.50	-	-	-
ASC_2	0.50	0.50	-	-	-
First coefficient	-1.02	-1.01	0.22	-0.20	-0.28
Second coefficient	-3.29	-3.00	1.07	-2.93	19.74

The first coefficient was assumed normal distributed, whereas the second was assumed (negatively skewed) lognormal. Mean values and variances were different for the two distributed terms, further the distributed coefficients were assumed independent. The mean value for the first coefficient is the prespecified value for β1 =-1, the variance for the first coefficient the prespecified value for $\sigma 1 = 0.5^2$.

The mean value for the second coefficient β2 can be calculated from E(β2 +ξ)= E(β2)+ E(ξ), where $E(\xi) = \exp(\zeta)\sqrt{\exp(\sigma^2)}$, $\log \xi \sim N(\zeta, \sigma^2) = N(0, 0.7)$. Hereby, the mean value of the second coefficient is -2-1.42 = -3.42.

As the data was simulated and subsequently some records have been removed due to sign change of either the β coefficients or the explanatory variable(s), estimated moments are listed below in table 1. As can be seen the moments (mean and variance) are close to the specified values. Further, it was validated that the effect of the distributed terms was large enough to detect a change of chosen alternative[5].

The 'true' coefficient values for the NLN data is given in table 1. The actual mean (the mean of the generated values) is 3.29, that is, these are rather close and the difference is in part due to the random generation and in part to that records with wrong sign of attributes or coefficients are removed. Also, the distribution of the lognormal coefficient is 'wider' than the normal coefficient, as the variation is 1.07 compared to the 0.22 for the normal.

5 Estimation methods

The paper takes advantage of different estimations techniques. First, the MNL model is estimated by ordinary maximum likelihood using both Alogit and Biogeme software. Secondly, mixed models were estimated based on different a priori assumptions of shapes of distributions. These models were estimated twice, by means of Alogit (pseudo random draws) and biogeme (Halton sequences).

[5] The data construction and model testing is further described in Sørensen (2003a).

The current version of Biogeme (version 07, beta version) enables the modeller to specify generalised utilities, where nonlinear expressions and distributed terms are possible. The distributed term is specified as a normal distributed term. However, transformations of this distributed terms is enabled whereby e.g. the Lognormal, χ^2 are possible distributions by use of simple transforms; other distributions may be obtained by use of inverse transformation. Hereby, a random variable X drawn from a uniform distribution U can be 'transformed' to follow a distribution with (invertible) cumulative density function B by $B^{-1}(X)$. The inverse transform is further described in e.g. Ross (1997). If other than the normal distributions are required some recalculation of the parameter estimates is required before genuine mean and variance estimates for the distribution are obtained.

Common for the two programs is the need for a random number generator, as an underlying concept of the simulations performed. Random numbers may be generated by various means, where probably the most common is pseudorandom numbers. An alternative way to produce the random numbers is Halton numbers where (negative) correlation is introduced which significantly reduces the number of draws required, see e.g. Train (2003); Bhat (1997).

6 Estimation results

Results of the MNL model estimation are shown in table 2. Two measures of model fit are listed, the ρ^2 and the likelihood. Parameter recovery, as an indicator of model fit is only of value, if the models actual scale is known since only the logit scale times β and not itself is estimated.

The MNL model did not correctly estimate the coefficients, as expected, as distributed coefficients were used to generate the data. However, significance levels of the coefficient estimates (ttest 28 to 60) combined with ρ^2 =0.53 – indicate that the model does explain a part of the variation. The loglikelihood value is 4,935. The deviance from the 'true' values cannot be explained by scaling as neither of the ratios are recovered.

Table 2: MNL Results for the 2 attribute data sets

	Est.	σ	t
α_1	1.182	0.0418	28.3
α_2	1.223	0.0419	29.2
β_1	-0.394	0.00829	-47.5
β_2	-1.060	0.0177	-59.8
$\mathcal{L}$	-4935		
ρ^2	0.53		
Obs	9,762		

6.1 MSL results

For the synthetic data the correct specified model was estimated by means of the MSL method. In a practical application of MSL the 'correct' shape of distribution is not known. One way to proceed has been to apply a symmetrical (normal) and a skewed distribution and assert which lead to the best model fit in terms of likelihood value and sound knowledge of the planning situation at hand. To replicate this modeler behaviour a large number of maliciously specified models were also estimated. These included assuming

- wrong shape of distribution

- assume distribution on nondistributed coefficients
- not assume distribution on distributed coefficients
- combinations of the above.

The list of possible distributions to apply is endless, here the author limited the list to combinations of fixed coefficients (F), normal distribution (N), lognormal distribution (LN)and the χ^2 distribution (χ^2). For latter two, an explicit sign was added, whereby the number of possible combinations of distribution and sign added up to a total of 36 combinationsoccurs.For all the shown models 500 Halton draws[6] have been used in the estimation, where each model had a run time of 59 hours on a 2GHz PC. Below references to the distributions are structured as χ^2 if the first coefficient is assumed to follow a normal distribution and the second a negatively skewed χ^2 distribution, etc.

Table 3: MSL Results for a sequence of estimations based on different shapes of distributions.

		F		N		LN		$-LN$		χ^2		$-\chi^2$	
F	ASC_1	1,182	28,3	1,337	26,9	1,295	27,4	1,359	26,3	1,182	28,3	1,182	28,3
	ASC_2	1,222	29,2	1,412	27,9	1,364	28,4	1,437	27,3	1,223	29,2	1,223	29,2
	β_1	-0,394	-47,5	-0,460	-40,0	-0,445	-42,1	-0,467	-38,4	-0,394	-47,5	-0,394	-47,5
	β_2	-1,060	-59,8	-1,613	-31,4	-2,521	-60,3	-0,055	-2,7	-1,060	-59,8	-1,060	-59,8
	σ_1												
	σ_2			1,195	16,4	0,736	24,6	1,105	16,2	0,000	0,0	0,000	NaN
	$\mathcal{L}/\rho^2$	-4934,6	0,54	-4855,3	0,55	-4867,6	0,55	-4851,5	0,55	-4934,6	0,54	-4934,6	0,54
N	ASC_1	1,231	27,3	1,402	25,2	1,376	25,6	1,417	25,0	1,231	27,3	1,231	27,3
	ASC_2	1,282	28,0	1,491	25,8	1,459	26,3	1,500	26,0	1,282	28,0	1,282	28,0
	β_1	-0,374	-42,3	-0,572	-23,0	-0,544	-24,4	-0,544	-22,8	-0,374	-42,3	-0,374	-42,3
	β_2	-1,104	-51,8	-1,325	-33,3	-2,560	-50,8	-0,411	-12,7	-1,104	-51,8	-1,104	-51,8
	σ_1	-0,261	-7,5	0,378	7,4	0,337	6,9	0,288	4,9	-0,261	-7,5	-0,261	-7,5
	σ_2			0,598	14,1	-0,409	-22,1	0,911	11,2	0,000	0,0	0,000	NaN
	$\mathcal{L}/\rho^2$	-4924,8	0,54	-4851,3	0,55	-4856,8	0,55	-4845,1	0,55	-4924,8	0,54	-4924,8	0,54
LN	ASC_1	1,234	27,2	1,398	25,4	1,350	25,9	1,415	25,6	1,229	27,4	1,229	27,4
	ASC_2	1,285	27,9	1,487	26,0	1,433	26,5	1,496	26,7	1,278	28,2	1,278	28,2
	β_1	-1,480	-80,8	-1,580	-61,2	-1,556	-66,9	-1,460	-109,1	-1,383	-150,3	-1,383	-150,3
	β_2	-1,106	-51,3	-1,323	-33,5	-2,308	-62,4	-0,417	-13,3	-1,101	-52,5	-1,101	-52,5
	σ_1	0,268	7,8	0,340	8,6	0,332	9,1	-0,288	-6,9	-0,252	-7,7	-0,252	-7,7
	σ_2			0,600	14,0	0,386	16,7	0,903	11,8	0,000	0,0	0	NaN
	$\mathcal{L}/\rho^2$	-4924,7	0,54	-4851,3	0,55	-4868,9	0,55	-4841,6	0,55	-4925,1	0,54	-4925,1	0,54
$-LN$	ASC_1	1,235	27,1	1,434	25,1	1,376	25,6	1,410	25,2	1,429	24,9	1,411	25,1
	ASC_2	1,286	27,8	1,520	26,0	1,458	26,2	1,489	26,2	1,521	25,5	1,492	26,1
	β_1	0,637	69,9	0,566	45,3	0,472	25,3	0,464	21,0	0,575	46,3	0,564	43,6
	β_2	-1,107	-50,7	-1,675	-27,1	-2,559	-50,9	-0,165	-6,3	-1,818	-25,9	-0,704	-25,8
	σ_1	0,267	7,4	0,384	9,3	-0,346	-6,6	-0,312	-4,4	0,406	10,1	0,234	4,3
	σ_2			-0,662	-16,1	-0,409	-22,1	-0,652	-15,7	-0,612	-29,9	-0,979	-21,1
	$\mathcal{L}/\rho^2$	-4924,5	0,54	-4838,3	0,55	-4856,9	0,55	-4844,9	0,55	-4852,2	0,55	-4847,2	0,55
χ^2	ASC_1	1,182	28,3	1,331	26,7	1,274	27,4	1,370	26,4	1,182	28,3	1,182	28,3
	ASC_2	1,223	29,2	1,406	27,5	1,341	28,2	1,447	27,6	1,223	29,2	1,223	29,2
	β_1	-0,394	-47,5	-0,456	-39,5	0,434	-42,6	-0,467	-39,7	-0,394	-47,5	-0,394	-47,5
	β_2	-1,060	-59,8	-1,246	-38,6	2,220	-75,8	-0,133	-5,9	-1,060	-59,8	-1,060	-59,8
	σ_1	0,000	0,0	0	NaN	0	NaN	0,000	0,0	0,000	0,0	0,000	0,0
	σ_2			0,583	14,1	0,371	16,0	-0,658	-15,9	0,000	0,0	0	NaN
	$\mathcal{L}/\rho^2$	-4934,6	0,54	-4862,0	0,55	-4882,5	0,54	-4848,5	0,55	-4934,6	0,54	-4934,6	0,54
$-\chi^2$	ASC_1	1,182	28,3	1,331	26,7	1,274	27,4	1,370	26,4	1,182	28,30	1,182	28,3
	ASC_2	1,223	29,2	1,406	27,5	1,341	28,2	1,447	27,6	1,223	29,20	1,223	29,2
	β_1	-0,394	-47,5	-0,456	-39,5	-0,434	-42,6	-0,467	-39,7	-0,394	-47,50	-0,394	-47,5
	β_2	-1,060	-59,8	-1,246	-38,6	-2,220	-75,8	-0,133	-5,9	-1,060	-59,84	-1,060	-59,8
	σ_1	0,000	NaN	0,000	NaN	0,000	NaN	0,000	0,0	0,000	NaN	0,000	NaN
	σ_2			0,583	14,1	0,371	16,0	-0,658	-15,9	0,000	0,00	0,000	NaN
	$\mathcal{L}/\rho^2$	-4934,6	0,54	-4862,0	0,55	-4882,5	0,54	-4848,5	0,55	-4934,6	0,54	-4934,6	0,54

[6] Previously the models were estimated using 100 pseudorandom numbers, results are provided in Sørensen (2003b).

In table 3 the results of 35 MSL and the MNL model runs are shown, in a matrix style. Recall, N-LN is the 'correct' model. The rows indicate shape of distribution for first coefficient, the columns for the second.For each model the coefficient estimates (left) and tvalues (right) and below the likelihood and ρ^2, are provided.

For interpretation of estimates recall, that the coefficients are structured as $\beta + \xi$ and the ξ is the distributed term. For the distribution, is the estimate of the std. error for the mean zero normal distributed term multiplied by the variable. The mean of coefficient is the value.For the LN distribution, is the estimate of the std.error for the mean zero normal distributed term that was used to generate the lognormal term. Hence, the variance of the ξ is $\exp(\sigma^2)(\exp(\sigma^2) - 1)$ and $E(\xi) = \sqrt{\exp(\sigma^2)}$ is added to the mean of the coefficient due to the skewness of the lognormal distribution. Similarly, for the χ^2 term, which was constructed as the square of a mean zero normal term.

Generally, all models involving the χ^2 distribution did either not have estimates significantly different from zero, or were clearly dominated by other models.

An attempt to replicate a real model building process, by successively adding distribution to one coefficient at a time would and letting a model be determined as 'better' provided that all estimates were significantly different from zero and significant likelihood ratio test for model fit. Depending on how the process was initiated, different final models would be the case!
To demonstrate this, four different distribution selection paths could be followed. These are

Path 1: $FF \rightarrow F - LN \rightarrow LN - LN$
Path 2: $FF \rightarrow NF \rightarrow N - LN \rightarrow LN - LN.$
Path 3: $FF \rightarrow LNF \rightarrow LN - LN.$
Path 4: $FF \rightarrow -LNF \rightarrow -LNN.$

Three of the four paths end in the same, though wrong, combination of distributions namely the LN-LN . The shape for the second coefficient is correctly recovered, possible explanations for the recovery of a skewed distribution for the first coefficient is mixing up with the positively skewed extremevalue residual term.

The fourth path, ends up in a -LNN which is 'doublewrong' as the skewed coefficient is described by a symmetrical distribution, and the symmetrical is described by a skewed distribution. The residual terms does not seem to be the one to blame here, due to differences in sign of the skewness.

The correct combination of distributions is crossed by the second path, although even here a seemingly better combination can be found (LN - LN).

However, it is also the case that wrongly specified mixed logit models did improve the likelihood value significantly compared to the MNL, and especially that the tvalues were high for these models. The difference between the models and the correct specification could be detected mainly by the likelihood values, and neither the ρ^2 were virtually identical nor the tvalues which were in the same range for most parameters across model specifications.

If these results hold in general, there may be a problem with the way model formulations are formed/ tested, as the likelihood and tvalue is generally used as key statistics. It also appears that there are some problems estimating the correct values of the alternative specific constants as well as the coefficients – beyond the impact of the logit scale.

In 'real' data, the significance of a normal variance term may be caused by some attribute(s) left unaccounted for (e.g. data was not available), where the contribution from this particular attribute to explaining the variation, may be described by a normal distributed term. However, in

the case of synthetic data, the formation of choice is known exactly (by the construction of the utility function). In principle, some other formation of data may result in the same data set, estimating close to the same coefficients, though this chance is remote.

7 SODA test for shape of distribution

A question raised by the above numerous model estimations of different model specifications is could the correct shape of distribution(s) have been determined prior to MSL estimation? In the following, a method for determination of the shape of distributions prior to the MSL estimation is described and results are provided. The method for Shape Of Distribution Assessment (SODA) method, is a tool to assist in determining the shape of distribution prior to a MSL estimation but is not an alternative to MSL estimation. The method is briefly described interested readers are referred to Sørensen (2003a).

The method was not proposed as an alternative to MSL, rather as a complimentary method, by which the shape of distribution could be determined prior to the MSL estimation. Hereby, the MSL estimation would be based on the correct shape of distribution and not, as has been demonstrated above some a priori assumption of shape of distribution.

The purpose of the proposed method is to obtain an assessment of the shape of the possible underlying distribution of distributed terms – where there is no priori assumption on shape of distribution! The method seeks to uncover the empirical distribution of the data by repeated model estimations within a discrete choice framework. Preferably, a model should be estimated for each data record and estimates compared to assess the shape of the distribution over individuals (choice situations). The estimations are performed on smaller subsets of data due to identifiability of the models, unless for very special rigorous data sets.

Determination of the shape of the distribution was the main purpose, that is, whether the empirical distribution is unimodal, bell shaped, skewed or any other systematic (describable) shape. The actual parameters of the distribution(s) can subsequently be determined by MSL estimation, conditional on the empirical distribution.

For implementation of the SODA method an algorithm has been provided which is given below with each step described afterwards.

Step 1	Initialisation: Determine number of groups G (or elements in each group, as these figures are inversely related), number of repetitions (fixed G) R and number of different group sizes T
Step 2	Generate random groups
Step 3	Estimate models (a vector of coefficients) for each group
Step 4	Determine (simultaneous) distribution of coefficients
Step 5	Repeat steps 2 through 4, R times
Step 6	Repeat steps 1 through 5, T times
Step 7	Determine (simultaneous) distribution of coefficients based on pooled data

The initialisation (step 1) serves to set the stage. One of the keys to the SODA method is the division of the data sample into groups; the number of groups is inversely related to the number

of observations in each group. Though, at the same time the number of groups should be as large as possible to enable the model estimation to detect possible differences between observations (larger groups tend to pull the coefficient estimates towards the estimates for the model based on the whole sample).
A recommended level for number of repetitions is 10 to 15 (for 100500 groups); this serves to smooth the revealed distribution.For highernumber of groups, the number of repetitions may be reduced. The recommended number of different group sizes is 24; which serves to determine whether the revealed distribution is dependent hereof (neither of the to date tested data set have shown such dependency).

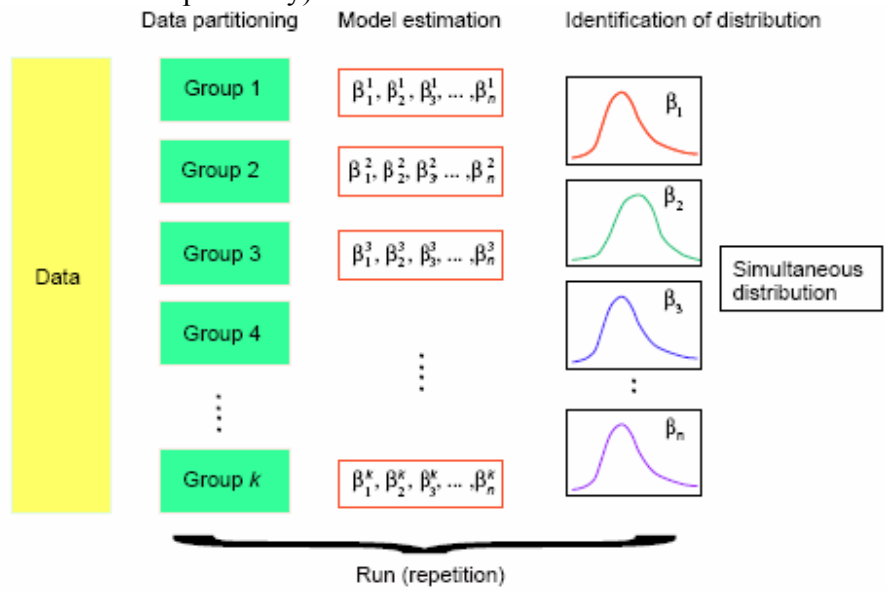

Figure 2: Method for assessment of distribution

After initialisation step 2 proceeds with grouping (data partitioning) of the data. This partitioning is performed by random sampling to preserve the properties of the data. Other means of data partitioning (cluster analysis, proportional to some variable) were not recommended to use as they would make the consecutive model estimation conditional on the criterion used for the data partitioning. At this point it should be noted that the initial groups should be neither too small nor too large. The first case would cause inability to estimate models for each of the groups while the latter would make the assumption of homogeneity within the groups become more critical. In general, the number of groups should be maximised with regards to that no group size is below some threshold. In practice, this is obtained by equal group sizes.

In step 3, estimation of the coefficient for each group follows. The same probability model (e.g. MNL) and the same specification of the utility function is applied for all groups. If the tstatistics for some group, indicate that a coefficient should be excluded, this is ignored, as this would lead to a larger error, than the lack of significance of one coefficient at this stage of the model will give rise to.

Based on the coefficient estimates histograms are produced for each of the coefficients, in step 4. The objective is to get an intuition on which shape applies to each of the coefficient distributions; shapes may differ for different coefficients. Based on the graphical evidence, the list of possible shapes can be narrowed, e.g. the shape is unimodal, flat, thick tailed, etc. When the shape is determined most attention should be directed to the region around the mode, as the

method has a tendency to produce outliers (some numerically high coefficient estimates). The determination of shape, is further aided by the calculation of the first four central moments; the mean, the variance, the skewness and the kurtosis.

Finally, in step 7 all the coefficient estimates are pooled and a search for distribution is conducted (identical to that in step 4). The pooling of data implicitly assigns an equal weight to all the coefficient estimates. The methodis illustrated in figure 2 working from left to right.

8 SODA results

This section describes the results from application of the SODA method to the synthetic data sets. The charts of distribution of β and -β2 (note, the axis has been reversed) are shown in figure 3 and 4). For both the charts, the horizontal axes are identical, wherefore differences in shape of distributions and spread can easily be seen. The blue dotted respective the red line are the fitted normal and lognormal density functions. However, it should be kept in mind that the data was generated as one normal and one lognormal distributed term, where records with wrong sign of attributes and/ or coefficients had been removed.

The chart for β1 in figure 3, indicate that the first coefficient is signrestricted which complies with the data generation. The shape is unimodal with some variation around the mode. The lognormal curve does not fit the density mode nor the area around the mode particularly well, hence it is not believed that this is the correct distribution. However, neither the normal curve as plottet here seems to fit the data, although this is partly due to a restriction in the applied statistical software. Shifting the blue curve to the left (approximately 0.2 on the axis) enables the normal curve to fit the mode and the region around.

However, the chart for β2 is more clear. The distribution is unimodal and skewed and the nonnormality seems obvious. The lognormal density function fits the data very well whereas the normal curve is not doing a particular fine job. Furthermore, the differences in the shape of distributions, as well as the difference in the variance levels (var(β2)> var(β)) have both been recovered.

Jointly, the two shapes of distribution has been identified and can be used a a priori information for a MSL estimation, as proposed[7].

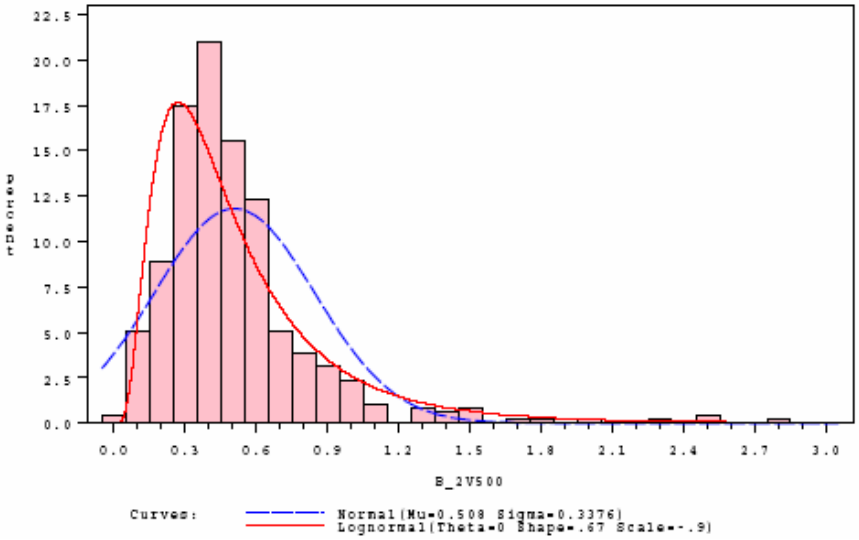

Figure 3: Distribution of β3

[7] Readers interested in applications of the SODA method on real data, are referred to Sørensen (2003a).

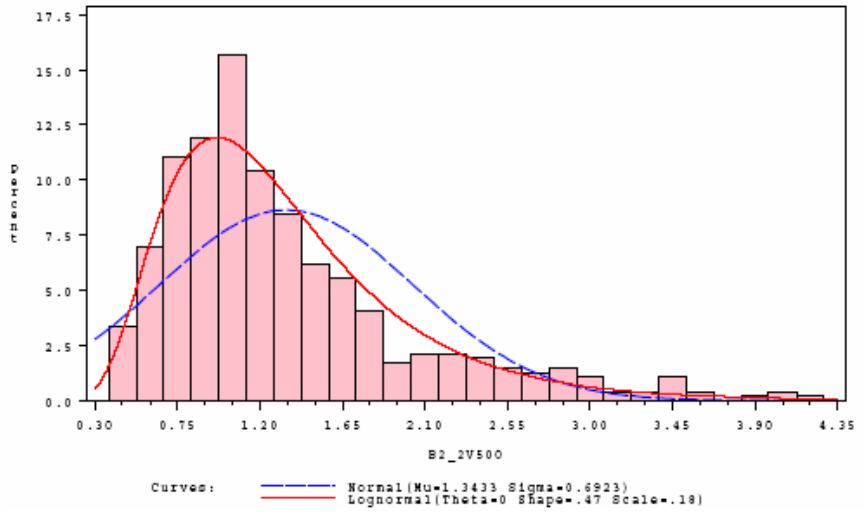

Figure 4: Distribution of β2

9 Conclusion

The paper has demonstrated that specification and estimation of a mixed logit model is not a simple task. The number of alternative specifications of shape of distributed terms grows rapidly with number of coefficients expected to be distributed and number of considered shapes of distribution.

Even if a great deal of these alternative specifications are tested by means of mixed model estimation and results are compared, the correct combination of shape of distributions is not guaranteed to be found. The importance of correctly specifying the probability model and the utility function was emphasised by the test results, as erroneously specified models could not always be detected from the coefficient estimates/ model results.

A synthetic data set was constructed and mixed models based on 36 different a priori assumptions of shape of distributions were estimated. Comparisons of estimation results showed that half of the model estimations did provide estimates significantly different from zero and improved fit of data in terms of likelihood value compared to the 'fixed coefficient' model. At this point, only few different shapes of distributions have been tested. Had more shapes of distributions been tested, it may have turned out that other distributions had performed better, but the N LN would under no circumstances have had come out as the best.

Hereby, estimating a model with nice tvalues, does not imply the 'best' model specification has been found as several differently shaped distributed terms can have a standard error significantly different from zero. Further, depending on how alternative specifications were introduced e.g adding distribution to one coefficient at a time, did result in different final model specifications!

As the comparisons of the model results showed most of the distributed terms or rather the estimate of the standard error for their distribution, turned out to be significantly different from zero! This implies that estimating a model and recovering a standard error for some distributed term significantly different from zero, does not imply the 'best' model specification has been found as numbers of differently shaped distributed terms can have a standard error significantly different from zero.

However, little comfort is left to the modeller as irrespective of by distributions the coefficients were assumed to follow, the effect in likelihoodvalue was positive and all models had likelihoodvalues significantly better than the 'fixed coefficient' model.

To assist in the search for the correct model specification the SODA method was described and applied to the data. By application of the SODA method to the data the 'correct' shapes of distribution were pointed at – sufficiently clear to enable the modeller to chose which shape of distribution to assume for the subsequent MSL estimation.

The importance of correctly specifying the probability model and the utility function is emphasised bythe test results, as erroneously specified models cannot always be detected from the coefficient estimates/ model results.

References

Ben Akiva, M., Bolduc, D., and Bradley, M., 1993. Estimation of travel choice models with random distributed values of time. Transportation Research Record, 1413:88–97.

Bhat, C. R., 1997. Covariance heterogeneity in nested logit models: econometric structure and application to intercity travel. Transportation Research, Part B, 31B (1)11–21.

Boyd, J.H., and Mellman, R.E., 1980. The effect of fuel economy standards on the us automotive market: A hedonic demand analysis. Transportation Research, Part A., 14A (56) 367–78.

Cardell, N., and Dunbar, F., 1980. Measuring the societal impacts of automobile downsizing. Transportation Research, Part A, 14A (56) 423–34.

Hajivassiliou, V., and Ruud, P., 1994. Classical estimation methods for ldv models using simulation. In Handbook of econometrics. Elsevier Sciences, Eds. Engle and McFadden.

Hensher, D.A., and Reyes, A.J., 2000. Trip chaining as a barrier to the propensity to use public transport. Transportation, 27 (4) 341–361.

McFadden, D., and Train, K.E., 2000. Mixed mnl models for discrete response. Journal of Applied Econometrics, 15 (5) 447–70.

Nielsen, O.A., Hansen, C.O., and Daly, A., 2001. A largescale model system for the CopenhagenRingsted railway project, chapter 35, pages 603–26. Travel Behaviour Research, The Leading Edge. Pergamon Press, Australia, Edited by D. Hensher.

Nunes, L.C., Cunha e S`a, M. A.M., Ducla-Soares, M., Rosado, M.A., and Day, B.H., 2001. Identifying nonconsistent choice behavior in recreation demand models. Econometric Letters, 72 (3) 403–410.

Peitz, M., 1995. Utility maximization in models of discrete choice. Economic Letters, 49 91–94.

Revelt, D., and Train, K.E., 2000. Customerspecific taste parameters and mixed logit. Working Paper 00274, Department of Economics, University of California, Berkeley.

Ross, S.M., 1997. Simulation. Academic Press Inc., New York, USA, 2nd edition.

Siikamäki, J., and Layton, D., 2001. Pooled models for contingent valuation and contingent ranking data. Working Paper, University of California, Davis.

Sørensen, M.V., 2002. Error components in demand estimation. In Proceedings of European Transport Conference, Cambridge, UK, September, CDrom.

Sørensen, M.V., 2003a. Discrete Choice Models. Estimation of passenger traffic. PhD thesis, Danish Technical University, DK2800 Kgs. Lyngby, Denmark.

Sørensen, M.V., October 2003b. Msl for mixed logit model estimation on shape of distributions. In Proceedings of European Transport Conference, Strasbourg, France, CDrom.

Sørensen, M.V., 2004. Discrete choice models. similarities and differences. working paper.

Sørensen, M.V., and Nielsen, O.A., July 2001. Assessing distributions of error components in a complex traffic model. In Proceedings of ninth World Conference on Transport Research, Seoul, Korea.

Sørensen, M.V., and Nielsen, O.A., 2002. Empirical distribution of error components. Submitted to journal.

Train, K.E., 1999. Recreation demand models with taste variation. Land Economics, 74(2) 230–239.

Train, K.E., 2001. A comparison of hierarchical bayes and maximum simulated likelihood for mixed logit. Working paper, Department of Economics, University of California, Berkeley.

Train, K.E., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, New York, NY, USA.