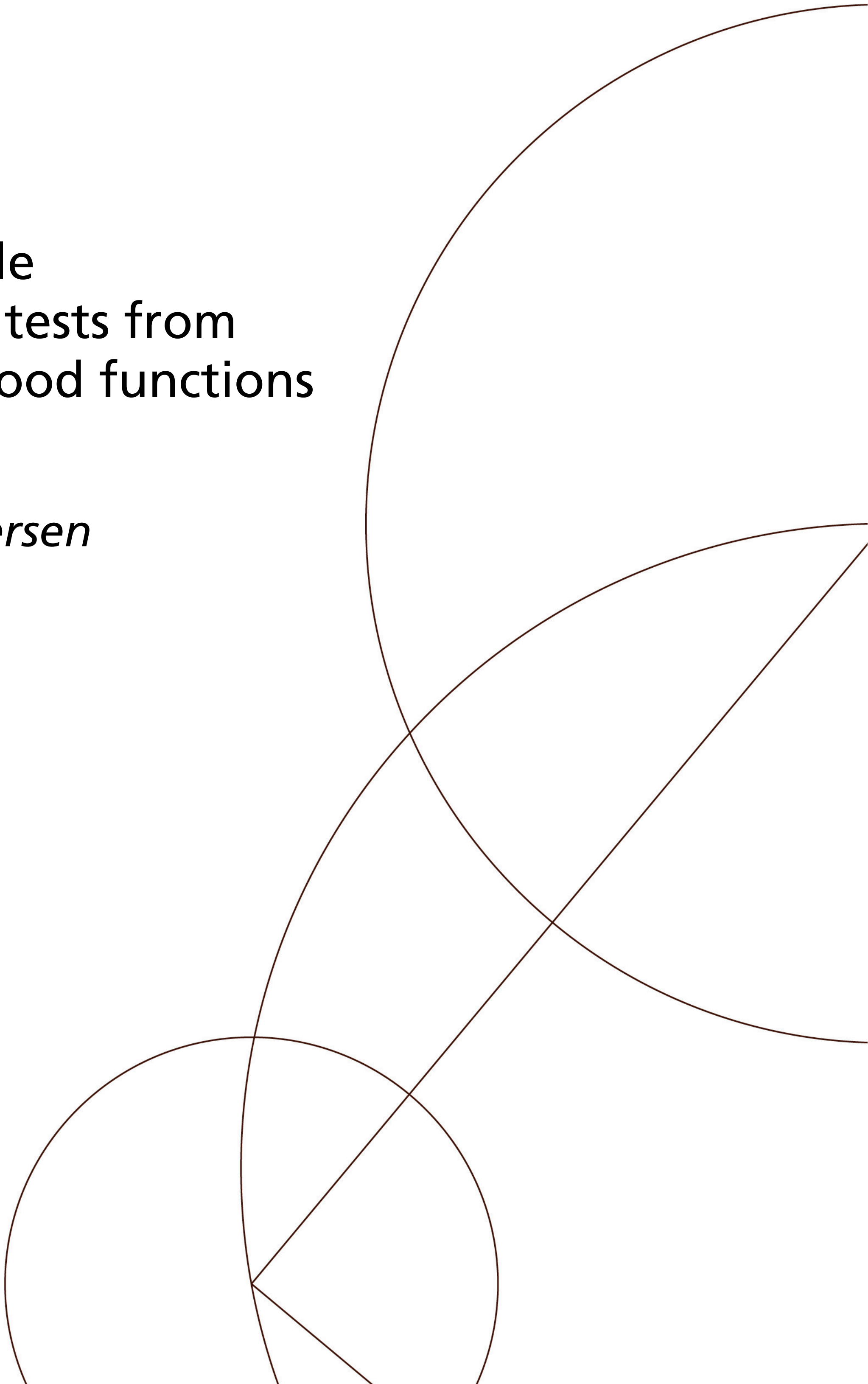
DEPARTMENT OF FOOD AND RESOURCE ECONOMICS
UNIVERSITY OF COPENHAGEN

IFRO Working Paper



Obtaining reliable Likelihood Ratio tests from simulated likelihood functions

Laura Mørch Andersen

2013 / 1

IFRO Working Paper 2013 / 1

Obtaining reliable Likelihood Ratio tests from simulated likelihood functions

Author: Laura Mørch Andersen

Department of Food and Resource Economics (IFRO)
University of Copenhagen
Rolighedsvej 25
DK 1958 Frederiksberg DENMARK
www.ifro.ku.dk (www.foi.life.ku.dk)

Obtaining reliable Likelihood Ratio tests from simulated likelihood functions[*]

Laura Mørch Andersen[†]

January 2013

Abstract

It is standard practice by researchers and the default option in many statistical programs to base test statistics for mixed models on simulations using asymmetric draws (e.g. Halton draws). This paper shows that when the estimated likelihood functions depend on standard deviations of mixed parameters this practice is very likely to cause misleading test results for the number of draws usually used today. The paper shows that increasing the number of draws is a very inefficient solution strategy requiring very large numbers of draws to ensure against misleading test statistics. The paper shows that using one dimensionally antithetic draws does not solve the problem but that the problem can be solved completely by using fully antithetic draws. The paper also shows that even when fully antithetic draws are used, models testing away mixing dimensions must replicate the relevant dimensions of the quasi-random draws in the simulation of the restricted likelihood. Again this is not standard in research or statistical programs. The paper therefore recommends using fully antithetic draws replicating the relevant dimensions of the quasi-random draws in the simulation of the restricted likelihood and that this should become the default option in statistical programs.

JEL classification: C15; C25.

Key words: Quasi-Monte Carlo integration; Antithetic draws; Likelihood Ratio tests; simulated likelihood; panel Mixed MultiNomial Logit; Halton draws

---

[*] I thank Kenneth Train, David Revelt and Paul Ruud for making their MMNL software available at Train's MMNL homepage: elsa.berkeley.edu/software/abstracts/train0296.html, and for allowing me to use a later version, which allows for correlation between mixed parameters.

The research was funded by The Danish Social Science Research Council ('FSE').

[†] Contact information: Laura Mørch Andersen, Institute of Food and Resource Economics, University of Copenhagen, www.foi.dk e-mail: LA@foi.ku.dk

1. Introduction

Models allowing for heterogeneity have been developing rapidly thanks to advances in computational speed and understanding of simulation methods for approximating integrals, (see e.g. Ben-Akiva et al., 1993; Ben-Akiva and Bolduc, 1996; Berry et al., 1995; Bhat, 1996; Brownstone and Train, 1999 or Geweke 1996). One way of introducing heterogeneity is to assume that value of one or more parameters follow a specified distribution. This is known as 'mixing' parameters, and the result of the estimation is the moments characterising the mixing distribution rather than a single value of the mixed parameter. Calculating the likelihood of a mixed model means that a conventional likelihood must be integrated over all possible values of the mixed parameters weighted by the mixing density. Often, this integral does not have a closed form and the integral is therefore approximated by Quasi-Monte Carlo integration. This means that random values of the mixed parameter are drawn quasi randomly from the underlying distribution and used to calculate the numerical integral which is then used as an approximation.

Many models are estimated by simulated maximum likelihood, and restrictions easily tested using Likelihood Ratio tests. As an example, the Mixed MultiNomial Logit (MMNL, McFadden and Train, 2000) is becoming an attractive way for researchers to introduce heterogeneity into discrete models, and Mixed MultiNomial Logit is now available in many different software packages. Chang and Lusk (2011) compare the accuracy of Mixed MultiNomial Logit estimation in SAS, NLOGIT-LIMDEP and a user-written add-in module for Stata. They find it curious that both SAS and the Stata module allows the estimated standard deviations of the mixing distributions to be negative, and that both packages advice the user to reverse the sign in these cases. There is however nothing strange about the negative values of the estimated standard deviations of the mixing distributions. If the mixed distribution depends on the value of a variance, it is common to maximize the likelihood value over the standard deviation which is the square root of the variance. This relationship between the variance and the standard deviation means that the likelihood function will be symmetric around zero for the standard deviation, and the simulated likelihood must therefore also be symmetric around zero in this dimension.[1]

This paper illustrates that if the conventional likelihood is symmetric the simulated mixed likelihoods will always be symmetric, but if the conventional likelihood is asymmetric this is not true unless the draws for the Quasi Monte Carlo integration are also symmetric around zero. Antithetic draws have been suggested as a variance reduction technique for Monte Carlo integration of asymmetric functions, especially for Bayesian inference (see e.g. Geweke 1988 or 1996). However, the technique is not generally used in mixed models which also use Monte Carlo integration, and the simulated mixed likelihoods are therefore usually not symmetric. As the number of draws increases the degree of symmetry will increase, but as it will also be illustrated in the following, the degree of symmetry is not always sufficient within the range of draws usually applied. One example of an asymmetric conventional likelihood is the likelihood of a logit model, which will be used to illustrate the problem in this paper.

Antithetic Halton draws have also been suggested as an instrument for faster computation of Quasi-Monte Carlo integrals, allowing for more precise point estimates within a reasonable time frame (recently in Sidharthan and Srinivasan, 2010). In the present paper we focus on the reliability of Likelihood Ratio tests of mixed models instead of precision of point estimates or estimation speed. If Mixed MultiNomial Logit models are estimated without antithetic draws, Likelihood Ratio tests may be compromised, which again may lead to false conclusions. This paper illustrates why asymmetric draws are likely to lead to false Likelihood Ratio tests, and why antithetic draws solve this problem. The purpose of this paper is to

---

[1] The statistical software Biogeme also sometimes report negative standard deviations, and the results of estimations using Biogeme vary depending on the sign of the starting values for the standard deviations. As will be illustrated in the following, this is problematic because standard deviations with identical absolute values should lead to the same likelihood value independent of the sign, and therefore also to the same result when maximizing the likelihood.

support the idea of using fully antithetic draws for simulated likelihoods, and not least to warn researchers who use standard estimation procedures for e.g. Mixed MultiNomial Logit, that their Likelihood Ratio tests may be invalidated by the asymmetric nature of the draws.

Using data simulated under a Mixed MultiNomial Logit model, we illustrate how asymmetry of the simulated likelihood function causes the likelihood to depend on the signs of the estimated Choleski factorization, and that *the problem of inconsistent Likelihood Ratio tests caused by the asymmetry of the quasi random draws is completely removed when one uses fully antithetic draws instead of conventional asymmetric draws.* We also show that the same mechanisms appear in a real data set with invalidating implications for Likelihood Ratio tests.

Having solved the problem of inconsistent Likelihood Ratio tests caused by asymmetric draws, the paper turns to another problem which could still invalidate Likelihood Ratio tests, even when antithetic draws are used. When restricting the number of dimensions of the mixing distribution from *n* to *n*-1 most standard procedures simply estimate the restricted model using the *n*-1 first set of quasi random draws, irrespective of which of the *n* original dimensions is restricted. As illustrated in this paper, keeping track of which dimension is restricted, and removing exactly this dimension when estimating the restricted model, may lead to better Likelihood Ratio tests.

The problem and solution presented in this paper not only applies to Mixed MultiNomial Logit models, but also to other models estimated by maximum simulated likelihood and the paper therefore provides a valuable contribution to the ongoing struggle to improve simulation methods. Halton draws are used in this paper, but the properties of the antithetic draws can be generalized to other types of draws.

The structure of the paper is: The above section introduced the problems associated with Likelihood Ratio tests performed on simulated log-likelihood values, if these are simulated using asymmetric draws, and Section 2 outlines the standard way of estimating and testing within panel Mixed MultiNomial Logit models. Section 3 explains why asymmetric draws used in Quasi-Monte Carlo integration of mixed likelihoods may invalidate Likelihood Ratio tests, and section 4 illustrates the problem on simulated data by comparing results in the true optimum using conventional Halton draws. Section 5 illustrates the problem using real data and section 6 introduces antithetic Halton draws. Section 7 presents the encouraging results of using this type of draws and Section 9 concludes.[2]

2. Estimation and Testing in Panel Mixed Logit Models

In a conventional logit (McFadden, 1973) it is assumed that all individuals have the same utility function, but in a Mixed MultiNomial Logit (MMNL or MXL) model[3] (McFadden and Train, 2000), it is assumed that (part of) the individual utility is drawn from a distribution. This means that the individual utility is known to the individual, but only the distribution is assumed to be observable to the econometrician. The mixed likelihood function is then the likelihood function of the conventional multinomial logit model integrated over all possible values of β, which in a panel mixed logit becomes (Revelt and Train, 1998):

$$L(\theta)=\prod_{i=1}^{I} L_i(\theta)=\prod_{i=1}^{I}\left(\int p\left(y_i \middle| \beta_i, x_i\right) f\left(\beta_i \middle| \theta\right) d\beta_i\right) \tag{1}$$

where $L(\theta)$ is the likelihood of the mixed logit given the mixing distribution of β given by θ, $L_i(\theta)$ is the likelihood for individual i, I is the number of individuals, $p\left(y \middle| \beta, x\right)$ is the likelihood of a conventional logit model given β and $f\left(\beta \middle| \theta\right)$ is the density of β given θ. The likelihood function is maximized over θ.

Calculating the likelihood function in equation (1) is very cumbersome, especially if β follows a multivariate distribution, but the problem can be reduced significantly by using Quasi-Monte

---

[2] Appendix A presents the simulated data used in this paper and Appendix B illustrates the differences in log-likelihood values in optimum in different quadrants.

[3] Also known as Random Parameter Logit (RPL).

Carlo integration.[4] This is often done using quasi random Halton sequences which were first presented by Hammersley (1960) and Halton (1960). The efficiency of Halton sequences compared to random draws is discussed in detail in both Train (1999) and Baht (2001). Both find that Halton sequences greatly improve accuracy, allowing for far fewer draws and faster computation. According to Hensher and Greene (2003), there is no standard for the number of draws needed, but they find that 100 draws appears to be a 'good' number. In order to validate the model Hensher and Greene suggest that the models are estimated over a range of draws from 25 to 2,000.[5]

When optimizing a mixed likelihood function by Quasi-Monte Carlo integration the optimization routine uses the same set of quasi random draws for all potential values of the means and standard deviations of the mixing distribution. This is done in order to ensure that the simulated likelihood values for different values of means and standard deviations are as comparable as possible. As shall be illustrated in the following, the property of identical draws are in some cases violated if the quasi random draws are asymmetric, and as will also be illustrated in the following, this may have serious consequences for the inference of the estimated models.

In many cases the purpose of estimating a likelihood function is twofold: Maximizing the likelihood function leads to the set of parameters which fit the data best, and comparing the best likelihood values of different models makes it possible to infer whether the models are significantly different. The latter is done by Likelihood Ratio (LR) tests based on the

---

[4] see e.g. Morokoff and Caflisch, 1995, for asymptotic properties.

[5] Clearly, reliable estimation, validation and inference techniques are a prerequisite for sound models and analysis. The use of simulated likelihoods are bound to induce some approximation error, and it is therefore important to validate the results, e.g. by varying the starting values of the parameters and checking the stability of the results. One example of a simulation error is investigated in Chiou and Walker (2007), who illustrated that a low number of draws in the simulation of the integral may lead to unidentified estimates. Ben-Akiva and Boulduc (1996), Walker (2001), Hensher and Greene (2003) and Walker et al. (2007) confirm that in general the validity of the results is greatly influenced by the number of draws, which should therefore also be varied.

difference between the restricted and the unrestricted likelihood values (see e.g. Greene, 1997). In theory a statistical model has one and only one maximum log-likelihood value, but in some cases the value of the simulated maximum log-likelihood varies with the starting values. This phenomenon is usually ascribed to local maxima, usually occurring if the variation in data is not large enough to support the statistical model. If local maxima is the problem both the value of the likelihood function and the estimated parameters will vary with the starting values, and the problem is often easily recognized. This paper however point to a more serious type of variation in the value of the maximum log-likelihood. We find that the use of asymmetric draws for Quasi-Monte Carlo simulation of mixed likelihoods may lead to cases where the parameters of the statistical model are uniquely identified, but the maximum likelihood function varies. The absolute level of the log-likelihood function is of no interest, but if the difference in the value of the simulated log-likelihood given different starting values of the parameters is above e.g. two, testing hypotheses may easily lead to false conclusions.[6] The varying values of the log-likelihood function may first of all lead to falsely accepted or rejected hypotheses. Secondly, it may also falsely indicate that data are not informative enough to support the model, and therefore lead to unnecessary reductions in model complexity.

3. The effect of asymmetric draws on simulated log-likelihood values

If one parameter in a conventional logit model is mixed with e.g. the normal distribution, two moments characterizing the distribution are estimated, a mean and a standard deviation. When estimating the mean and the standard deviation of the mixing distribution, these are usually both maximized over the entire real axis $\mathbb{R}$. Actually, the mixing distribution depends on the *variance* rather than the standard deviation, and since the variance is the square of

---

[6] If the difference between the unrestricted and the restricted log-likelihood is above 1.92 the null hypothesis will be rejected at the five percent level in a test with one degree of freedom.

the standard deviation the true mixed likelihood function of the standard deviation will be symmetric around zero.

In most cases however, the true value of the likelihood value of a Mixed MultiNomial Logit model is not possible to obtain, and instead the value is often simulated using Quasi-Monte Carlo integration. This means that the value of the likelihood function of the conventional likelihood is calculated for a number of quasi randomly drawn β's and the average of these values is used as an approximation of the integral in equation (1). Formally the likelihood function for the entire sample becomes:

$$L(\theta)=\prod_{i=1}^{I}\left(\int p\left(y_i|\beta_i,x_i\right)f\left(\beta_i|\theta\right)d\beta_i\right)\approx\prod_{i=1}^{I}\left(\frac{1}{R}\sum_{r=1}^{R}\left(p\left(y_i|\beta_{ir},x_i\right)\right)\right) \tag{2}$$

where R is the number of draws and the $R\cdot I$ β_{ir}'s are drawn quasi randomly from the mixing distribution given by θ.

If the mixing distribution is one dimensional then for each individual *i*, the r^{th} draw is created as

$$\beta_{ir}\big|_{\theta}=\theta_{std}\gamma_{ir}+\theta_{mean} \tag{3}$$

where $\beta_{ir}\big|_{\theta}$ is β_{ir} given θ, θ_{std} is the standard deviation of the mixing distribution, θ_{mean} is the mean of the mixing distribution and γ_{ir} is a quasi random draw from a standard normal distribution with mean zero and standard deviation one. In the following we will investigate the effect of symmetry versus asymmetry of the draws β_{ir} around the mean θ_{mean}, and for simplicity θ_{mean} is therefore set to zero. When θ_{mean} is zero, $\beta_{ir}\big|_{\theta}=\theta_{std}\gamma_{ir}$ for all i and all r, and therefore:

$$\beta_{ir}\big|_{\theta}=-\beta_{ir}\big|_{-\theta} \tag{4}$$

where $\beta_{ir}\big|_{-\theta}=-\theta_{std}\gamma_{ir}$.

If the conventional likelihood *p* is symmetric in β_{ir} then $p\left(y_i\middle|\beta_{ir}|_\theta, x_i\right) = p\left(y_i\middle|-\beta_{ir}|_\theta, x_i\right)$, which because of (4) also means that $p\left(y_i\middle|\beta_{ir}|_\theta, x_i\right) = p\left(y_i\middle|-\beta_{ir}|_{-\theta}, x_i\right) = p\left(y_i\middle|\beta_{ir}|_{-\theta}, x_i\right)$ for all values of i, r and θ, and therefore that the Quasi-Monte Carlo integral in equation (2) is independent of the sign of the standard deviation. If p is not symmetric in β_{ir}, then the Quasi-Monte Carlo integral is only identical for θ and $-\theta$ if for every combination of i and r there exists a value *s* so that $p\left(y_i\middle|\beta_{ir}|_\theta, x_i\right) = p\left(y_i\middle|\beta_{is}|_{-\theta}, x_i\right)$. If the quasi random draws γ_{ir} are symmetric around zero then this is always the case, but if the quasi random draws are not symmetric around zero then $\beta_{ir}|_\theta \neq \beta_{ir}|_{-\theta}$, at least for some combinations of i's and r's.

This means that the simulated mixed likelihood of θ is not identical to the simulated mixed likelihood of $-\theta$ $\left(\prod_{i=1}^{I}\left(R^{-1}\sum_{r=1}^{R}\left(p\left(y_i\middle|\left(\beta_{ir}|_\theta\right), x_i\right)\right)\right) \neq \prod_{i=1}^{I}\left(R^{-1}\sum_{r=1}^{R}\left(p\left(y_i\middle|\left(\beta_{ir}|_{-\theta}\right), x_i\right)\right)\right)\right)$ when an asymmetric likelihood is mixed using asymmetric draws, because the draws are not identical on the positive and the negative half of the real axis $\mathbb{R}$, and the results using θ or $-\theta$ therefore are incomparable. The sign of the standard deviation therefore influences the value of the simulated mixed likelihood, even though the sign has no influence on the variance of the mixing distribution. Clearly, the problem decreases as the number of draws increases, and the symmetry of the draws therefore increases, and the important question is therefore whether the variation in likelihood values caused by asymmetric draws poses a real problem within the range of draws usually used today. As will be illustrated in the following sections, this is unfortunately the case.

If the mixing distribution is multivariate, the draws are created as $\beta_{ir}\middle|\theta_{chol}, \theta_{mean} = Q\gamma_{ir} + \theta_{mean}$ where Q is the triangular Choleski factorization and θ_{mean} is a vector of means.[7] In the case of an *n*-dimensional mixing, the variance-covariance matrix may be obtained by 2^n different Choleski factorizations with different combinations of signs of the elements of the triangular

---

[7] The Choleski factorization (Q) is a triangular matrix with the property $QQ' = \Omega$, where Ω is the variance-covariance matrix (Train 2003). If the variance-covariance matrix is diagonal (i.e. no correlations) the Choleski matrix is merely a diagonal matrix of standard deviations.

matrix. This means that if an asymmetric likelihood function such as the conventional logit model is mixed with a distribution which depends on a Choleski factorization of a variance-covariance matrix, using asymmetric draws, the combination of signs of the estimated Choleski factorization will wary between quadrants, but all lead to the same variance-covariance matrix. The values of the estimates are therefore not affected by the lack of symmetry of the quasi random draws, but the optimal value of the mixed likelihood may differ between the 2^n quadrants. As shall be illustrated in the following section, this problem can in some cases invalidate Likelihood Ratio tests.

4. Illustrating the problem using simulated data

In order to investigate the magnitude of the problem under controlled conditions, a hypothetical data set has been simulated. The data are panel data with 1,000 individuals each making 20 choices between 4 alternatives. The utility of the alternative specific constant is zero for the alternative which is used as base; the utility of the remaining alternatives follows a three-dimensional normal distribution with no correlation and nonzero means. In the case of three mixings, the variance-covariance matrix is estimated in $\mathbb{R}^3$, which means that the number of different quadrants is $2^3 = 8$, and the likelihood function must therefore be symmetric in all eight quadrants.

One of the virtues of simulated data is that the true mean and variance-covariance of the mixing distribution are known,[8] and in the rest of this section, the likelihood values will be evaluated in the true mean and variance-covariance of the simulated data for each of the eight quadrants, and compared between the different quadrants.[9] The evaluation is done by

---

[8] The finite nature of the simulated data means that the true (realized) means and standard deviations of utility in the sample are not identical with the values used in the simulation of the data. Appendix A describes the differences.

[9] The purpose of this section is to illustrate the potential problem using simulated data, ignoring the optimization errors caused by different optimization routines. We use therefore use the term ‘evaluated’ rather than ‘estimated’ because the likelihood function is not *optimized* over each quadrant, but instead *calculated* by Monte Carlo integration in the optimal point which is known because the data is simulated.

Quasi Monte Carlo integration, using standard Halton draws. In these calculations, the probability of finding an optimum in a given quadrant is treated as equal for all quadrants. In actual estimations the probability of ending up in a given quadrant may well vary, and the results in this section therefore only illustrate the *magnitude* of the problems that may *potentially* arise from actual estimations. As will be illustrated in section 5, estimations on actual data lead to results in all eight quadrants, so the problem also exists when the modes of the distribution of the mixed parameters are optimized rather than known a priory.

Table 1 compares the results from the eight quadrants for increasing numbers of draws. The Choleski factorization of the variance-covariance matrix of the mixing distribution in the different quadrants vary by the combination of signs of the Choleski factorization, but all leads to the same (true) variance-covariance matrix and all have the same (true) means. For 100 draws, the highest difference between the log-likelihood values of different quadrants is 9.26 which is definitely not zero as it should theoretically be. As mentioned in section 2, differences of this magnitude can ruin Likelihood Ratio tests completely. The difference between the quadrants decreases as the number of draws increases, simply because the distance between draws is reduced, but it does not disappear within a feasible span of draws, and certainly not for the low number of draws recommended in Hensher and Greene (2003).[10]

---

[10] The difference in likelihood values between quadrants is generally higher outside the optimum. This problem decreases to some extent with the number of draws, but as illustrated in Andersen (2008), the difference still does not disappear even with 7,500 draws, and the problem is sometimes smaller for 5,000 draws than for 7,500 draws. This has to do with the degree of symmetry of the Halton draws. As illustrated in Andersen (2008), the degree of symmetry increases as the number of draws increases, but not monotonically. Increasing the number of draws by a few thousand may therefore lead to set of draws with a lower degree of symmetry, and thereby a bigger difference between the likelihood values in different quadrants. This might be the reason why Train (2003, page 230) find lower standard deviations of the estimated parameters using 100 Halton draws compared to 1,000 Halton draws. This is not explored in the present paper, but would be interesting to investigate further in the future.

Table 1 Variation in simulated log-likelihood, by number of draws, simulated data

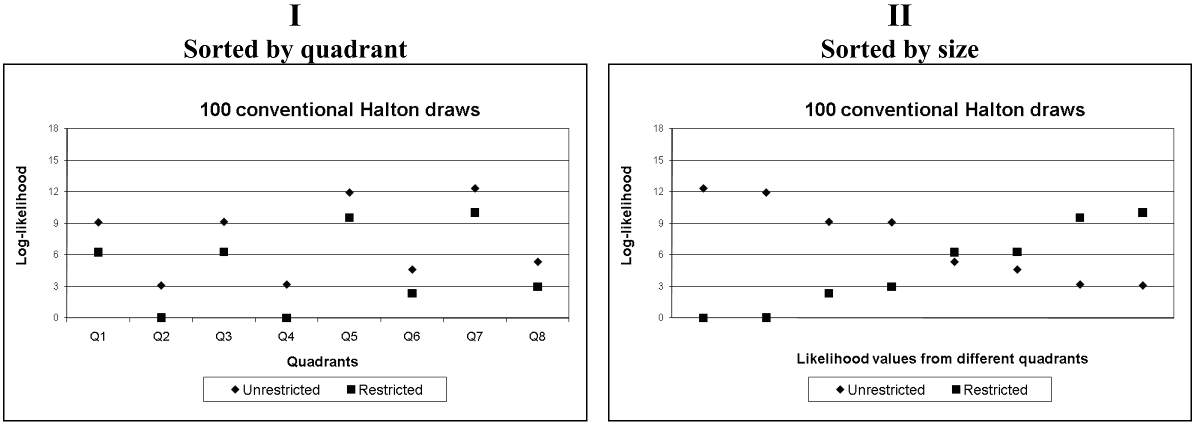
	Number of draws per individual					
	100	**500**	**1,000**	**1,500**	**5,000**	**7,500**
Highest absolute difference in simulated log-likelihood, evaluated in true means and true variance-covariance matrix in different quadrants	9.26	4.44	1.09	0.88	0.47	0.19

Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants. Simulated data, 1,000 individuals, 20 observations per individual. Data is defined in Appendix A. Calculations conducted in the MMNL GAUSS program developed by Train, Revelt and Ruud, using conventional Halton draws.

As mentioned above, one of the problems caused by the difference between the values of the log-likelihood function evaluated at different quadrants is that it influences the results of Likelihood Ratio tests. The rest of this section investigates the effect on a Likelihood Ratio test of the null hypothesis that the mean utility of alternative B is zero. Note that the base alternative A has both mean *and* standard deviation equal to zero. Testing away the mean of alternative B is therefore *not* the same as testing whether the utility of alternative B is the same as the utility of the base alternative A.

The large variation in the value of the log-likelihood function means that the value of the restricted model in one quadrant may be higher than the value of the unrestricted model in another quadrant, but never within quadrants. Figure 1 shows the log-likelihood values for the unrestricted and the restricted model using 100 Halton draws. **I** shows the relationship between the two models in each quadrant and **II** ignores the quadrants and sorts the eight values by size. Especially from **II** it is evident that the value of the restricted model will sometimes be higher than the value of the unrestricted model, leading to negative values of the LR test statistic.

Figure 1 Differences between log-likelihood values of unrestricted and restricted model



I
Sorted by quadrant

II
Sorted by size

Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants. Only the difference between likelihoods is interesting, and the lowest estimated log-likelihood value (from the restricted model in Q4) is therefore subtracted from all the estimated values. Simulated data, 1,000 individuals, 20 observations per individual. Data is defined in Appendix A. Calculations conducted in the MMNL GAUSS program developed by Train, Revelt and Ruud, using conventional Halton draws. The null hypothesis is that the mean utility of alternative B is zero.

Table 2 shows that for 100 conventional Halton draws, the LR test statistic will become negative in 20 of the 64 different combinations of restricted and unrestricted log-likelihood values, corresponding to 31 per cent of the cases. The problem decreases with the number of draws, but is still present at 1,000 draws. Appendix B repeats Figure 1 for all the different numbers of draws presented in Table 2.

Table 2 Testing away one mean using conventional Halton draws

	Number of draws per individual:					
	100	**500**	**1,000**	**1,500**	**5,000**	**7,500**
Share of negative LR values	.31	.31	.05	.00	.00	.00
Results of positive LR values						
Lowest p-value	.00	.00	.03	.05	.08	.12
Highest p-value	.64	.66	.77	.65	.27	.20
Standard deviation of p-values[a]	.13	.16	.15	.13	.05	.02

Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants. Simulated data, 1,000 individuals, 20 observations per individual. Data is defined in Appendix A. Calculations conducted in the MMNL GAUSS program developed by Train, Revelt and Ruud, using conventional Halton draws. The null hypothesis is that the mean utility of alternative B is zero.
a: Standard deviation calculated from the p-values of the positive LR values from the 64 different combinations of quadrants.

Table 2 also summarizes the results of the Likelihood Ratio tests that can be performed on the positive LR test statistics. The p-values vary from zero to 64 per cent for 100 draws leading to a standard deviation of 13 per cent. Note that these tests are all performed on the *same* data set. Had the test been performed on different realizations of data with identical values of the mean and the variance-covariance, the test should have been accepted on 10 per cent of the data sets at the ten percent level, but when the tests are performed at the same dataset the results should all be identical. The differences are caused by the asymmetry of the

Halton draws used in the Quasi-Monte Carlo integration of the likelihood, not by statistical properties of the test.

The standard deviation of the p-values presented in Table 2 is of course deeply problematic, because it means that the result of the Likelihood Ratio test is likely to be unreliable. Table 3 shows the probability of rejecting the null hypothesis at different significance levels.

Table 3 Probability of rejecting the null hypothesis

	Number of draws per individual:					
	100	**500**	**1,000**	**1,500**	**5,000**	**7,500**
At the 1 per cent level	.55	.25	.00	.00	.00	.00
At the 5 per cent level	.89	.45	.07	.05	.00	.00
At the 10 per cent level	.91	.59	.28	.30	.17	.00

Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants. Simulated data, 1,000 individuals, 20 observations per individual. Data is defined in Appendix A. Calculations conducted in the MMNL GAUSS program developed by Train, Revelt and Ruud, using conventional Halton draws. The null hypothesis is that the mean utility of alternative B is zero.

Table 3 shows that using 100 draws, 55 per cent of the positive combinations of unrestricted and restricted log-likelihood values reject the null hypothesis at the 1 per cent significance level and in 91 per cent of the cases the null hypothesis is rejected at the 10 per cent level. Using 7,500 draws, the model is never rejected. Table 3 therefore shows that the problem decreases as the number of draws increases but even for 5,000 draws the null hypothesis will sometimes be accepted at the 10 per cent level, and other times rejected.

5. An example using real data

The problem described above has also been experienced on real data. The example below is based on 10,971 observations from 848 individuals, choosing between four different alternatives.[11] The utility of the non-base alternatives is assumed to follow a tree-dimensional normal distribution with correlation. In this example the true values of the means and variance-covariance of the mixing distribution are not known, and the model is therefore *optimized* using 52 different sets of starting values.

---

[11] The data is used to estimate willingness to pay for eggs with different levels of animal welfare, see Andersen (2011), where antithetic draws are used in the estimation.

The optimized log-likelihood values of 52 different sets of starting values have been sorted into quadrants by the sign of the estimated Choleski factorization, and Table 4 shows the optimized values of the log-likelihood function in the eight different quadrants, along with the probability of finding a maximum in each quadrant. The estimated log-likelihood values differ significantly between the eight quadrants, but except for one value in Q2 they are identical within quadrants for a given number of draws (not shown). Comparing the quadrants of the starting values and of the optimized results shows that there is apparently no connection between the quadrant of the starting point and quadrant of the final result (not shown), so the quadrant of the estimation results cannot be influenced a priori.

Table 4 Maximum simulated log-likelihood values by quadrant. Real data, estimated optima

	Number of draws per individual:							
	1,000		**1,500**		**5,000**		**10,000**	
	Prob. of quadrant*	**log-likelihood**[**]	**Prob. of quadrant**	**log-likelihood**	**Prob. of quadrant**	**log-likelihood**	**Prob. of quadrant**	**log-likelihood**
Q1	19%	10.69	23%	7.83	17%	11.88	15%	10.23
Q2	8%	0.00[***]	12%	8.64	8%	11.98	15%	12.58
Q3	25%	6.64	6%	11.27	17%	11.79	13%	11.45
Q4	19%	8.72	8%	11.32	23%	10.55	21%	11.19
Q5	15%	5.24	19%	11.90	12%	12.80	13%	11.64
Q5	6%	7.05	15%	13.05	6%	12.03	6%	10.99
Q7	4%	4.29	8%	13.76	10%	12.24	13%	12.51
Q8	4%	10.88	10%	13.85	8%	11.37	2%	11.43
Largest difference		*10.88*		*6.02*		*2.26*		*2.35*

* This is the probability that the estimated Choleski factorization lies in this quadrant. The estimation is performed 52 times with different sets of starting values.

** Only the difference between likelihoods is interesting, and the lowest estimated log-likelihood value (-8,388.20, from the estimation with 1,000 draws in Q2) is therefore subtracted from all the estimated values, in order to make the table easier to read.

*** One of the four results in this quadrant differs from the others by 0.274.

Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants with correlation. Real data, unbalanced panel, 848 individuals, 10,971 observations. Estimations conducted in the MMNL GAUSS program developed by Train, Revelt and Ruud, using conventional Halton draws.

It is clear from Table 4 that the difference between the quadrants is smaller for 5,000 draws than for 1,500 draws. The problem thus decreases with the number of draws, but even with 10,000 draws (which is in most cases too time consuming) the problem is still present, as the maximized log-likelihood still varies by 2.35 between quadrants, a difference which is large enough to invalidate Likelihood Ratio tests.[12] The problem of unreliable Likelihood Ratio tests

---

[12] As mentioned in footnote 6, a Likelihood Ratio test with one degree of freedom leads to rejection of the null hypothesis if the difference between the unrestricted and the restricted log-likelihood is 1.92 or higher.

which was illustrated using simulated data in the previous section therefore also appears on real data. In this example, the variation in likelihood values is sufficiently large to disturb Likelihood Ratio tests, even for high numbers of draws. It is important to note that whereas the log-likelihood values vary problematically much between the quadrants of the Choleski factorization, the resulting variance-covariance matrices are identical between quadrants, so the problem only affects Likelihood Ratio tests, not the validity of the estimated parameters.

6. Antithetic Halton draws

If the model includes more than one mixed parameter, symmetry in one dimension is not enough. If the number of mixed parameters is *n* – and if perfect symmetry is the goal – for each point in a given quadrant a corresponding point must be present in all of the other 2^n-1 quadrants. The problem is solved by creating antithetic Halton draws. As in Train (2003), the draws are created so that each point is 'mirrored' into the 2^n-1 other dimensions. We call these draws fully antithetic because they are antithetic in all dimensions.

For a case with three mixed parameters a Halton draw $d_1 = \begin{bmatrix} d_1^1 & d_1^2 & d_1^3 \end{bmatrix}$ (each between zero and one) is drawn, and then paired with 7 mirrors in the following way:

$$\begin{bmatrix} d_{11} \\ d_{12} \\ d_{13} \\ d_{14} \\ d_{15} \\ d_{16} \\ d_{17} \\ d_{18} \end{bmatrix} = \begin{bmatrix} d_1^1 & d_1^2 & d_1^3 \\ 1-d_1^1 & d_1^2 & d_1^3 \\ d_1^1 & 1-d_1^2 & d_1^3 \\ d_1^1 & d_1^2 & 1-d_1^3 \\ 1-d_1^1 & 1-d_1^2 & d_1^3 \\ 1-d_1^1 & d_1^2 & 1-d_1^3 \\ d_1^1 & 1-d_1^2 & 1-d_1^3 \\ 1-d_1^1 & 1-d_1^2 & 1-d_1^3 \end{bmatrix} \tag{5}$$

The Haltons must be symmetric for each individual in the panel, and it is therefore important that each 'set' of symmetric draws is assigned to one individual only, and not distributed over different individuals. The number of draws per individual in a model with *n*-dimensional mixing must therefore be a multiple of 2^n. In the case of 1,500 draws and three mixings this means that the number of draws must be e.g. $63 \cdot 2^3 = 1,504$ instead of 1,500 to ensure symmetry. Antithetic draws always have perfect symmetry, and therefore always skewness coefficient equal to zero (see e.g. Greene, 1997, for a definition of the skewness coefficient).

Capelari and Jenkins (2006) provide Stata programs for calculating multivariate normal probabilities by simulation, and allow for antithetic Halton draws. However, the antithetic draws are only one dimensionally antithetic, ignoring the 2^{n-1} other dimensions. This means that these draws will still allow for 2^{n-1} different local maxima, and that this type of one dimensionally antithetic draws therefore do not solve the problem of unreliable Likelihood Ratio tests. The following section shows that the fully antithetic draws described above solve the problem of simulation driven local maxima, and therefore provides reliable Likelihood Ratio tests.

7. Results of antithetic Halton draws

When the simulated likelihood function for the simulated data is calculated using antithetic Halton draws, the difference between the log-likelihood values from different quadrants is always zero as desired, and the Likelihood Ratio test of the null hypothesis therefore no longer varies. However, the result still changes as the number of draws increases. Table 5 presents the p-values for the simulated data which were also presented in Table 2 above, combined with the results of the antithetic draws.

Table 5 Testing away one mean using conventional or antithetic Halton draws, simulated data

	Number of draws per individual:					
	100	500	1,000	1,500	5,000	7,500
Conventional Halton draws (as in Table 2)						
Lowest p-value	.00	.00	.03	.05	.08	.12
Highest p-value	.64	.66	.77	.65	.27	.20
Antithetic Halton draws:						
Lowest and highest p-value	.01	.08	.11	.09	.15	.15

Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants. Simulated data, 1,000 individuals, 20 observations per individual. Data is defined in Appendix A. Calculations conducted in the MMNL GAUSS program developed by Train, Revelt and Ruud, using standard or antithetic Halton draws. The null hypothesis is that the mean utility of alternative B is zero.

Table 6 compares the differences in log-likelihood values on real data presented in Table 4 with the results of antithetic draws, and clearly demonstrates the effect of the antithetic Halton draws.

Table 6 Difference in simulated log-likelihood between quadrants, by number of draws, real data

	Number of draws per individual			
	1,000	**1,500**	**5,000**	**10,000**
Conventional Haltons (as in Table 4)				
Highest absolute difference in simulated log-likelihood	10.88	6.02	2.26	2.35
Antithetic Haltons				
Highest absolute difference in simulated log-likelihood	0.000196			

Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants with correlation. The estimation is performed 52 times with different sets of starting values. Real data, unbalanced panel, 848 individuals, 10,971 observations. Estimations conducted in the MMNL GAUSS program developed by Train, Revelt and Ruud, using either conventional Halton draws or antithetic Halton draws.

The precision of the optimization is set to 10^{-4}, and the highest difference between two results using antithetic draws is now lower than twice this level, and thereby completely acceptable.[13] Differences of this magnitude will have absolutely no effect on Likelihood Ratio tests, and the antithetic Halton draws therefore solve the problem of instability in the simulated likelihood of the Mixed MultiNomial Logit. At least the part caused by lack of symmetry of the likelihood function.[14]

8. Testing away mixing dimensions

Even when the problem of symmetry is solved by using full dimensionally antithetic draws, the problem of comparing log-likelihood values of models with different dimensions still remains. In a model with two mixed parameters (β_1 and β_2) the Halton draws will be based on two primes, e.g. 2 and 3 (2 representing β_1 and 3 representing β_2). If one of the mixed parameters (e.g. β_1) is restricted to be fixed (standard deviation restricted to zero), the dimension of the log-likelihood function is decreased by one, and the Halton draws will be based on only one prime. The standard choice would be the first prime, i.e. 2, independent of

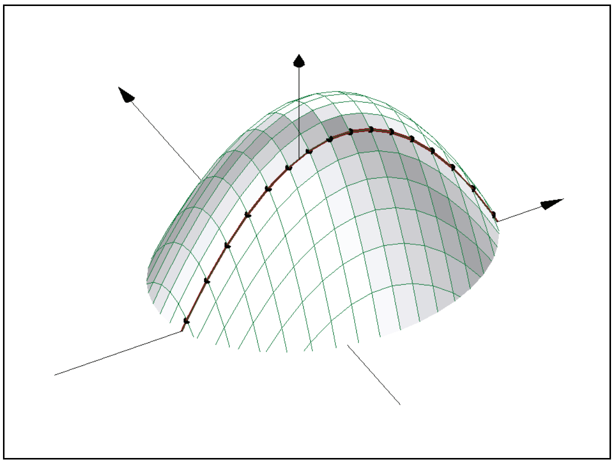
---

[13] The precision of the optimization indicates how close to zero the gradient of the log-likelihood function must be to be perceived as a maximum.

[14] Note, however, that local maxima may still occur if the model cannot be empirically identified by the data. The stability of the simulated log-likelihood should therefore still be investigated by estimations using different sets of starting values of the parameters, and standard estimation procedures should therefore also allow the user to control the starting values.

which dimension is restricted. Figure 2 illustrates the simulated conventional likelihood function which is to be integrated to form the likelihood function of the mixed logit. The heavy black line shows the likelihood function when one of the parameters is restricted to zero, and the dots on this line show the points in which the one-dimensional log-likelihood function would be evaluated for the given grid.

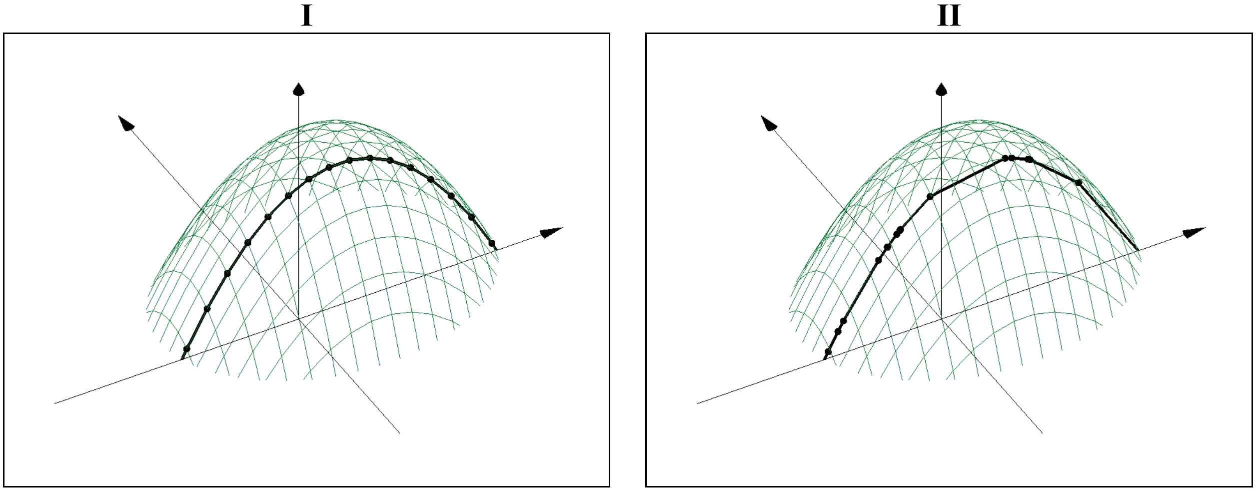
Figure 2 Simulated log-likelihood function in one and two dimensions



The figure describes a hypothetical log-likelihood function on a two dimensional parameter space. The heavy black line shows the likelihood function when one of the parameters is restricted to zero.

The symmetry of antithetic Haltons is needed to ensure that the log-likelihood functions of the different quadrants are identical, but as illustrated in Figure 3, the choice of prime may also matter. Figure 3 describes the same hypothetical log-likelihood function as Figure 2. The dots show the points in which the one-dimensional log-likelihood function would be evaluated for different draws. The dots in **I** illustrate a case where the one-dimensional draws correspond with the two-dimensional grid, and **II** illustrates a case where the one-dimensional draws are not part of the two-dimensional grid. The area under the one-dimensional likelihood function is clearly not the same in **I** and **II**.

Figure 3 Different one-dimensional likelihood functions given by different draws



The figure describes the same hypothetical log-likelihood function as in Figure 2. **I** illustrates a case where the one-dimensional draws correspond with the two-dimensional grid, **II** illustrates a case where the one-dimensional draws are not part of the two-dimensional grid.

To investigate the size of the problem we return to the three dimensional mixing on the simulated data used above (1,000 individuals and 20 observations per individual, defined in Appendix A). The restriction is now placed on the standard deviation of the utility of alternative C instead of the mean utility of alternative B, which was restricted in the mean-restriction case above. The utility of alternative C has a mean of 0.9981 and a standard deviation of 0.0984. The restricted model in this example assumes that the standard deviation is zero, but places no bounds on the mean. This means that the restricted model does not assume that the utility of alternative C is the same as the utility of the base alternative A. Table 7 shows the results of evaluating the log-likelihood function in the true parameters of the restricted model, using different primes for the antithetic Haltons.

Table 7 Log-likelihood values in the optimum of the restricted model using antithetic Haltonsfe

	Number of draws per individual:					
	100	**500**	**1,000**	**1,500**	**5,000**	**7,500**
Antith. Haltons based on 2 and 3	-18,466	-18,413	-18,412	-18,410	-18,409	-18,408
Antith. Haltons based on 2 and 5	-18,475	-18,415	-18,410	-18,411	-18,409	-18,408
Difference between results based on different primes	8.37	2.98	-1.64	0.92	-0.70	-0.85

Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants. Simulated data, 1,000 individuals, 20 observations per individual. Data is defined in Appendix. Calculations conducted in the MMNL GAUSS programme developed by Train, Revelt and Ruud, using conventional Halton draws. The restricted model assumes that the standard deviation of the utility of alternative C is zero.

The differences in Table 7 may seem small compared to the absolute values of the likelihood functions, but remember that when restricting a single parameter, a difference of 1.92

between the restricted and the unrestricted likelihood will lead to acceptance of the restricted model at the five percent level. The difference between likelihoods based on different primes is therefore substantial, and for Likelihood Ratio tests which are 'close' to being accepted it will be important to keep track of the relationships between primes and mixing dimensions.

9. Conclusion

This paper shows that Likelihood Ratio test in mixed models estimated by standard methods cannot always be trusted, and presents a solution to this problem. When a model with *n*-dimensional mixing is estimated by Quasi-Monte Carlo integration using asymmetric draws, the estimated variance-covariance matrix may be obtained from 2^n different Choleski factorizations, located in different quadrants of the 2^n dimensional parameter space, and these different factorizations may lead to different values of the optimized log-likelihood, even though they all lead to the same variance-covariance matrix. The paper shows that if the solution to an unrestricted and a restricted model is found in different quadrants, the Likelihood Ratio test is not reliable, and the paper also demonstrates that using fully antithetic draws eliminates this problem. Antithetic draws are therefore strongly recommended for future use in standard estimation procedures for mixed models.

The paper also illustrates that even when fully antithetic draws are used, testing restrictions on the number of mixed parameters may lead to false inference if the relationship between primes and mixed parameters are not maintained in the restricted model.

Some estimation procedures allow for one-dimensionally antithetic draws, but as explained above, one-dimensionally antithetic draws only removes one dimension of the problem, leaving 2^{n-1} different optima which still may lead to false Likelihood Ratio tests. In the three dimensional mixings used in this paper, the dimension of the full set of antithetic draws is only eight and it is therefore possible to use fully antithetic draws, but in cases of higher dimensional mixings this may not be possible. Sidharthan and Srinivasan (2010) suggest drawing a limited number of antithetic draws randomly from the full set of antithetic draws. They conclude that 5 antithetic draws are enough even in the 30 dimensional case where the true number of antithetic draws is 2^{30} which is well above one billion. Their conclusion is based on evaluations of time reductions and precision of parameter estimates, and might be different if the evaluation criterion was the reliability of Likelihood Ratio tests. This could be an interesting route of further research.

In many applications the dimension of the mixing distribution will be small enough to use fully antithetic draws, and in these cases fully antithetic draws will ensure reliable Likelihood Ratio tests. Fully antithetic draws are therefore recommended for ‘small’-dimensional mixings.

References:

- Andersen, L. M. (2008): Antithetic Halton draws, in *Information Provision to Consumers as an Instrument of Environmental Regulation*, PhD Series No. 130 - 2008, Department of Economics, University of Copenhagen.
- Andersen, L. M. (2011): Animal Welfare and Eggs – Cheap Talk or Money on the Counter?, *Journal of Agricultural Economics*, **62**: 565–584.
- Baht, C. R. (2001): Quasi-random maximum simulated likelihood estimation of the Mixed MultiNomial Logit model. *Transportation Research Part B*, **35**: 677-693.
- Ben-Akiva, M., D. Bolduc and M. Bradley (1993): Estimation of travel choice models with randomly distributed values of time. *Transportation Research Record*, **1413**: 88-97.
- Ben-Akiva, M. and D. Bolduc (1996): Multinomial probit with a logit kernel and a general parametric specification of the covariance structure. *Working paper, Department of Civil Engineering, Massachusetts Institute of Technology.*
- Berry, S., J. Levinsohn and A. Pakes (1995): Automobile Prices in Market Equilibrium. *Econometrica*, **63** (4): 841-890.
- Bhat, C. (1996): Accommodating variations in responsiveness to level-of-service measures in travel model choice modelling. *Working paper, Department of Civil Engineering, University of Massachusetts at Amherst.*
- Brownstone, D. and K. Train (1999): Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics*, **89:** 109-129.
- Cappellari, L. and S. P. Jenkins (2006): Calculation of Multivariate Normal Probabilities By Simulation, With Applications to Maximum Simulated Likelihood Estimation. *IZA Discussion Paper No. 2112*. Available at SSRN: http://ssrn.com/abstract=900829 (accessed February 2012).
- Chang, J. B. and J. L. Lusk (2011): Mixed logit models: Accuracy and software choice. Software review in Journal of Applied Econometrics, **26**: 167–172.
- Chiou and Walker (2007): Masking identification of discrete choice models under simulation methods. *Journal of Econometrics*, 141: 683–703.
- Geweke, John (1988): Antithetic Acceleration of Monte Carlo Integration in Bayesian Inference. *Journal of Econometrics* **38:** 73-89.
- Geweke, John (1996): Monte Carlo Simulation and Numerical Integration, in *Handbook of Computational Economics*, Volume L Edited by H.M. Amman, D.A. Kendrick and J. Rust.
- Greene, W. H. (1997): *Econometric analysis*. Third edition. Prentice Hall, Upper Saddle River, New Jersey, USA.
- Halton, J. H. (1960): On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, **2**: 84-90.
- Hammersley, J. M. (1960): Monte Carlo methods for solving multivariable problems. *Annals of the New York Academy of Sciences*, **86**: 844-874.

- Hensher, D. A. and W. H. Greene (2003): The Mixed Logit model: The state of practice, *Transportation,* **30**: 133–176.
- McFadden, D. (1973): Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (ed.): *Frontiers in econometrics*. Academic Press, New York. USA, 105-142.
- McFadden, D. and K. Train (2000): Mixed MNL models for discrete response. *Journal of Applied Econometrics*, **15**: 447-470.
- Morokoff, W. J. and R. E. Caflisch (1995): Quasi-Monte Carlo Integration. *Journal of Computational Physics*, **122**: 218-230.
- Revelt, D. and K. Train (1998): Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of Economics and Statistics*, **80**: 647-657.
- Sidharthan, R. and K. K. Srinivasan (2010): Random Coefficient Mixed Logit Models Based on Generalized Antithetic Halton Draws and Double Base Shuffling. Transportation Research Record: Journal of the Transportation Research Board, **2175**: 1–9.
- Train, K. (1999): Halton sequences for mixed logit. *Working paper*, Department of Economics, University of California, Berkeley.
- Train, K. (2003): *Discrete Choice Methods with Simulation*. Cambridge University Press.
- Walker (2001): *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables*. Ph.D. dissertation, Massachusetts Institute of Technology.
- Walker, J. L., M. Ben-Akiva and D. Bolduc (2007): Identification of Parameters in Normal Error Component Logit-Mixture (NECLM) Models. *Journal of Applied Econometrics*, **22**: 1095-125.

Appendix A: Simulated data

The number of individuals is 1,000 and each individual makes 20 choices. The number of alternatives is 4.

The simulated utility of the four alternatives is defined as:

$$mean\begin{bmatrix} u_A \\ u_B \\ u_C \\ u_D \end{bmatrix} = \begin{bmatrix} 0 \\ 0.1 \\ 1 \\ 2 \end{bmatrix}, \quad std.dev\begin{bmatrix} u_A \\ u_B \\ u_C \\ u_D \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0.1 \\ 3 \end{bmatrix}, \quad corr\begin{bmatrix} u_A \\ u_B \\ u_C \\ u_D \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ & 1 & 0 & 0 \\ & & 1 & 0 \\ & & & 1 \end{bmatrix} \tag{6}$$

which means that the simulated covariance is defined as:

$$\text{cov}\begin{bmatrix} u_A \\ u_B \\ u_C \\ u_D \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ & 1 & 0 & 0 \\ & & 0.01 & 0 \\ & & & 9 \end{bmatrix} \tag{7}$$

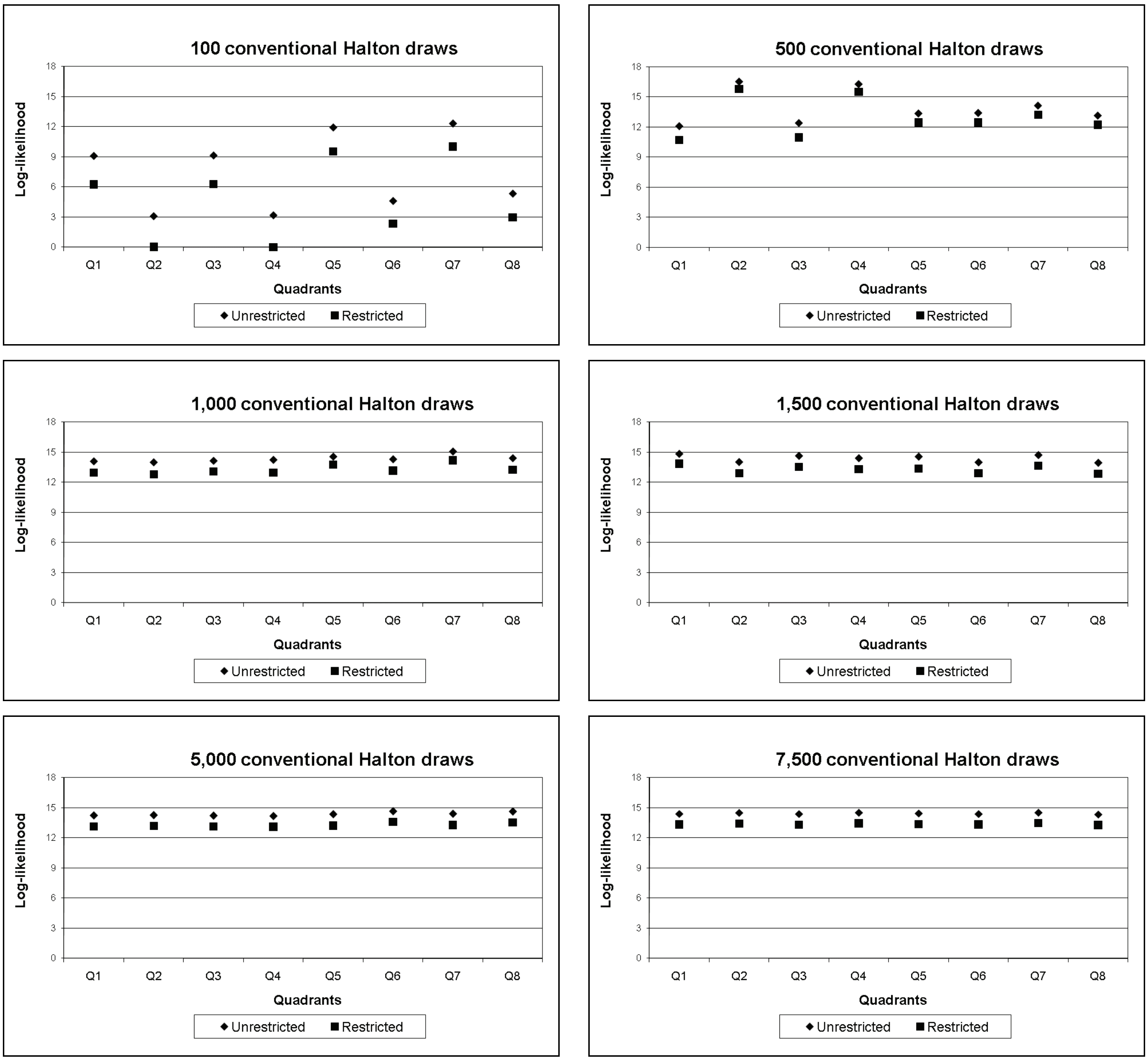
Alternative A is used as the base. The mean and the standard error of the base alternative is zero and the realization of the utility is:

$$mean\begin{bmatrix} u_B \\ u_C \\ u_D \end{bmatrix} = \begin{bmatrix} 0.0834 \\ 0.9981 \\ 2.0682 \end{bmatrix}, \quad \text{cov}\begin{bmatrix} u_B \\ u_C \\ u_D \end{bmatrix} = \begin{bmatrix} 0.9868 & 0.0021 & -0.0633 \\ & 0.0097 & -0.0004 \\ & & 9.2070 \end{bmatrix} \tag{8}$$

The values in (6) are used as the *true* parameter values in the paper.

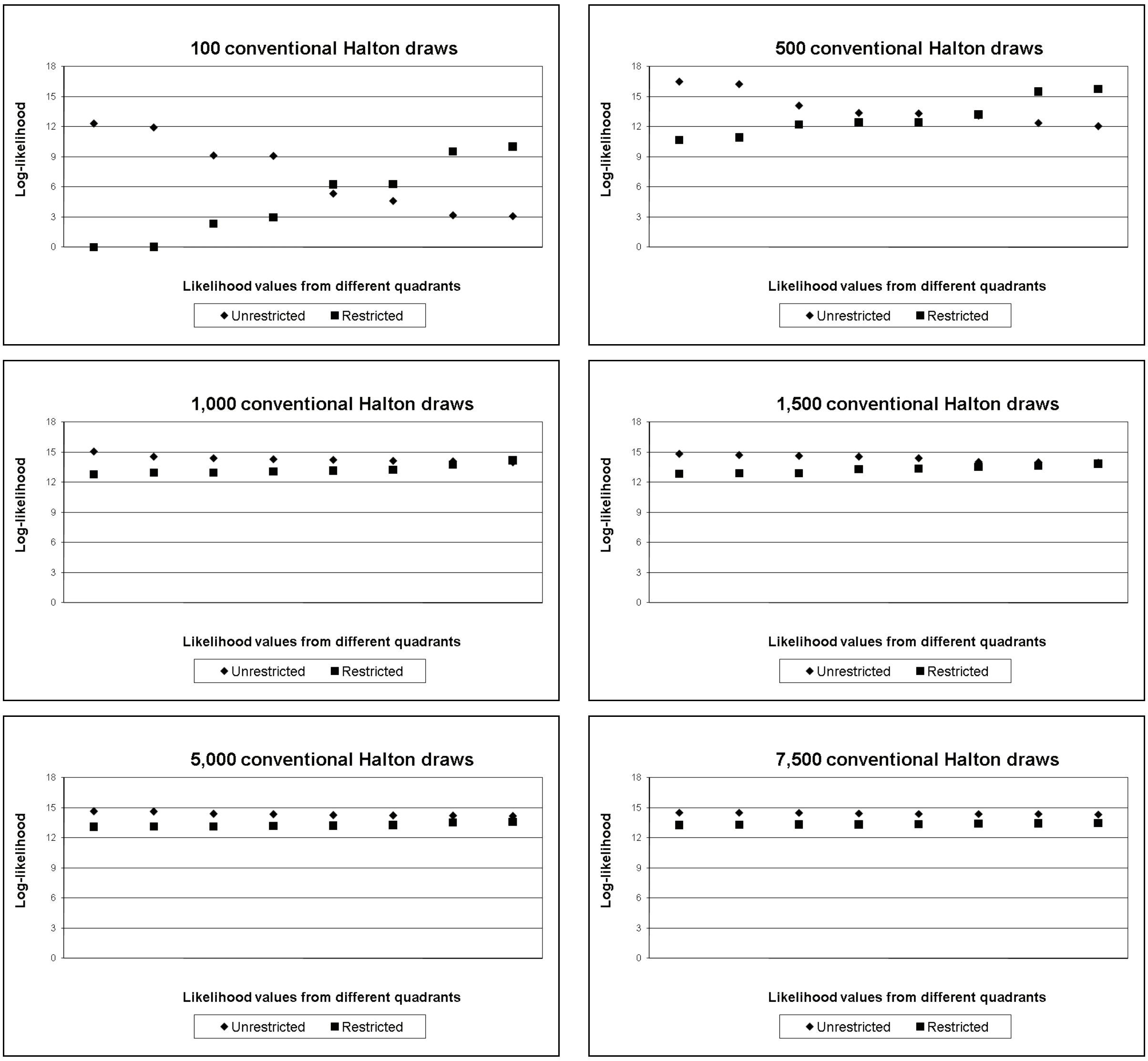
Appendix B: Illustration of differences in log-likelihood values in optimum from different quadrants

Table 8 Log-likelihood function evaluated in the true parameter values of the simulated data, by quadrant



Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants. Only the difference between likelihoods is interesting, and the lowest estimated log-likelihood value (from the restricted model in Q4, using 100 draws) is therefore subtracted from all the estimated values. Simulated data, 1,000 individuals, 20 observations per individual. Data is defined in Appendix A. Calculations conducted in the MMNL GAUSS program developed by Train, Revelt and Ruud, using conventional Halton draws. The null hypothesis is that the mean utility of alternative B is zero.

Table 9 Log-likelihood function evaluated in the true parameter values of the simulated data, by size



Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants. Only the difference between likelihoods is interesting, and the lowest estimated log-likelihood value (from the restricted model in Q4, using 100 draws) is therefore subtracted from all the estimated values. Simulated data, 1,000 individuals, 20 observations per individual. Data is defined in Appendix A. Calculations conducted in the MMNL GAUSS program developed by Train, Revelt and Ruud, using conventional Halton draws. The null hypothesis is that the mean utility of alternative B is zero.