On the use of a Modified Latin Hypercube Sampling (MLHS) Method in the Estimation of a Mixed Logit model for vehicle choice

Stephane Hess
(corresponding author)
Centre for Transport Studies
Imperial College London
London SW7 2AZ
United Kingdom
Tel: +44(0)20 7594-6105
Fax: +44(0)20 7594-6102
E-mail: stephane.hess@imperial.ac.uk

Kenneth E. Train
Department of Economics
University of California
549 Evans Hall # 3880
Berkeley CA 94720-3880
Tel: 415-291-1023
Fax: 415-291-1020
E-mail: train@econ.berkeley.edu

John W. Polak
Centre for Transport Studies
Imperial College London
London SW7 2AZ
United Kingdom
Tel: +44(0)20 7594-6089
Fax: +44(0)20 7594-6102
E-mail: j.polak@imperial.ac.uk

ABSTRACT

Quasi-random number sequences have been used extensively for many years in the simulation of integrals that do not have a closed-form expression, such as Mixed Logit and Multinomial Probit choice probabilities. Halton sequences are one example of such quasi-random number sequences, and various types of Halton sequences, including standard, scrambled, and shuffled versions, have been proposed and tested in the context of travel demand modeling. In this paper, we propose an alternative to Halton sequences, based on an adapted version of Latin Hypercube Sampling. These alternative sequences, like scrambled and shuffled Halton sequences, avoid the undesirable correlation patterns that arise in standard Halton sequences. However, they are easier to create than scrambled or shuffled Halton sequences. They also provide more uniform coverage in each dimension than any of the Halton sequences. A detailed analysis, using a sixteen-dimensional Mixed Logit model for choice between alternative-fuelled vehicles in California, was conducted to compare the performance of the different types of draws. The analysis shows that, in this application, the Modified Latin Hypercube Sampling (MLHS) outperforms each type of Halton sequence. This greater accuracy combined with the greater simplicity make the MLHS method an appealing approach for simulation of travel demand models and simulation-based models in general.

INTRODUCTION

The analysis and modeling of travelers' choices is one of the most important areas of transportation research. Predominantly, the models used in this area of research are members of the family of discrete choice models. This family can be split into two main groups of models; models with a (convenient) closed-form expression, and models that require simulation to approximate the value of the integrals that they are based on. For a detailed review of existing model types, along with in-depth descriptions of the different approaches to simulation, see Train (2003).

While the simulation-based methods allow for more realistic modeling of the actual behavior of decision-makers, notably by allowing for random taste variation in the population, their use was initially very limited, due to the high computational cost of the simulation processes required during estimation and application. Over recent years however, significant advancements in the performance of computers have led to renewed interest in simulation-based models, and have for example led to the Mixed Logit model becoming one the most widely used models in transportation research.

Besides the advancements in computer technology, improvements have also been made to the simulation process, notably in the use of alternatives to pseudo-random numbers. The use of these "cleverly-crafted" sequences, known as *quasi-random* number sequences, can lead to significant improvements in the precision of simulation processes, thanks to the improved quality (uniformity) of coverage of the multi-dimensional area of integration. This in turn leads to a requirement for fewer draws to be used in the simulation process, considerably easing the computational cost of simulation. Over the years, many different types of such *quasi-random* number sequences have been proposed, one of the most popular being Halton sequences. In their

original form, Halton sequences were found to greatly outperform pseudo-random numbers in the estimation of choice models, at least when the number of dimensions is small (Bhat, 1999, Train, 1999). However, untransformed Halton sequences create undesirable correlations over dimensions, which can be particularly problematic for higher-dimensional integration. Two main methods of transforming Halton sequences have been proposed to eliminate the multi-dimensional correlation. Scrambled Halton draws permute the digits on the original elements of a multidimensional sequence, while shuffled Halton sequences permute the order of the original elements.

In this paper, we propose an alternative to Halton sequences, based on the use of one-dimensional sequences of evenly spaced draws that are randomly shuffled before being combined into multi-dimensional sequences. These sequences obtain more uniform coverage in each dimension than Halton sequences of any form, since even spacing gives the most uniform coverage possible. The shuffling of the one-dimensional sequences does not affect the spacing in each dimension and yet avoids correlation over dimensions in the same way as shuffled Halton sequences. Also, these new sequences are easier to create than the scrambled and shuffled Halton sequences. To analyze the performance of this new type of sequence, we fit a Mixed Logit model to a data set for choice between gas, electric and hybrid vehicles, and compare the ability of the shuffled uniform sequences to recover the "true" parameter values to that of the different types of Halton sequences.

The remainder of this paper is organized as follows. In the following two sections, we give a brief review of Monte-Carlo and Quasi-Monte Carlo integration. The development of our proposed approach is described in the fourth section, the data and models used in the estimation

are described in the fifth section, and the sixth section compares the performance of the different types of approaches in terms of their ability to recover the "true" values of the parameters.

MONTE-CARLO INTEGRATION AND MAXIMUM SIMULATED LIKELIHOOD ESTIMATION

Equation [1] shows a very general form of integral that is used in different types of discrete choice models, as well as other areas of research.

$$\int_x g(x) f(x) dx \qquad \ldots[1]$$

The integral is over the range of x, with density function for x given by $f(x)$, and the dimensionality of the integral depends on the dimensionality of x. The definition of $f(x)$ implies that equation [1] is in fact equal to the expected value of $g(x)$, say $\bar{g}$. The integral in equation [1] can easily be approximated by using the following approach:

1. Draw a value x^r from the density function $f(x)$, where r indicates the current iteration, with $r=1$ in the first iteration. The draws are based on "uniform" draws in the 0-1 interval, which are then transformed in order to represent draws for a given choice of $f(x)$; for details, see Train (2003).
2. Calculate $g(x^r)$, the value of the integrand in the current iteration.
3. Repeat steps 1 & 2 a high number of times, then average the results.

With R repetitions (draws), the integral is thus approximated by

$$\breve{g} = \frac{\sum_{r=1}^{R} g\left(x^r\right)}{R}, \qquad \ldots[2]$$

where, with independent draws $\{x^1,\ldots,x^R\}$, the estimator is unbiased by construction. Furthermore, the variance of the estimator decreases as R increases, such that the simulation error decreases as R increases.

Integrals of the form shown in equation [1] are used in Mixed Multinomial Logit models (MMNL) to calculate the choice probabilities for the different alternatives. We have:

$$P_{ni} = \int_{\varepsilon} g_i(\varepsilon, z_n) f(\varepsilon;\beta) d\varepsilon \,, \qquad \ldots[3]$$

where P_{ni} is the probability of decision-maker *n* choosing alternative *i* and z_n is a matrix of the (fixed) attributes of the different alternatives as faced by decision-maker *n*. The vector ε varies over decision-makers and reflects the idiosyncratic aspects of decision-maker *n*'s preferences; these terms are distributed in the population with density $f(\varepsilon;\beta)$, where β is a vector of parameters to be estimated that represents, for example, the mean and variance of preferences in the population. The function $g_i(\varepsilon, z_n)$, which is a Multinomial Logit (MNL) probability term (for alternative *i*, as faced by decision-maker *n*), is integrated over the density of random terms to obtain the MMNL choice probability. For a given value of the parameter vector β, the choice probabilities for each alternative in the choice set are calculated for each decision-maker; with *I* alternatives, the log-likelihood is then given by:

$$LL(\beta) = \sum_{n=1}^{N} \sum_{i=1}^{I} d_{ni} \ln\left(P_{ni}\right), \qquad \ldots[4]$$

where d_{ni} is a dummy variable that is set to *1* if decision-maker *n* is observed to choose alternative *i* and *0* otherwise. In estimation packages, equation [4] is simply replaced by the sum of the logged probabilities of the chosen alternatives, requiring the calculation of only *N* choice probabilities, as opposed to *N·I*, as in equation [4].

The standard approach to finding estimates for β is to use maximum likelihood estimation, such that with maximum likelihood estimator (MLE) $\hat{\beta}$, we have that

$$\frac{dLL\left(\hat{\beta}\right)}{d\hat{\beta}} = 0 \,. \qquad \ldots[5]$$

In the presence of integrals that do not have a closed-form expression (as in the MMNL model), the choice probabilities P_{ni} are replaced by the simulated choice probabilities $\breve{P}_{ni}$ in equation [4], leading to the simulated log-likelihood function. The simulated choice probabilities are obtained by:

$$\breve{P}_{ni} = \frac{\sum_{r=1}^{R} g_i\left(\varepsilon^r, z_n\right)}{R}, \quad \ldots[6]$$

where the different values of ε^r are independent draws from $f(\varepsilon;\beta)$, for a given value of β. The simulated log-likelihood is then given by:

$$SLL(\beta) = \sum_{n=1}^{N}\sum_{i=1}^{I} d_{ni} \ln\left(\breve{P}_{ni}\right), \quad \ldots[7]$$

with maximum simulated likelihood estimator (MSLE) given by $\breve{\beta}$.

It has been shown (Lee, 1995) that if R rises faster than the square root of the number of observations ($\sqrt{N}$), MSL estimation is asymptotically equivalent to ML estimation. The use of a fixed number of draws R induces simulation bias and variance, which is lower with higher R (c.f. Train, 2003).

ALTERNATIVES TO SIMPLE PSEUDO-RANDOM NUMBER SEQUENCES

The problem with using short sequences of pseudo-random numbers (also called pseudo-Monte Carlo, or PMC, sequences) is the uneven coverage of the area of integration that can arise by the randomness in the sequences. This can lead to greater simulation error in the choice probabilities, leading to poorer estimation of the parameter vector β.

Quasi-random number (or quasi Monte Carlo, QMC) sequences are designed to provide more uniform coverage of the area of integration. The use of these QMC sequences can lead to significant improvements in the precision of the simulated probabilities and MSL estimation,

hence leading to lower requirements in the number of draws used, with corresponding reductions in the computational cost of model estimation and application.. The only QMC sequence that has seen widespread use in Mixed Logit modeling in general, and transportation research in particular, is the Halton sequence (Halton, 1960). In the following subsections, we present a brief description of the Halton sequence, in its different forms, and discuss some of the findings regarding its use, especially in the area of transportation research. We then briefly discuss some alternative types of sequence that could be used in the estimation of MMNL models.

Halton sequences

Halton sequences are constructed according to a deterministic method that uses a prime number as its base. The one-dimensional Halton sequence based on prime p (≥ 2) fills the 0-1 space by dividing this space into p segments, and by systematically filling in the empty spaces, using cycles of length p that place one draw in each segment.

Formally, the i^{th} element in the Halton sequence based on prime p is obtained by taking the radical inverse of integer i in base p by reflection through the radical point. We have:

$$i = \sum_{l=0}^{L} b_l(i) p^l \,, \qquad \ldots[8]$$

with $0 \leq b_l(i) \leq p-1$ and $p^L \leq i < p^{L+1}$, and use the values for $b_0(i), \cdots, b_L(i)$ that solve equation [8], in writing the resulting Halton element in base p as:

$$\varphi_p(i) = 0.b_0(i)b_1(i)..b_L(i) \qquad \ldots[9]$$

This can be rewritten as a decimal number by:

$$\varphi_p(i) = \sum_{l=0}^{L} b_l(i) p^{-l-1} \qquad \ldots[10]$$

The deterministic Halton sequence can be randomized in several ways. Tuffin (1996) and Bhat (1999) suggest random shifting, which is implemented as follows: (1) take a draw from a

uniform density between 0 and 1, (2) add this draw to each element of the Halton sequence, and (3) substract 1 from any resulting element that exceeds one. The process shifts the entire sequence over by the amount of the random draw, with the elements that are shifted out of the unit line "wrapping back" to re-enter the unit line from zero. We use this form of randomization in our application. Another procedure, suggested by Wang and Hickernell (2000), is to eliminate the first G elements of the sequence, where G is chosen randomly. Multi-dimensional Halton sequences are constructed by combining one-dimensional sequences generated from different primes. Randomization of the one-dimensional sequences provides randomization of the multi-dimensional sequence.

A well-known problem with multi-dimensional Halton sequences is that, especially when using high primes, the individual Halton sequences can be highly correlated, leading to an uneven distribution of draws in the multi-dimensional area of integration (poor coverage), and consequently poor simulation and estimation performance. While most readily recognized with high primes, these problems actually arise to some degree whenever the ratio of any two prime numbers used is close to an integer value. The way in which Halton sequences are constructed implies that the effects of such correlation are gravest in the case of high primes, where the number of cycles used is consequently lower. For sequences based on choices of primes that do not lead to substantial problems with correlation, standard Halton sequences can lead to significant improvements in estimation efficiency compared to pseudo-random number sequences, and have been used successfully in different areas of science, including transportation; see for example Bhat (1999), Train (1999), Bhat (2000) and Hess and Polak (2004).

Two main methods for reducing the correlation in Halton sequences have been discussed in the literature; scrambling and shuffling. The former approach scrambles the sequences by using permutations of the coefficients $b_l(i)$ in the radical inverse function (equations [9] and [10]). The resulting Scrambled Halton sequence for prime *r* is written as:

$$\varphi_{sp}(i) = \sum_{l=0}^{L} \sigma_p(b_l(i))\, p^{-l-1} \quad \ldots[11]$$

where σ_p is the operator of permutations for the possible values of $b_l(i)$ in base *p*. Different methods for producing the permutations have been proposed; the most commonly used method was proposed by Braaten and Weller (1979), other approaches are given by Hellekalek (1984), Kocis and Whiten (1997) and Tuffin (1998). In this paper, we use the approach proposed by Braaten and Weller (1979) and implemented in transportation by Bhat (2002).

The Shuffled Halton sequence was introduced by Morokoff and Caflisch (1994). The procedure consists of randomly reordering the elements of each one-dimensional sequence before combining them into a multi-dimensional sequence. The generation of a Shuffled Halton sequence uses a one-dimensional standard Halton sequence of length *N*, generated from prime p:

$$H_p = \langle \varphi_p(1), \ldots, \varphi_p(N) \rangle \quad \ldots[12]$$

where $\varphi_p(i),\ i = 1, \ldots, N$ is defined as in equation [10]. Differently shuffled versions of this sequence are then used in different runs, with the sequence used in the j^{th} run being given by:

$$\mathrm{H}_{p,\sigma(j)} = \mathbf{H}(\mathrm{I}_j, \mathrm{H}_p), \quad \ldots[13]$$

where $\mathbf{H}$ is a function that creates a sequence with the elements of H_p arranged in the order given in the vector *Ij*, where *Ij* is a random permutation of the ordered index vector, and where $\sigma(\cdot)$ is a permutation operator that yields the permutation *Ij* of the index vector *I* in the j^{th} run. The use of different permutations of the index vector for different dimensions disrupts the cyclical ordering

in the different dimensions in different ways and hence manages to reduce correlation between the individual sequences. Although the performance of the shuffled Halton sequence is clearly dependent on the specific shuffling order used in a given run, it has been shown that the shuffled Halton sequence does have the potential to significantly reduce the impact of correlation, and can offer very stable performance over runs in the estimation of MMNL models (Hess et al., 2003).

QMC Sequences Designed in Multiple Dimensions

As described above, multi-dimensional Halton sequences are created by combining one-dimensional sequences. Several QMC sequences have been proposed that are created in the multi-dimensional space directly. Generally, these sequences obtain more uniform spacing of points over the multiple dimensions while attaining less uniformity in each one dimension.

One such method is the systematic sampling approach proposed by McGrath (1970). This approach produces a *K*-dimensional uniform grid of *N* points by randomly drawing *N/M* points in one of the *M K*-dimensional parts of the grid and translating these points into the remaining *M-1* grid areas. Alternatively, the draws in each sub-area are drawn randomly. The precision of either approach increases with *M*. The main difficulty with these grid methods is that of finding appropriate values of *M* (and *N*), such that *N/M* and $\sqrt[K]{M}$ are integer values. Clearly, the best performance with this method is obtained in the case where a division can be used such that each sub-cube of the 0-1 cube contains exactly one draw. This is only possible in the rare case where $\sqrt[K]{N}$ is an integer value. Alternatively, it is possible to use a division of the 0-1 hypercube into hyper-rectangles, rather than hyper-cubes. As an example, with *N=100* in three dimensions, we can divide the 0-1 interval into five segments along two dimensions, and into four segments along the third dimension.

In any case, the uniformity in each one-dimension is generally less with the grid approach than under Halton sequences or the MLHS that we discuss below. For example, in a two-dimensional case with N=100, the grid can divide each dimension into 10 segments. Each segment in one dimension contains 10 points which are either exactly the same in that dimension (if one random draw is used for the entire grid) or are randomly placed within this segment (if different random draws are used for each sub-area of the grid). In contrast, MLHS divides the unit line into 100 segments with one point in each.

Another type of QMC sequences designed in multiple dimensions are *(t,m,s)*-nets, as discussed for example by Niederreiter (1992). These nets constitute a general class that includes Sobol, Faure, Niederreiter, Niederreiter-Xing, and other sequences. The specification and construction of *(t,m,s)*-nets are different for different numbers of dimensions and points. They are restrictive in the number of points that can be used for any given number of dimensions, and, like the grid procedures, generally attain more uniform multi-dimensional coverage at the expense of uniformity along each individual dimension[1]. Sándor and Train (2004) compare four kinds of (*t,m,s*)-nets with Halton sequences in the estimation of MMNL models. They find that two of the (*t,m,s*)-nets performed better than Halton sequences, and the other two performed worse.

MODIFIED LATIN HYPERCUBE SAMPLING

In this section, we describe our proposed Modified Latin Hypercube Sampling (MLHS) approach. The motivation for the approach arises from the concepts developed with respect to Halton sequences. In single dimensions, Halton sequences attain good coverage because they are

[1] An exception to this trade-off occurs for nets with *t=0*, which is difficult to attain in high dimensions.

constructed of sub-sequences of evenly spaced points, with each sub-sequence filling in spaces left unfilled by previous sub-sequences. Halton sequences were developed in the context of series, which are defined as being sequences whose length can be extended without changing the original points. This requirement allowed theoretical analysis of the asymptotic properties of the series, such as measuring the uniformity of coverage as the length of the series extends without bound. However, for any given length of the sequence, the points are not evenly spaced, and hence the one-dimensional coverage is not as uniform as possible. This is an effect of the sequential nature of these sequences, making their performance depend heavily on the exact number of draws used; this can explain the fact that increases in sequence length do not necessarily lead to constant improvements in simulation performance, as for example observed by Hess et al. (2003).

With a given number of points, the most uniform coverage possible is necessarily attained by spacing those points evenly. Such a sequence of evenly-spaced points cannot be extended without changing the original points and therefore does not qualify as a series in the way that a Halton sequence does. This could be seen as a disadvantage, since the researcher cannot add points without recalculating the original points. However, for the purposes of numerical simulation rather than theoretical analysis of asymptotic properties, some preset number of points is always used, and with a given number of points, the best coverage is attained by spacing them evenly. This constitutes the first aspect of our proposed approach.

The second aspect arises with respect to combining several one-dimensional sequences to create multi-dimensional sequences. When one-dimensional Halton sequences are combined into multi-dimensional vectors, the cycling over sub-sequences in each dimension creates correlation over dimensions, which scrambling and shuffling have been proposed to address. Combining

evenly spaced points in each dimension creates perfect correlation over dimensions. However, the same kind of shuffling that was proposed to eliminate correlation in the case of Halton sequences can be applied in the case of uniform sequences. The points in each dimension are re-ordered randomly (i.e. shuffled), before combining the one-dimensional sequences into a multi-dimensional sequence. As is the case with shuffled Halton draws, the shuffling of the uniform vectors does not change the coverage in any of the one-dimensional sequences. Each shuffled sequence of evenly spaced points therefore provides more uniform coverage in each dimension than any type of Halton sequence, since the coverage was more uniform before shuffling.

We now describe the sequence more formally. It can easily be seen that the most basic way of guaranteeing equal distances between draws is to set the j^{th} draw to be equal to $\frac{j}{N+1}$; the difference between any two adjacent draws is now equal to $\frac{1}{N+1}$, as is the difference between the extreme points and the border of the 0-1 interval. However, this approach is clearly impractical as it prohibits the use of different draws in different dimensions, as well as in different runs.

The problem thus lies in devising a method that leads to equal distances between adjacent draws without requiring the draws to be the same in different dimensions and runs. We propose an approach that starts with a sequence of draws defined by:

$$\varphi(j) = \frac{j-1}{N}, \quad j = 1, \cdots, N \qquad \ldots[14]$$

A random number x is then drawn, such that $0 < x < 1/N$; this can be obtained by drawing a pseudo-random number[2] contained in the interval $]0,1[$, thus using strict upper and lower limits, and dividing this number by N. The elements in the final sequence are then generated by:

$$\Psi(j) = \varphi(j) + x, \quad j = 1, \cdots, N. \qquad \ldots[15]$$

In the resulting sequence, the distances between adjacent draws are all equal to $1/N$, satisfying the condition of equal spacing. Furthermore, the combined distance between the two extreme points and the respective borders of the 0-1 interval is also equal to $1/N$; if the interval is regarded as being cyclical, the distance between the two extreme points is thus also $1/N$, so that all "adjacent" distances are equal to $1/N$. Multi-dimensional sequences are constructed by simple combination of randomly shuffled one-dimensional sequences.

The addition of the random variate x has the desired effect that the draws used are different across dimensions and runs, while still keeping the distance between adjacent draws at $1/N$, such that the quality of one-dimensional coverage remains unchanged. Two different approaches can be used in the shifting of the draws; the use of the same shifting (same value of x) in different runs of the approach will lead to the same one-dimensional draws but different multi-dimensional draws due to the shuffling (as with shuffled Halton draws), while the use of a different shifting in different runs will lead to different one-dimensional as well as multi-dimensional draws.

---

[2] The spread of the values of x across sequences can be made more uniform by basing the draws of x themselves either on a quasi-random approach (e.g. Halton draws), or by using a pseudo-random number generator with proven properties (of which the computational statistics literature contains many examples).

The procedure is very similar to the Latin Hypercube Sampling (LHS) proposed by McCay *et al*. (1979) – hence the name, MLHS. LHS is the same as the MLHS approach described above except that for LHS, a different draw is taken for each element in each dimension, rather than using the same draw for all elements in each dimension. That is, for LHS, equation 15 is changed to

$$\Psi(j) = \varphi(j) + x_j, \quad j = 1, \cdots, N$$

where x_j is a separate draw from a uniform between 0 and 1/N. The use of NK different uniform draws as opposed to K uniform draws makes the LHS method computationally more expensive, although, given the low cost of producing random draws, these differences in speed are trivial. The more important differences arise in the actual substantive effects of using a common draw for each point in a given dimension, as opposed to using separate draws. The motivation for MLHS was to attain a sequence that has more uniform coverage in each dimension than Halton sequences. An evenly spaced set of points necessarily attains that goal and is the most uniform set possible. Using one draw to shift all points retains this uniformity. In contrast, when using a separate draw for each point, as in LHS, the resulting set of points is less uniform than evenly spaced points and can potentially (depending on the realized draws) be less uniform than the Halton sequence. It should be noted however that the LHS method allows for the possibility of a type of cancelling out, by which a draw that moves one point to the higher end of its segment is counteracted by a draw that moves another point to the lower end of its segment. Inherently, however, the goal of uniformity operates against the potential for cancelling out in this way.

It is of interest to briefly compare our MLHS approach to the systematic sampling method described in the previous section, hence highlighting the differences between a sequence designed in multiple dimensions and a sequence based on the combination of one-dimensional

sequences. Indeed, it should be highlighted again that, with a method based on the combination of randomly shuffled one-dimensional sequences, there is no guarantee of good multi-dimensional coverage, even in the case of very good one-dimensional uniformity (although this clearly increases the scope for good multi-dimensional coverage). As described in the previous section, the one-dimensional quality of coverage of a grid method increases as M and hence $\sqrt[K]{M}$ (the number of intervals per dimension) increases; this quality of coverage will however necessarily be inferior to that obtained in single dimensions by our MLHS method, as our approach effectively uses N intervals along each dimension with one draw per interval. Only in the case where $K=1$ and $M=N$ will the systematic sampling approach lead to the same degree of one-dimensional uniformity of coverage. As the systematic sampling approach does not rely on any shuffling of the one-dimensional draws, but explicitly creates a grid of multi-dimensional points, the multi-dimensional quality of coverage should be superior to that of MLHS; this however presumes sufficiently high values for M (with correspondingly small cells and low numbers of draws per cell, resulting in more uniform coverage due to a lower effect of the random positioning of draws), which are often not possible.

The MLHS approach has a disadvantage in that it is not easily possible to accurately define the asymptotic properties of the method, since it is not a series. This is a drawback in that it does not allow us to compare the method asymptotically with other approaches. However, some comparisons can be made. Firstly, the coverage offered by MLHS is necessarily better than that of the shuffled Halton sequence, given that both approaches rely on random shuffling of one-dimensional sequences in the construction of multi-dimensional sequences, where with our new approach, the coverage offered by the one-dimensional uniform sequences is superior to that of the one-dimensional Halton sequences. Secondly, the same reasoning can be used in the

comparison with pseudo-random sequences. Indeed, multi-dimensional pseudo-random number sequences are in fact simply combinations of one-dimensional pseudo-random sequences, where the ordering of points in these one-dimensional sequences is similarly random, just as in the case of shuffled uniform vectors. The fact that the one-dimensional ordering of points is thus random in both cases means that, asymptotically, any differences between the two approaches in terms of the resulting multi-dimensional quality of coverage depend solely on the quality of coverage of the one-dimensional sequences, which is better in the case of uniform vectors.

EMPIRICAL APPLICATION

Data

The data used in this paper is for potential customers' choice between three types of vehicles in California; gas internal combustion (ICV), electric (EV) and gas-electric hybrid (HV). The data is described in detail by Train and Hudson (2000), and has also been used by Sándor and Train (2004) and Train and Sonnier (2003). A total of 500 respondents are included in the data set, and the respondents were presented with a number of different choice situations (up to 15) between three vehicles, where the individual choice sets do not necessary contain one vehicle of each of the three types of vehicles considered. The total number of observations in the data set is 7437. Two different data sets were available for the original study; a basic data set and an "enhanced" data set. The present analysis uses the latter data set, in which the respondents were provided with prior information on EV vehicles and on air quality in California. This has been observed to have a significant positive effect on the attitude of respondents towards EV vehicles (Train and Hudson, 2000).

Each alternative used in the data set is described by a total of 6 attributes:

- Car type (ICV, EV or HV)

- Body type (10 different types, ranging from mini car to mini van)
- Purchase price ($1000's)
- Operating cost ($/month)
- Performance (grouped into high, medium and low performance)
- Range (100's of miles between refueling/recharging)

Although the performance of the vehicles is simply divided into three levels, the respondents were actually provided with more detailed information on top speed and seconds needed to reach 60mph; these were however directly linked to the three levels of performance. In addition to this, the range attribute was set to constant values for ICV and HV vehicles; the reason for including this attribute was simply to gauge the effect on respondents' choices of increases in the range of EV vehicles.

Choice of model

The model used in the estimation process is a Mixed Logit model (c.f. Revelt and Train, 1998), allowing for random taste variation across decision-makers and also taking into account the repeated choice nature of the data set (panel data). The dummy variables for type of vehicle take the role of alternative specific constants; for reasons of identification, no coefficient was associated with ICV vehicles, such that the coefficients associated with the dummy variables of the other two types of vehicles represent the net impact of unmeasured variables (including general attitude of respondents) on the utility of EV and HV vehicles relative to ICV vehicles. Similarly, some normalization was performed for the other variables included in the model. For the body type, midsize car was taken as the base, whereas for the performance variables, medium performance was chosen as the base. Finally, as the range variable had been kept constant for ICV and HV vehicles, a coefficient associated with this attribute was only estimated for EV cars.

An important question arises with regards to what distributional assumptions should be made for the different coefficients. It can safely be assumed that the coefficients associated with the different attributes for car type, body type and performance are strongly dependent on the individual respondents' taste parameters and need thus not be restricted to having the same sign for all members of the population; accordingly, a Normal distribution was used for these coefficients. It is however very risky to simply associate a normally distributed coefficient with the other three types of attributes (price, operating cost and range) without some prior testing regarding the advantages of using a distribution that imposes some prior condition on the sign of the coefficients. For a discussion of the risks involved, and on guidance for which approach should be used, see for example Train and Sonnier (2003), Hensher and Greene (2001), Hess and Polak (2004) and Hess et al. (2004). The log-normal distribution is the standard choice in such scenarios; it is restricted to be strictly positive such that it can be used directly for desirable attributes (having strictly positive coefficients), and in combination with a sign change for undesirable attributes, such that a positive coefficient is associated with the negative value of an undesirable attribute. Previous applications using this data set have either assumed a fixed (negative) value for the price coefficient and a Normal distribution for all other coefficients (Train and Hudson, 2000 and Sándor and Train, 2004) or used a lognormal distribution for those coefficients that should have the same sign over all decision-makers (Train and Sonnier, 2003).

We started off by using Normal distributions for all coefficients and compared the log-likelihood of the resulting model to three models that use a lognormal distribution for one of the above-mentioned three coefficients. The biggest increase in the log-likelihood (from –6309.6 to –6297.2 with no additional parameters) was achieved by the model using a lognormal distribution for the price coefficient. The fact that the improvements offered by the log-normal

distribution are more important for the price coefficient than for the operating cost coefficient should come as no surprise; high purchase price is in this case clearly less desirable than high operating cost (different units), such that a wrongly-signed coefficient for the price parameter can be seen as a more significant misspecification than for the operating cost parameter. The next step consisted of using a lognormal distribution for either the operating cost coefficient or the range coefficient. Surprisingly, the latter leads to a marginally bigger improvement in the log-likelihood (from –6297.2 to -6294.3 compared to –6295.7). Finally, the best fit is given by a model using a lognormal distribution for all three coefficients (price, cost and range) and a Normal distribution for the remaining 13 coefficients. The use of this model results in a log-likelihood of –6286.1, which is a significant improvement over the original model (LL=-6309.6) as it comes without changing the number of parameters. This model now uses a random distribution for all 16 coefficients, leading to 32 parameters to be estimated. To the authors' knowledge, this is the first application comparing the performance of the different types of Halton sequences in a Mixed Logit application with this high a number of dimensions.

Generation of true parameters

In order to produce "true" values of the parameters for use as reference points in the comparison of the different types of draws, multiple runs with a high number of pseudo-random draws were used. A stability analysis was conducted to determine an appropriate number of draws to be used in this process. The results showed that beyond 2,000 draws, there were no significant changes in the log-likelihood function and parameter estimates over the different runs used; this was thus judged to be a high enough number of draws to guarantee stable estimation results, without unnecessarily increasing the computational cost of the estimation process. Ten separate runs were however still used at this stage to allow for the effects of randomness in the draws. From

the results, mean values of the parameters were calculated over runs, in addition to standard errors of the parameters (calculated as the square root of the average of the squared standard errors from individual runs). The standard errors are related to the shape of the log-likelihood function at the maximum. A high standard error for a parameter indicates a flat shape of the log-likelihood function for this parameter at the overall maximum of the function; slight changes in the value of the parameter have little effect on the value of the log-likelihood function. The converse is the case for low standard errors. The inverse of these standard errors can be used as weights for the parameters in the calculation of estimation performance (when comparing different types of draws), such that a higher simulation error is tolerated for parameters that have a higher standard error (c.f. Sándor and Train, 2004).

The results from this estimation are summarized in Table 1, which gives the estimates of the different parameters, along with their standard errors, and also gives the corresponding mean and standard variation values for those parameters based on the lognormal distribution. For these three coefficients, the estimated parameters are the mean c and standard deviation s of the log of the coefficient (rather than mean and the standard deviation of the coefficient itself). The mean and standard deviation of the actual coefficient can be calculated as $\mu = \exp\left(c + \frac{s^2}{2}\right)$ and $\sigma = \mu\sqrt{\exp\left(s^2\right) - 1}$ respectively.

The estimated values of the coefficients indicate the changes in utility following a unit change in the respective attribute, with all other attributes kept constant. The results produced in the present analysis are broadly consistent with those produced by Train and Hudson (2000). The positive value of the mean of the price and cost coefficients in conjunction with the negative value of the attributes means that an increase in price and/or operating cost leads to a decrease in

utility; this is clearly consistent with normal consumer behavior. The high value of the standard deviations for these parameters (along with the corresponding t-statistic) also supports the decision not to use a fixed value for the parameters – there is strong variation in price and cost sensitivity across decision-makers. The use of a lognormal distribution for the range coefficient implies that all respondents appreciate range increases; for this parameter, the variation in the population is however far less important, as can be seen from the low standard deviation. A similar observation was made by Train and Hudson (2000).

With medium performance used as the base, the signs of the coefficients for high and low performance are as expected, indicating that high performance is preferred to medium performance while low performance leads to lower utility than medium performance. The results suggest that the effect of moving from low to medium performance is around 2.5 times as important as the effect of moving from medium to high performance. Similar results were obtained by Train and Hudson (2000) when using the enhanced sample, whereas with the non-enhance sample, the ratio of coefficients was close to 1. This implies that the extra information given to respondents in the enhanced sample lessens their desire for increases in performance past the medium level.

The estimate for the coefficient associated with the dummy variable for electric vehicles can be most easily interpreted by summing it with the range coefficient (which was set to zero for all non-electric vehicles); this shows that, ceteris paribus, an electric vehicle needs a range of 353 miles to be valued equally highly as an ICV vehicle. With inferior range, EV vehicles are valued lower than ICV vehicles, while HV vehicles are valued more highly than ICV vehicles.

In terms of body type, only the mid-size SUV is preferred to the mid-size car (used as base); smaller vehicles are valued more lowly, possibly due to a combination of lower comfort

and lower safety performance, while larger vehicles are possibly valued more lowly due to extra strain they put on air quality (remembering that the associated higher operating cost in taken into account separately). It should also be noted that the mean of the coefficient associated with large SUV's is insignificant (low t-statistic), while its standard deviation is significant.

SIMULATION PERFORMANCE WITH DIFFERENT TYPES OF DRAWS

To compare the performance of the different types of methods, the root-squared-error (RSE) between the estimated value and the "true" value of each parameter was calculated, thus testing the ability of the draws to recover the "true" parameters. In order to account for the shape of the log-likelihood function, the RSE values were expressed as a proportion of the standard error of the true parameter (c.f. Sándor and Train, 2004). The same approach was used for the alternative statistics of standard deviation and bias. Ten runs were used for the non-deterministic methods (shuffled Halton draws, MLHS and pseudo-random vectors), and the errors were averaged over runs. For the Halton-based approaches, the first sixteen eligible primes (2 to 53) were used.

Table 2 shows the average RSE as a proportion of the standard error, where the average is over the 32 parameters that were estimated, and with experiments being conducted for sequences of length 50, 100, 200 and 500. Due to space constraints, only the overall measures are reproduced here; more detailed results are available from the first author on request. Similarly, for those methods where multiple runs were used, only the mean performance over runs is given in Table 2; it is however worth noting that the performance of the shuffling-based methods was found to be very stable over runs (reflecting earlier results by Hess et al., 2003).

The first observation that can be made from Table 2 is that the performance of the standard Halton draws is surprisingly good, despite the high correlation between some of the dimensions. Another interesting observation is that the use of scrambled Halton sequences

actually leads to poorer prediction performance than the use of standard Halton sequences for three out of the four sequence lengths (50, 100 and 200 draws). Except for the longest choice of sequence length (500 draws), the standard and scrambled Halton draws are outperformed by the shuffled Halton draws, suggesting that, especially with short sequences, the shuffling approach can offer improvements over the other two approaches. The MLHS method leads to better performance than any of the Halton sequences. This result holds for all four choices of length of sequence.

Finally, it should be noted that the use of pseudo-random draws leads to surprisingly good relative performance, given earlier results, for example by Bhat (1999) and Train (1999). Stated more directly, the alternative methods did not provide very much improvement over pseudo-random draws. A possible reason for the small improvement could be the high dimensionality of the application compared with those of Bhat and Train. As stated above, the Halton and MLHS procedures are not specifically designed in multiple dimensions; rather they are all designed as one-dimensional sequences that are combined to create multi-dimensional sequences. It might be the case that, as the number of dimensions rises, the importance of one-dimensional coverage diminishes. Or, stated alternatively, any unstructured combination of one-dimensional sequences may start to more closely resemble a purely random sequence as the number of dimensions rises. This conclusion suggests the need for comparisons with sequences like systematic sampling and *(t,m,s)*-nets, which, as stated above, provide better uniformity over dimensions by relinquishing uniformity in each dimension.

Very similar results were obtained when using the standard deviation or bias measures instead of RSE, as well as when using different lengths of sequences. It should be noted that the differences in performance between the single methods are rather small, and probably too small

to allow us to generalize the results. The advantage of our proposed approach may also be specific to this application, and further research is needed to compare the performance of the different methods in other applications (including systematic analyses using synthetic datasets), and also using different statistics in the comparison, such as for example the t-statistics produced with the different approaches. Along these lines, one could also attempt to compare the entire variance-covariance matrix across methods. Nevertheless, the results in this study are interesting and intuitive. They indicate that if a multi-dimensional sequence is going to be created by randomly combining one-dimensional sequences, then performance is enhanced by obtaining more uniform coverage in each dimension, which is attained by MLHS relative to shuffled Halton sequences, and by MLHS and shuffled Halton Halton sequences relative to pure pseudo-random sequences.

SUMMARY AND CONCLUSIONS

In this paper, we have proposed a modified Latin Hypercube Sampling (MLHS) approach for use in the estimation of MMNL models.

A practical application was conducted to compare its performance with that of various types of Halton sequences and pure pseudo-random sequences in simulation-based estimation, using a Mixed Logit model for choice between differently fuelled vehicles. The initial estimation conducted to generate "true" values of the parameters shows that while information campaigns have the potential to reduce customers' dislike of EV vehicles, there is still a preference for ICV and HV vehicles. Indeed, ceteris paribus, EV vehicles only become competitive with unrealistically appealing range characteristics (>353 miles).

The different types of sequences were then used to estimate the models with a lower number of draws, with the aim of assessing the ability of the different types of sequences to

recover the "true" values of the parameters. MLHS was found to outperform the other approaches in the current analysis. However, the overall differences in performance between the different methods are relatively small, suggesting that more extensive simulation experimentation is required to reach more reliable conclusions. Also, the comparisons are only among methods that create multi-dimensional sequences by combining one-dimensional sequences. Comparisons are needed between these methods and QMC sequences like grids and *(t,m,s)*-nets that are designed in multiple dimensions. Nevertheless, the initial results are promising particularly given the simplicity of implementing MLHS in any number of dimensions.

ACKNOWLEDGMENTS

The authors would like to acknowledge the fact that, independently of the research presented in this paper, the idea of using MLHS was also put forward by John Bowman in an e-mail to the first author, shortly after work on this analysis was started. The authors would also like to thank Bill Greene for suggestions on how to efficiently code the shuffling approach. Finally, the authors would like to thank Chandra Bhat and two anonymous referees for helpful comments and suggestions on an earlier version of this paper.

REFERENCES

Bhat, C. R. (1999), "Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model", *Transportation Research*, 35B(7), pp.677-693.

Bhat, C. R. (2000), "A multi-level cross-classified model for discrete response variables", *Transportation Research*, 34B(7), pp. 567-582.

Bhat, C. R. (2002), "Simulation Estimation of Mixed Discrete Choice Models Using Randomized and Scrambled Halton Sequences", *Transportation Research*, forthcoming.

Braaten, E. & Weller, G. (1979), "An improved low-discrepancy sequence for multidimensional quasi-Monte Carlo integration", *Journal of Computational Physics*, 33, pp. 249-258.

Halton, J. (1960), "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals", *Numerische Mathematik*, 2, pp. 84-90.

Hellekalek, P. (1984), "Regularities in the distribution of special sequences", *Journal of Number Theory*, 18, pp. 41-55.

Hensher, D. & Greene, W. H. (2001), "The Mixed Logit Model: The State of Practice and Warnings for the Unwary", *Institute of Transport Studies*, The University of Sydney, Sidney, Australia.

Hess, S., Polak, J.W. & Daly, A. (2003), "On the performance of the Shuffled Halton sequence in the estimation of discrete choice models", paper presented at the *European Transport Conference, Strasbourg*.

Hess, S. & Polak, J.W. (2004), "Mixed Logit estimation of parking type choice", paper presented at the 83rd Annual Meeting of the Transportation Research Board, Washington D.C.

Hess, S., Bierlaire, M. & Polak, J.W. (2004), "Estimation of value of travel-time savings using Mixed Logit models", Transportation Research A, *forthcoming*.

Kocis, L. & Whiten, W. J. (1997), "Computational investigations of low-discrepancy sequences", *ACM Transactions on Mathematical Software*, 23(2), pp. 266-294.

Lee, L. (1995), "Asymptotic bias in simulated maximum likelihood estimation of discrete choice models", *Econometric Theory*, 11, pp. 437-483.

McGrath, E. (1970), *Fundamentals of Operations Research*, West Coast University Press, San Francisco.

McKay, M.D., Conover W. J. & Beckman R. J. (1979), "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code", *Technometrics*, 21, 239-245.

Morokoff, W.J. & Caflisch, R.E. (1994), "Quasi-random sequences and their discrepancies", *SIAM Journal on Scientific Computing*, 15(6), pp. 1251-1279.

Niederreiter, H. (1992), "Random Number Generation and Quasi-Monte Carlo Methods", SIAM, Philadelphia.

Revelt, D. & Train K. (1998), "Mixed logit with repeated choices", *Review of Economics and Statistics*, 80, pp. 647–657.

Sándor, Z. & Train, K. (2004), "Quasi-Random Simulation of Discrete Choice Models", *Transportation Research*, Part B, Vol. 38, pp. 313-327.

Train, K. (1999), "Halton sequences for mixed logit", technical paper, Department of Economics, University of California, Berkeley.

Train, K. (2003), *Discrete Choice Methods with Simulation,* Cambridge University Press, Cambridge, MA.

Train, K. & Hudson K. (2000), "The impact of information in vehicle choice and the demand for electric vehicles in California", project report, National Economic Research Associates.

Train, K. & Sonnier, G. (2003), "Mixed Logit with Bounded Distributions of Correlated Partworths", forthcoming, *Applications of Simulation Methods in Environmental Resource Economics*, A. Alberini and R. Scarpa, eds., Kluwer Academic Publisher.

Tuffin, B. (1996), "On the use of low-discrepancy sequences in Monte Carlo methods," *Monte Carlo Methods and Applications*, Vol. 2, pp. 295-320.

Tuffin, B. (1998), "A new permutation choice in Halton sequences", in H. Niederreiter et al., eds., *Monte-Carlo and quasi-Monte-Carlo methods 1996, Lecture notes in statistics 127*, Springer, New York, pp. 427-435.

Wang, X. and Hickernell, F., (2000), "Randomized Halton sequences," *Mathematical and Computer Modelling*," Vol. 32, pp. 887-899.

TABLE 1: True parameters, generated by 10 runs using 2,000 pseudo-random number draws each

Parameter	Statistic	Value	Std. error	Mean for log-normal parameters	Std. Dev. for log-normal parameters
Price	c	-2.54594	0.059417	0.102375	0.085966
	s	0.73051	0.04285		
Op.Cost	c	-3.53995	0.108101	0.041745	0.04318
	s	0.85295	0.078472		
Range	c	-0.58633	0.247207	0.560528	0.068711
	s	0.122127	0.223853		
Electric	Mean	-1.97947	0.207765		
	Std. Dev.	1.277689	0.129988		
Hybrid	Mean	0.79084	0.102068		
	Std. Dev.	1.140494	0.098965		
High Perf.	Mean	0.183616	0.057337		
	Std. Dev.	0.608566	0.092543		
Low Perf.	Mean	-0.49192	0.058369		
	Std. Dev.	0.55102	0.100139		
Mini Car	Mean	-2.98284	0.234245		
	Std. Dev.	1.935799	0.308884		
Sm. Car	Mean	-1.32572	0.165322		
	Std. Dev.	1.11959	0.285104		
Lrg. Car	Mean	-0.46289	0.173889		
	Std. Dev.	1.182731	0.269354		
Sm. SUV	Mean	-0.79594	0.162877		
	Std. Dev.	0.757905	0.276538		
Mid SUV	Mean	0.331235	0.151385		
	Std. Dev.	0.777742	0.327428		
Lrg. SUV	Mean	-0.1597	0.237397		
	Std. Dev.	1.580258	0.409971		
Com. PU	Mean	-1.28973	0.175784		
	Std. Dev.	1.038175	0.281801		
Full PU	Mean	-0.77123	0.190277		
	Std. Dev.	1.588612	0.313687		
Minivan	Mean	-0.4788	0.187064		
	Std. Dev.	1.500402	0.254235		

TABLE 2: Mean prediction error when using different types of quasi-random number sequences (mean RSE as proportion of standard errors)

Draws	Standard Halton	Scrambled Halton	Shuffled Halton	MLHS	Pseudo-random number draws
50	0.75791	0.76823	0.75774	0.73766	0.76266
100	0.58074	0.59639	0.55211	0.50165	0.56638
200	0.37616	0.38775	0.36821	0.36296	0.38580
500	0.22777	0.21768	0.22801	0.21518	0.24476