Halton Sequences for Mixed Logit

By

Kenneth Train[1]
Department of Economics
University of California, Berkeley

July 22, 1999
Revised August 2, 1999

Abstract: The simulation variance in the estimation of mixed logit parameters is found, in our application, to be lower with 100 Halton draws than with 1000 random draws. This finding confirms Bhat's (1999a) results and implies significant reduction in run times for mixed logit estimation. Further investigation is needed to assure that the result is not quixotic or masking other issues.

I. Introduction

Choice probabilities in mixed logit models take the form of a multidimensional integral over a mixing distribution (see, e.g., Brownstone and Train, 1999). The integral does not have a closed form in general, and so it must be evaluated numerically. In applications, the integral has been approximated though simulation using random draws from the mixing distribution. A large number of draws is usually needed to assure reasonably low simulation error in the estimated parameters. The large number of draws translates into long computer run-times. Estimation can require 2-3 hours for moderately sized models, and run-times of 10-20 hours are not uncommon.

Numerous procedures have been proposed in the numerical analysis literature for taking "intelligent" draws from a distribution rather than random ones (see, e.g., Sloan and Wozniakowski, 1998; Morokoff and Caflisch, 1995.) The procedures offer the potential to

[1] Acknowledgements: The data for this analysis were collected by the Electric Power Research Institute (EPRI.) I am grateful to Ahmad Faruqui and EPRI for allowing me to use the data and present the results publicly. Andrew Goett and Kathleen Hudson, who had previously used these data, provided me datafiles in easily useable form, which saved me a considerable amount of time. I am also grateful to Chandra Bhat for sharing with me his GAUSS code for creating Halton sequences.

reduce the number of draws that are needed for mixed logit estimation, thereby reducing run times, and/or to reduce the simulation error that is associated with a given number of draws. In the only application of these procedures to mixed logits to my knowledge, Bhat (1999a)[2] tested Halton sequences for mixed logit estimation and found their use to be vastly superior to random draws. In particular, he found that the simulation error in the estimated parameters was lower using 100 Halton numbers than 1000 random numbers. In fact, with 125 Halton draws, he found the simulation error to be half as large as with 1000 random draws and smaller than with 2000 random draws.

In this paper, we examine Halton sequences in another application of mixed logit. Our results confirm Bhat's and illustrate the reasons for the improvement. In our application, simulation variance in the estimated parameters is found to be considerably smaller with 100 Halton numbers than 1000 random draws. The reasons for the improvement are twofold. First, the Halton numbers are designed to give fairly even coverage over the domain of the mixing distribution. With more evenly spread draws for each observation, the simulated probabilities vary less over observations, relative to those calculated with random draws. In our application, variance over draws in the simulated probability for an observation is half as large with 100 Halton draws than 1000 random draws, and is a third as large with 125 Halton draws than 1000 random draws. Second, with Halton sequences, the draws for one observation tend to fill in the spaces that were left empty by the previous observations. The simulated probabilities, therefore, become negatively correlated over observations (even when the data for each observation are the same.) This negative correlation reduces the variance in the log-likelihood function. In our application with 125 Halton draws for each observation, we obtained a correlation of -0.44 between the simulated probability for one observation and that for the immediately previous observation, while the correlation using random draws is essentially zero.[3]

[2] Another paper by Bhat (1999b) also uses Halton draws in mixed logit estimation but does not describe his tests against random draws.

[3] GAUSS code to estimate mixed logits using Halton draws is available from my website at http://elsa.berkeley.edu/~train.

II. Mixed Logit

Mixed logit models take the form:

$$P_{in} = \int L_{in}(\beta)\, f(\beta|\theta) d\beta$$

$$L_{in}(\beta) = \exp(\beta' x_{in}) / \sum_j \exp(\beta' x_{jn}).$$

where P_{in} is the choice probability for observation n and alternative i, $L_{in}(\beta)$ is the logit formula evaluated with coefficients β, and $f(\beta|\theta)$ is the density of β, which has parameters θ. Essentially, the mixed logit is a mixture of logits with f as the mixing distribution. The goal is to estimate the parameters θ of the mixing distribution.

In different applications, the parameters θ take on different meaning. The most common interpretation is based on random coefficients: Utility is specified as $U_{in} = \beta_n' x_{in} + e_{in}$ with agent-specific coefficients β_n that represent that agent's tastes. The researcher does not observe, and cannot estimate, the coefficients for each agent but knows that the coefficients vary in the population, with density f. For example, the coefficients may be distributed normally in the population, with mean θ_1 and variance θ_2. In this case, the goal is to estimate the mean and variance of tastes in the population.

The choice probabilities are evaluated numerically through simulation. Take R draws from density f, and label these draws β^r, r=1,...,R. For each β^r, calculate the logit formula. The simulated probability is the average of these calculated logits:

$$SP_{in} = (1/R) \sum_{r=1,\ldots,R} L_{in}(\beta^r) .$$

SP_{in} is an unbiased estimate of P_{in} whose variance decreases as R rises. The simulated log-likelihood function is created from the simulated probabilities, $SLL(\theta)=\sum_n \ln(SP_{in})$ where i denotes the chosen alternative for each observation n. The estimated parameters are those that maximize SLL. Properties of this estimator, based on smooth simulated probabilities, are given by Lee (1992) and Hajivassiliou and Ruud (1994.) Due to the non-linearity of the log transformation, $\ln(SP_{in})$ is not an unbiased estimate of $\ln(P_{in})$, such that the estimator based on maximizing SLL is biased. However, the bias decreases as the number of draws increases, and when the number of draws rises as fast as the square root of the number of observations, the estimator is consistent and equivalent to the classical maximum likelihood estimator.

In all previous applications of mixed logits to my knowledge, random draws have been used for the simulation. While the same draws could be used for each observation (see, e.g., Lee), the use of different draws for each observation allows simulation errors to cancel-out over observations. (This was the essential insight in McFadden's 1989 paper on method of simulated moments.) For our investigations in section V below, the models that are estimated with random draws use a different set of draws for each observation. The application of Halton sequences for simulation is described in the following section.

III. Halton Sequences

Halton sequences are best understood through an example. Start with a number that defines the sequence. For illustration, consider the number 3. A Halton sequence for number 3 is constructed as follows. Take the unit interval (0,1) and divide it into 3 parts. The dividing points become the first two elements of the Halton sequence: 1/3 and 2/3. Now take each of the three parts and divide them into 3 parts. The dividing points constitute the next elements in the Halton sequence: 1/9, 4/9 (which is 1/9 above 1/3), 7/9 (which is 1/9 above 2/3), and 2/9, 5/9 (which is 2/9 above 2/3), and 8/9 (which is 2/9 above 2/3). The unit interval has now been divided into nine parts. Divide each of these nine parts into thirds. The dividing points are 1/27, 10/27, 19/27, 4/27, 13/27, 22/27, 7/27, 16/27, 25/27 (which are 1/9 added to zero and the previous numbers) and 2/27, 11/27, 20/27, 5/27, 14/27, 23/27, 8/27, 17/27, 26/27 (which are 2/9 added to zero and the previous numbers.) Each of the 27 spaces are then divided into three parts, and so on for as many numbers as needed in the sequence. Similar sequences are defined for other numbers, such as 2 (1/2, 1/4, 3/4, 1/8, 5/8, 3/8, ...) and 5 (1/5, 2/5, 3/5, 4/5, 1/25, 6/25, 11/25, ...).

Draws for mixed logit estimation are created as follows. Using prime numbers, create a Halton sequence for each dimension of the mixing distribution. For example, if the mixing distribution describes the distribution of three random terms, create three Halton sequences. Base these Halton sequences on prime numbers, since the Halton sequence for a non-prime number divides the unit space the same as each of the primes that constitute the non-prime. For each element of each sequence, calculate the inverse of the cumulative mixing distribution that is appropriate for that dimension.[4] For example, if the mixing distribution is normal, take the inverse cumulative normal of each element of each sequence. The resulting values are the Halton draws from the mixing distribution.

The length of each sequence is determined by the number of observations and the numbers of draws that the researcher decides to use. With N observations and R draws per observation, sequences of length (N*R)+10 are created. The first 10 elements of the sequence are discarded, since the early elements have a tendency to be correlated over Halton sequences with different primes. (For example, the first four elements of the sequences for 5 and 7 are 1/5,2/5,3/5,4/5 and 1/7, 2/7, 3/7, 4/7, which are highly correlated.) After discarding the first ten elements, use the next R elements for the first observation, the next R for the second observation, and so on.

As illustration, consider a mixed logit that is specified as containing three normally distributed coefficients. Halton sequences are created for the first three primes: 2, 3 and 5. A sequence of triplets is then created from these three sequences, where the first term is the Halton sequence for the first prime, the second term is the Halton sequence for second prime, and the third is the Halton sequence for third prime:

[4] As when taking random draws, the mixing distribution is re-expressed in terms of standardized, independent distributions; the inverse of these standardized distributions (one for each dimension of the original mixing distribution) are taken. For example, suppose f is N(b,W), meaning that the coefficients of the logit function are distributed normally with mean b and covariance W. The coefficients are re-expressed as b+Se where S is the Choleski factor of W and e consists of iid standard normal deviates. The inverse cumulative standard normal of the Halton sequences gives draws of e. Since this re-expression is necessary with random draws, it is not an extra task when using Halton draws.

< 1/2, 1/3, 1/5 >

< 1/4, 2/3, 2/5 >

< 3/4, 1/9, 3/5 >

< 1/8, 4/9, 4/5 > ...

The inverse cumulative standard normal is evaluated at each element to obtain a sequence of draws from the three-dimensional normal mixing distribution:

< 0.0 , -0.43, -0.84 >

< -0.67, 0.43, -0.25 >

< 0.67, -1.22, 0.25 >

< -1.15, -0.14, 0.84 > ...

Discard the first 10 triplets, and use the others in groups of R for each of the observations.

IV. Application

For our investigations we use a mixed logit model of residential customers' choice of energy supplier. Surveyed customers were presented with conjoint-type choice experiments. In each experiment, the customer was presented four alternative suppliers with different prices and other characteristics. The suppliers differed on the basis of price (fixed price at a given cents per kWh, time-of-day prices with stated prices in each time period, or seasonal prices with stated prices in each time period), the length of the contract (during which the supplier is required to provide service at the stated price and the customer would need to pay a penalty for leaving the supplier), and whether the supplier was their local utility, a well-known company other than their local utility, or an unfamiliar company. The data were collected by Research Triangle Institute (1997) for the Electric Power Research Institute and have been used by Goett (1998) to estimate mixed logits with random draws. We utilize a similar specification to Goett's, eliminating or combining variables that he found to be insignificant. Details on the data and survey design are provided by these authors.

Table 1 gives the estimation results for a model estimated with 125 Halton draws for each observation.[5] There are six explanatory variables, and five of them are specified to have normally distributed coefficients. The price coefficient is specified to be fixed, such that the distribution of willingness to pay for each nonprice attribute (which is the ratio of the attribute's coefficient to the price coefficient) is normally distributed.[6]

The estimated price coefficient is negative and highly significant. The estimated means of the coefficients of nonprice attributes are all highly significant. All but one of the estimated standard deviations of the coefficients are highly significant, and the one that is not highly significant nevertheless has a t-statistic over 1. These results imply that there is considerable heterogeneity in customers' preferences for energy suppliers, such that a mixed logit is a significantly more realistic representation than a standard logit. The estimated parameters imply:

- The average customer is willing to pay about a quarter-cent per kWh in higher price in order to have a contract that is shorter by one year. Stated conversely, a supplier that requires customers to sign-onto a four-year contract must discount its price by one cent to attract the average customer.
- There is considerable variation in customers' attitudes towards contract length, with 27% of customers preferring a longer contract to a shorter contract. A long-term contract constitutes insurance for the customer against price increases, with the supplier being locked into the stated price for the length of the contract. However, the contract prevents

[5] The standard errors in Table 1 are based on the robust formula $H^{-1}G\ H^{-1}$, where G is the outer product of the gradient and H is the Hessian (calculated as the second derivative of the log-likelihood function.) As McFadden and Train (1997) point out, this formula correctly incorporates simulation noise, unlike the commonly used (and easier to calculate) G^{-1}.

[6] There are several reasons for keeping the price coefficient fixed. (1) As Ruud (1996) points out, mixed logit models have a tendency to be unstable when all coefficients are allowed to vary. Fixing the price coefficient resolves this instability. (2) If the price coefficient is allowed to vary, the distribution of willingness to pay is the ratio of two distributions, which is often inconvenient to evaluate. With a fixed price coefficient, willingness to pay for an attribute is distributed the same as the coefficient of the attribute. (3) The choice of distribution to use for a price coefficient is problematic. The price coefficient is necessarily negative, such that a normal distribution is inappropriate. With a lognormal distribution (which assures that the price coefficient is always negative), values very close to zero are possible, giving very high (implausibly high) values for willingness to pay.

the customer from being able to take advantage of reductions in market prices, since the customer is locked into the stated price. Apparently, a considerable share of customers value the insurance against higher prices more than they mind losing the option to take advantage of potentially lower prices. The degree of customer heterogeneity implies that the market can sustain different lengths of contracts, with suppliers making profits by writing contracts that appeal to segments of the population.

- The average customer is willing to pay a whopping 2.5 cents per kWh more for its local supplier than for an unknown supplier. Hardly any customers are willing to pay more for an unknown supplier than their local utility (a phenomenon that could occur if customers greatly dislike their current local utility.) This finding has important implications for competition. It implies that entry in the residential market by previously unknown suppliers will be very difficult, particularly since the price discounts that entrants can potentially offer in most markets are fairly small. The experience in California, where only 1% of residential customers have switched away from their local utility after more than a year of open access, is consistent with this finding.
- The average customer is willing to pay 1.7 cents per kWh for a known supplier relative to an unknown one. Note, however, that the average willingness to pay for a known supplier is only 0.8 cents less than for the local utility. Furthermore, the estimated standard deviations imply that a sizeable share of customers would be willing to pay more for a known supplier than for their local utility, presumably because of a bad experience or a negative attitude toward the local energy utility. These results imply that companies that are known to customers -- such as their long distance carriers, local telecommunications carriers, local cable companies, and even retailers like Sears and Home Depot -- can be expected to be relatively successful in attracting customers for electricity supply, particularly compared to companies that were unknown prior to their entry as an energy supplier. To enhance competition, regulators might take steps to encourage entry by telecommunications and cable companies rather than preventing or delaying it.
- The average customer evaluates the TOD rates in a way that is fairly consistent with TOD usage patterns. The mean coefficient of the dummy variable for the time-of-day (TOD) rates implies that the average customer considers these rates to be equivalent to a fixed price of 9.8c per kWh. Note that 9.8c is the average price that a customer would pay under

the TOD rates if 80% of its consumption occurred during the day (between 8AM and 8PM) and the other 20% occurred at night. These shares, while a little high for the day, are not unreasonable. The estimated standard deviation is highly significant, reflecting heterogeneity in usage patterns and perhaps in customers' ability to shift consumption in response to TOD prices. The estimated standard deviation is larger than reasonable, however, implying that a non-negligible share of customers treat the TOD prices as being equivalent to a fixed price that is higher than the highest TOD price or lower than the lowest TOD price. (This anomaly is one of the drawbacks of specifying distributions, like the normal or lognormal, that have unbounded support.)

- The average customer seems to avoid seasonal rates for reasons beyond the prices themselves. The average customers treats the seasonal rates as being equivalent to a fixed 10c per kWh, which is the highest seasonal price. An possible explanation for this result relates to the seasonal variation in customers' bills. Consumption is usually highest in the summer, when air-conditioners are being run. Energy bills are therefore higher in the summer than in other seasons, even under fixed rates. The variation in bills over months, without commensurate variation in income, makes it harder for customers to pay for their summer bills. In fact, nonpayment for most energy utilities is most frequent in the summer. Seasonal rates, which apply the highest price in the summer, increase the seasonal variation in bills. Customers would rationally avoid a rate plan that exacerbates an already existing difficulty. If this interpretation is correct, then seasonal rates combined with bill-smoothing (by which the supplier carries a portion of the summer bills over to the winter) could provide an attractive arrangement for customers and suppliers alike.

V. Investigation of Simulation Variance

We investigated the properties of the simulated probabilities as follows. We took the first observation in the data set and simulated its probability 1000 times. We based the simulations on the estimated parameters from Table 1. We first simulated the probability using random draws. We then simulated the probability using Halton draws, as if the 1000 simulations were for 1000 observations. That is, we created Halton sequences of length (R*1000)+10, where R is the number of draws used for each observation, discarded the first 10, and used the rest in groups of R to simulate the probability 1000 times. As in estimation of Table 1, we used the primes 2, 3, 5, 7, and 11 for the Halton sequences.

Tables 2 and 3 give statistics for the simulated probability using various numbers of random and Halton draws. The second row of each table gives the variance of the simulated probability over the 1000 simulations. The variance is lower with 50 Halton draws than 100 or 500 random draws, and the variance with 100 Halton draws is less than half that with 1000 random draws. This improvement is presumably due to the fact that Halton sequences are constructed to provide fairly even coverage along each dimension for each observation.

With random draws, the variance decreases at a rate of approximately 1/R, where R is the number of draws. With the Halton draws, the rate of decrease is faster: doubling the number of draws decreases the simulation variance by a factor of about three. This difference is expected (see the discussion in Bhat, 1999a, and Morokoff and Caflisch, 1994) and reflects the fact that coverage with Halton draws becomes more even as the number of draws increases, such that the advantage of having more draws is accentuated by having them more evenly placed.

The third row in each table gives the covariance between each simulated probability and the immediately previous simulated probability. Recall that the simulated probabilities using Halton draws were calculated as if the 1000 simulations were for a sample of 1000 observations. (Each of these 1000 "observations" has the same data so that correlations in data over observations do not mask the correlation due to the simulation procedure). For random draws, the covariance is small and, for any number of draws, is either positive or negative depending on the particular outcome in that situation. The correlation with random draws never exceeds 0.002 in magnitude. With the Halton draws, the covariance is consistently negative. This negative covariance reflects the tendency of the Halton draws for one observation to fill in the empty spaces that were left with previous observations.

As expected, the negative covariance with Halton draws decreases in magnitude as the number of draws increases. This reduction is attributable to two factors. First, the variance decreases as the number of draws rises. Second, with more Halton draws for each observation, coverage is better for each observation, leaving less opportunity for filling in the empty spaces from previous observations. The correlation coefficient accounts for the first of these factors (since the correlation is the covariance as a proportion of the variance). As shown in the fourth row of Table 3, the correlation is highly negative for all numbers of draws. It decreases slightly in magnitude as the number of draws rises, though the pattern is not strong. This highly negative correlation (which also occurs for the log of the simulated probability) provides an advantage in model estimation because it reduces the simulation error in the log-likelihood function.

As expected, the average of the simulated probabilities is essentially the same for both types of draws and each number of draws.[7] The average of the log of the simulated probability decreases as the number of draws increases. This reduction reflects the reduction in the bias that arises from taking the log transformation. With Halton draws, the reduction occurs only in the fifth digit when 100 or more draws are used, suggesting perhaps that bias is small with this number of Halton draws.

Consider now the estimated parameters of the mixed logit model. The model of Table 1 was estimated repeatedly using 100 Halton draws, 125 Halton draws, and 1000 random draws. For each type and number of draws, the model was estimated five times using a different sets of draws each time. With random draws, different sets of draws were obtained by using a different seed for the random number generator. For the Halton draws, we obtained different sets of draws by cycling the order of the primes, starting with 2,3,5,7,11, then 3,5,7,11,2, and so on.

Tables 4 and 5 give the means and standard deviations, respectively, of the estimated parameters over the five sets of draws for each of the three procedures. Examining the means provides us information about bias. As discussed above, the maximum simulated likelihood estimator based on random draws is biased due to the log transformation of the simulated probabilities, with the bias decreasing as the number of draws increases. The systematic nature of the Halton draws can potentially induce bias both in the estimated probabilities and their

[7] Interestingly, this average differs over numbers of draws in the third digit for random draws and only in the fifth digit for Halton draws, due to the lower variance in the simulated probabilities using Halton draws.

logs. As shown in Table 4, the means of the estimated parameters are very similar with 100 Halton draws as 1000 random draws. A t-test on the difference between the means indicates that the hypothesis of no difference cannot be rejected for any of the coefficients at any reasonable level of significance. These results suggests that either (i) bias is negligible in both cases, or (ii) the extent of bias with 100 Halton draws is essentially the same as that with 1000 random draws. Similar results occur with 125 Halton draws.[8]

Consider now the standard deviations in Table 5. Using 100 Halton draws, the standard deviations are lower for all but one coefficient than with 1000 random draws. For eight of the eleven coefficients, the standard deviations are half as large. This finding confirms the results that Bhat obtained in his Monte Carlo study. Given that both sets of draws give essentially the same means, the lower standard deviations with the Halton draws indicates that a researcher can expect to be closer to the expected values of the estimates using 100 Halton draws than using 1000 random draws. An even stronger statement is possible. Label the mean estimates using 1000 random draws (i.e., the first column in Table 4) as b_{1000r}. With one exception,[9] the root mean squared error against b_{1000r} is lower for the estimates using 100 Halton draws than for the estimates using 1000 random draws. This result can be interpreted as following. Suppose a researcher is considering using 1000 random draws; there is some expectation of this estimator over different random draws. Our results suggest that the researcher can expect to be closer to this expectation using 100 Halton draws than 1000 random draws.

An interesting, and perplexing, phenomenon occurred in the estimations using 125 Halton draws. Column 3 gives the standard deviations for the estimated parameters using the cycle of primes described above (that is, cycling the order of 2, 3, 5, 7, and 11). These standard deviations are generally higher than those obtained with 100 Halton draws, contrary to expectations. (They were nevertheless lower than those based on 1000 random draws.)

[8] The fact that the means are similar with 125 Halton draws as with 100 would seem to suggest that the bias using these numbers of Halton draws is negligible, since one would expect the bias to decrease as the number of draws increases. However, more extensive testing is needed in this regard, particularly given the issue, discussed below, about the two cycles of 125 Halton draws.

[9] The exception is the estimated standard deviation of the coefficient for “known supplier.”

Examination of the individual runs indicated that the first four runs obtained very similar estimates but that the fifth run obtained considerably different estimates. For example, the estimated price coefficient in the first four runs was 0.862, 0.865, 0.863, and 0.864, respectively, while the fifth was 0.911. We reestimated this fifth run using prime 13 instead of 11. The results were similar to those in the first four runs. The standard deviations using this estimate, instead of the original fifth estimate, are given in the last column of Table 4. These statistics conform to expectations, in that the standard deviations are lower with 125 draws than 100.

The question remains, however, of what caused this anomalous result. We tried several different starting values under the concept that perhaps the aberrant estimate was a local maximum; however, the same estimate was obtained from all different starting values. We also examined the effect of outliers among the Halton draws in this run, but found that the estimate did not change appreciably when outliers were truncated.

It is perhaps fitting to close this paper with a recognition of the anomaly, since it emphasizes our limited understanding of Halton sequences for estimation. Our results indicate that Halton draws provide substantially better simulations for mixed logit than random draws. However, much remains to be investigated. For example, there is a potential relationship between the number of draws that are used for each observation and the primes that are used for the Halton sequences. With primes of 2 and 3 for two-dimensional integration, a type of cycling of the Halton numbers occurs every 6 draws. Is it desirable or undesirable in this situation to set the number of draws to a multiple of 6? With primes of 2, 3, 5, 7, and 11 for five-dimensional integration, cycling of the five Halton sequences occurs every 2310 draws. Is it desirable or undesirable to have the number of draws be an integer fraction of 2310, such as 231? We discarded the first 10 draws from the Halton sequences. Is the number that should be discarded related to the primes that are used and the number of draws for each observation? These and other questions warrant investigation as we start to use Halton and other “intelligent” draws in estimation.

References

Bhat, C., 1999a, "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model," working paper, Department of Civil Engineering, University of Texas, Austin.

Bhat, C., 1999b, "A Multi-Level Cross-Classified Model for Discrete Response Variables, forthcoming, *Transportation Research*.

Brownstone and Train, 1999, "Forecasting New Product Penetration with Flexible Substitution Patterns," *Journal of Econometrics*, Vol. 89, pp. 109-129.

Research Triangle Institute, 1997, *Predicting Retail Customer Choices Among Electricity Pricing Alternatives*, Electric Power Research Report, Palo Alto.

Goett, A., 1998, *Estimating Customer Preferences for New Pricing Products*, Electric Power Research Report TR-111483, Palo Alto.

Hajivassiliou, V. and P. Ruud, 1994, "Classical Estimation Methods for LDV Models using Simulation," in R. Engle and D. McFadden, eds., *Handbook of Econometrics*, Vol. IV, New York: Elsevier.

Lee, L.-F., 1992, "On Efficiency of Methods of Simulated Moments and Maximum Simulated Likelihood Estimation of Discrete Choice Models, *Econometric Theory*, Vol. 8, pp. 518-552.

McFadden, D., 1989, "A Method of Simulated Moments for Estimation of the Multinomial Probit without Numerical Integration," *Econometrica*, Vol. 57, pp. 995-1026.

McFadden, D. and K. Train, 1997, “Mixed MNL Models for Discrete Response,” forthcoming, *Applied Econometrics*.

Morokoff, W., and R. Caflisch, 1995, “Quasi-Monte Carlo Integration,” *Journal of Computational Physics*, Vol. 122, pp. 218-230.

Ruud, P., 1996, “Approximation and Simulation of the Multinomial Probit Model: An Analysis of Covariance Matrix Estimation,” working paper, Department of Economics, University of California, Berkeley.

Sloan, J. and H. Wozniakowski, 1998, “When Are Quasi-Monte Carlo Algorithms Efficient for High Dimensional Integrals?” *Journal of Complexity*, Vol. 14, pp. 1-33.

Table 1: Mixed Logit Model of Customers' Choice Among Energy Supplier
Simulation based on 125 Halton draws.

Variable	Estimate	Standard error	t-statistic	Willingness to pay
Price, in cents per kWh, for fixed rates (zero for seasonal and time-of-day rates.)				
Fixed coefficient	-0.862	0.104	8.32	-1.00
Length of contract, in years				
Mean coefficient	-0.197	0.033	6.02	-0.229
Standard deviation in coefficient	0.318	0.079	4.03	0.369
1 if supplier is local energy utility, 0 otherwise.*				
Mean coefficient	2.125	0.256	8.32	2.46
Standard deviation in coefficient	0.886	0.882	1.01	1.03
1 if supplier is a well-known company (other than local utility), 0 otherwise.*				
Mean coefficient	1.437	0.185	7.78	1.67
Standard deviation in coefficient	0.857	0.419	2.04	0.99
Dummy for time-of-day rates: 11c/ kWh 8AM-8PM and 5c/kWh 8PM-8AM.**				
Mean coefficient	-8.440	1.144	7.38	-9.79
Standard deviation in coefficient	2.552	0.603	4.23	2.96
Dummy for seasonal rates: 10c/kWh in summer, 8c/kWh in winter, and 6c/kWH in spring and fall.**				
Mean coefficient	-8.651	1.244	6.96	-10.0
Standard deviation in coefficient	1.888	0.651	2.90	2.19

Number of observations: 4308. Log-likelihood at convergence: -4944.32. Likelihood ratio index: 0.1721
* Base for comparison is "An unfamiliar company supplies electricity."
** In the conjoint-type experiments, only one time-of-day and one seasonal plan was offered, with no variation in the rates. The dummy variables identify these plans, and the coefficients reflect customers' preferences for these particular plans with their specified rates.

Table 2: Statistics for Simulated Probability using Random Draws

Draws	100	500	1000
Simulated probability			
Mean	0.413878	0.414079	0.413901
Variance * 1000	0.828779	0.163057	0.081277
Covariance * 1000	0.012513	-0.002996	0.000159
Correlation	0.015104	-0.018360	0.001952
Simulated log of probability			
Mean	-0.884613	-0.882173	-0.882364
Variance * 1000	4.884061	0.952550	0.474979
Covariance * 1000	0.088126	-0.017148	0.000546
Correlation	0.018050	-0.017989	0.001150

Table 3: Statistics for Simulated Probability using Halton Draws

Draws	50	75	100	125	200
Simulated probability					
Mean	0.413875	0.413855	0.413840	0.413836	0.413841
Variance * 1000	0.119541	0.093428	0.035200	0.028625	0.011843
Covariance * 1000	-0.048179	-0.046457	-0.008545	-0.012548	-0.005192
Correlation	-0.402697	-0.496987	-0.242782	-0.437699	-0.438208
Simulated log of probability					
Mean	-0.882541	-0.882513	-0.882376	-0.882367	-0.882308
Variance * 1000	0.703250	0.548890	0.205749	0.167306	0.069146
Covariance * 1000	-0.282532	-0.272518	-0.050015	-0.073341	-0.030289
Correlation	-0.401423	-0.496226	-0.243102	-0.437717	-0.437872

Table 4: Means of Parameter Estimates

	1000 random draws	100 Halton draws	125 Halton draws (cycle A)	125 Halton draws (cycle B)
Price	-0.8607	-0.8588	-0.8731	-0.8584
Contract length				
Mean	-0.1955	-0.1965	-0.2002	-0.1957
Std. Dev.	0.3092	0.3158	0.3227	0.3134
Local supplier				
Mean	2.0967	2.1142	2.1478	2.1121
Std. Dev.	1.0535	1.0236	1.0106	0.9358
Known supplier				
Mean	1.4310	1.4419	1.4572	1.4310
Std. Dev.	0.8208	0.6894	0.8449	0.8115
TOD rates				
Mean	-8.3760	-8.4149	-8.5653	-8.3946
Std. Dev.	2.4647	2.5466	2.6087	2.5225
Seasonal rates				
Mean	-8.6286	-8.6381	-8.7916	-8.6159
Std. Dev.	1.8492	1.8977	1.9534	1.8760

Table 5: Standard Deviations of Parameter Estimates

	1000 random draws	100 Halton draws	125 Halton draws (cycle A)	125 Halton draws (cycle B)
Price	0.0310	0.0169	0.0210	0.0120
Contract length				
Mean	0.0093	0.0045	0.0071	0.0029
Std. Dev.	0.0222	0.0108	0.0162	0.0050
Local supplier				
Mean	0.0844	0.0361	0.0555	0.0248
Std. Dev.	0.1584	0.1180	0.0943	0.1057
Known supplier				
Mean	0.0580	0.0242	0.0390	0.0200
Std. Dev.	0.0738	0.1753	0.0454	0.0406
TOD rates				
Mean	0.3372	0.1650	0.2543	0.1281
Std. Dev.	0.1578	0.0696	0.1414	0.0588
Seasonal rates				
Mean	0.4134	0.1789	0.2585	0.1426
Std. Dev.	0.2418	0.0679	0.1303	0.0742