

## Customer-Specific Taste Parameters and Mixed Logit

by

David Revelt and Kenneth Train  
Department of Economics  
University of California, Berkeley

September 9, 1999

Revised: November 23, 1999

**Abstract:** With flexible models of customers' choices among products and services, we estimate the tastes (part-worths) of each sampled customer as well as the distribution of tastes in the population. First, maximum likelihood procedures are used to estimate the distribution of tastes in the population using the pooled data for all sampled customers. Then, the distribution of tastes of each sampled customer is derived conditional on the observed data for that customer and the estimated population distribution of tastes (accounting for uncertainty in the population estimates.) The procedure provides the same type of information and is similar in spirit to hierarchical Bayes (HB.) The procedure is computationally attractive when it is easier to calculate the likelihood function for the population parameters than to draw from the posterior distribution of parameters as needed for HB. We apply the method to data on residential customers' choice among energy suppliers in conjoint-type experiments. The estimated distribution of tastes provides practical information that is useful for suppliers in designing their offers. The conditioning for individual customers is found to differentiate customers effectively for marketing purposes and to improve considerably the predictions in new situations.

**Acknowledgements:** We have benefited from comments and suggestions by Greg Allenby, Joel Huber, Rich Johnson, Daniel McFadden, and Peter Rossi. Of course, we alone are responsible for our representations and conclusions. The data for this analysis were collected by the Electric Power Research Institute (EPRI.) We are grateful to Ahmad Faruqui and EPRI for allowing us to use the data and present the results publicly. Andrew Goett and Kathleen Hudson, who had previously used these data, provided us data files in easily useable form, which saved us a considerable amount of time. For interested readers, software to estimate mixed logits is available free from Train's web site at <http://elsa.berkeley.edu/~train> .

# Customer-Specific Taste Parameters and Mixed Logit

by

David Revelt and Kenneth Train

## **I. Introduction**

A customer's taste parameters, or part-worths, represent the value that the customer places on each attribute of a product when choosing among products. Knowing the tastes of individual customers, as well as the distribution of tastes in the population, allows a firm to identify opportunities for targeted marketing, to design products that attract specific customers, and to identify groups of customers with similar tastes. Prediction of choices in new situations, which is important for assessing the market feasibility of new products, is also improved with information on customers' tastes.

We describe and apply a method for obtaining this information under highly flexible specifications of customer choice and distributions of tastes. First, the distribution of tastes in the population is estimated with maximum likelihood procedures on the pooled data for all sampled customers. Next, the position of each customer's tastes within the population distribution is inferred by examining the choices of that customer. Stated more precisely, the distribution of each customer's tastes is derived, conditional on the customer's observed choices and the distribution of tastes in the population. For convenience, we call the method "conditioning of individual tastes" or COIT. The procedure has been known and discussed before, mainly with respect to its computational difficulty (e.g., Allenby and Rossi, 1999). We demonstrate its relative simplicity when used in conjunction with mixed logit and other highly flexible choice models.

COIT is similar in spirit and provides the same type of information as Hierarchical Bayes models of discrete choice (e.g., McCulloch and Rossi, 1994; Allenby and Rossi, 1999.) Both procedures provide information on the parameters of the population distribution as well as individual customer's conditional taste densities. Hierarchical Bayes (HB) uses Gibbs sampling to obtain draws from the posterior distribution of the population

parameters and each customer's tastes. Calculation of the likelihood function is not needed in HB, whereas it is central to the COIT procedure. Therefore, in situations where the likelihood function is difficult to calculate, HB provides a computationally feasible alternative to COIT. On the other hand, HB relies on taking draws sequentially from the conditional posterior distributions of the population and individual parameters based on researcher-defined prior distributions. When specifying and drawing from the conditional posterior distributions is cumbersome, COIT provides a computationally convenient alternative. COIT can be considered as a collapsed Gibbs sampler that, when feasible, provides computational efficiency relative to standard Gibbs samplers like HB (Liu, 1994). Also, the conditioning provides a diagnostic tool for model specification and for differentiating local versus global maxima of the likelihood function.

We describe and apply the method with a mixed logit model (Bhat, 1998; Revelt and Train, 1998), though other behavioral specifications, such as probit, could be used. McFadden and Train (1999) show that any discrete choice model can be approximated to any degree of accuracy by a mixed logit. Also, the likelihood function for mixed logit is fairly easy to calculate using direct simulation methods. HB has been applied mainly to pure probit models,<sup>1</sup> since the hierarchical conditioning and Bayesian updating is relatively straightforward with normal distributions and since the calculation of the likelihood function, while feasible, is relatively difficult. Allenby and Rossi (1999) and Bolduc, *et al.* (1996) indicate the relative ease and reduction in computer time from estimating pure probits with HB rather than likelihood function procedures. With mixed logit, the balance shifts: the likelihood function is relatively easy to calculate and drawing from the posterior distribution with the variety of distributions that mixed logit allows, while feasible, is relatively cumbersome. For similar reasons, COIT can be attractive for mixed probit, which, like mixed logit, allows for a variety of mixing distributions and has an easy-to-calculate likelihood function (Brownstone and Train, 1999). Researchers who

---

<sup>1</sup> We use the term "pure probit" to refer to a choice model in which all the random terms are normally distributed. A mixed probit is a mixture of independent probits, where the mixing distributions takes any form, including non-normal distributions. A mixed probit with a normal mixing distribution is a pure probit.

are comfortable using the GHK simulator, and are not as comfortable with Bayesian methods, might find that COIT is an attractive method for estimating pure probits as well.

We apply the method to residential customers' choice of energy supplier. Firms that sell electricity in newly opened energy markets can benefit from knowing the importance that customers place on various attributes of their offers, such as the contract length and the type of pricing (e.g., time-of-day rates that apply a different price for electricity consumed at different times of the day and seasonal rates that charge different prices in different seasons.) Data were collected on a sample of residential energy consumers in soon-to-be-opened markets. Each surveyed customer was presented with a series of conjoint-type choice situations in which respondents were asked to choose among several different energy providers and rate schedules. We estimate a mixed logit on these data (excluding the last choice, which we retain for use in assessing the predictive ability of the models.) We calculate each customer's conditional distribution of tastes and predict the customers' choices in the last situation. The conditional expectation of each customer's taste parameters is found to vary considerably over customers and to provide valuable information for differentiating customers for targeted marketing and cluster analysis. Predictions based on the conditional distribution of each customer's tastes are found to be far more accurate than those based on the population distribution of tastes only.

Section II describes COIT within the context of mixed logit and illustrates its properties with a small Monte Carlo experiment. The application to electricity supply data is described in Section III.

## II. Mixed Logit, Conditional Taste Density, and Predictions of Customer Choices

### A. The Mixed Logit Model

Let each sampled customer face a choice among  $J$  alternatives in  $T$  choice situations.<sup>2</sup> The utility that each customer obtains in choice  $t$  from alternative  $j$  is:

$$U_{ij} = \beta'X_{ij} + \varepsilon_{ij}, \quad t = 1, \dots, T, j = 1, \dots, J,$$

where  $\varepsilon_{ij}$  is iid extreme value and independent of  $\beta$ , and the index for customer is omitted for notational convenience. The coefficient vector  $\beta$  is known to the customer but not to the researcher. It varies over customers with density  $g(\beta | \theta)$ , where  $\theta$  represents the parameters of this distribution.<sup>3</sup> For example, if  $\beta$  is normally distributed in the population,  $\theta$  would represent the mean and covariance.

In each choice situation, the customer chooses the alternative that provides the greatest utility. Let  $y_t$  denote the customer's chosen alternative in situation  $t$ , and let  $y = (y_1, \dots, y_T)$  denote the customer's sequence of choices. Since the  $\varepsilon_{ij}$ 's are distributed extreme value, the probability conditional on  $\beta$  that the customer chooses alternative  $i$  in situation  $t$  is standard logit (McFadden, 1973):

$$L(i, t | \beta) = \frac{e^{\beta'X_{ti}}}{\sum_j e^{\beta'X_{tj}}},$$

---

<sup>2</sup>  $J$  can of course vary over choice situations, but we do not represent this dependence in our notation.

<sup>3</sup> This specification assumes that  $\beta$  does not vary over  $T$ . As in most conjoint-type surveys, the customers in our dataset responded to all choice situations in one sitting suggesting that this assumption is reasonable.

and since the  $\varepsilon_{ij}$ 's are independent over choice situations,<sup>4</sup> the probability of the customer's sequence of choices, conditional on  $\beta$ , is the product of logits:

$$(1) \quad P(y | \beta) = L(y_1, 1 | \beta) \cdot \dots \cdot L(y_T, T | \beta) \quad .$$

The researcher does not observe  $\beta$ , and so these conditional probabilities are integrated over all possible values of  $\beta$ , using the population density of  $\beta$ ,

$$P(y | \theta) = \int P(y | \beta) g(\beta | \theta) d\beta \quad .$$

$P(y | \theta)$ , which is called the mixed logit choice probability, is the probability of the customer's sequences of choices conditional on the parameters of the population distribution,  $g(\beta | \theta)$ . It does not exhibit the restrictive "independence from irrelevant alternatives" property of standard logit, and McFadden and Train (1999) show that any choice model can be approximated arbitrarily closely by a mixed logit with the appropriate specification of  $g(\beta | \theta)$ .

The integral in the mixed logit probability generally does not have a closed form, and so it is approximated numerically through simulation. In particular,  $R$  draws of  $\beta$  are taken from the density  $g(\beta | \theta)$ . For each draw, the product of logits in equation (1) is calculated, and the results are averaged over draws. The simulated probability, denoted  $\tilde{P}(y | \theta)$ , is this average:

$$(2) \quad \tilde{P}(y | \theta) = \frac{1}{R} \sum_r P(y | \beta^r).$$

---

<sup>4</sup> Even though the  $\varepsilon_{it}$ 's are independent over  $t$ , utility is correlated over choice situations due to the common influence of  $\beta$ .

The population parameters  $\theta$  are estimated by inserting  $\tilde{P}(y|\theta)$  for each customer into the log-likelihood function and maximizing the function over  $\theta$ . Properties of the maximum simulated likelihood estimator are given by Hajivassiliou and Ruud (1994) and Lee (1992.)

## B. Conditioning of Individual Tastes (COIT)

The density  $g(\beta|\theta)$  describes the distribution of tastes in the population. We want to determine, insofar as possible, where each customer's  $\beta$  lies in this distribution. Let  $h(\beta|y,\theta)$  denote the density of  $\beta$  conditional on the customer's sequence of choices and the population parameters. By Bayes' rule,

$$h(\beta|y,\theta) = \frac{P(y|\beta) \cdot g(\beta|\theta)}{P(y|\theta)} .$$

We use this distribution to calculate the conditional expectations of functions of  $\beta$ . Of particular interest are the conditional expectation of  $\beta$  itself, and the logit choice probabilities for a new choice situation that the customer has not yet faced.

Formally, if  $k(\beta)$  is a function of  $\beta$ , then

$$\begin{aligned} E(k|y,\theta) &= \int k(\beta) \cdot h(\beta|y,\theta) d\beta \\ &= \frac{\int k(\beta) \cdot P(y|\beta) \cdot g(\beta|\theta) d\beta}{P(y|\theta)} \\ &= \frac{\int k(\beta) \cdot P(y|\beta) \cdot g(\beta|\theta) d\beta}{\int P(y|\beta) \cdot g(\beta|\theta) d\beta} . \end{aligned}$$

A simulated approximation to this expectation is:

$$(3) \quad \tilde{E}(k | y, \theta) = \frac{\sum_r k(\beta^r) \cdot P(y | \beta^r)}{\sum_r P(y | \beta^r)}$$

where  $\beta^r$  is the  $r$ -th draw from the population density  $g(\beta | \theta)$ . If  $k(\beta)$  is the identity function,  $k(\beta) = \beta$ , then equation (3) gives the conditional expectation of  $\beta$ , the customer's expected tastes.<sup>5</sup>

An interesting re-formulation occurs when  $k(\beta)$  is the logit probability from a new choice situation. Suppose the customer faces a new situation which we label  $T+1$ . Conditional on  $\beta$ , the probability that the customer chooses alternative  $i$  is  $L(i, T+1 | \beta)$ . The researcher does not know the customer's  $\beta$  but does know the density of  $\beta$  conditional on the customer's previous choices and the parameters of the population density. The expected probability for alternative  $i$  is simulated as:

$$\tilde{P}(i, T+1 | y, \theta) = \frac{\sum_r L(i, T+1 | \beta^r) \cdot P(y | \beta^r)}{\sum_r P(y | \beta^r)}.$$

Or if we let  $y^+$  denote  $(y_1, \dots, y_T, i)$ , then

$$(4) \quad \tilde{P}(i, T+1 | y, \theta) = \frac{\sum_r P(y^+ | \beta^r)}{\sum_r P(y | \beta^r)},$$

which is simply the simulated probability of the entire sequence of choices divided by the simulated probability of the sequence of previously observed choices.

---

<sup>5</sup> Instead of using equation (3), draws from  $h$  can be obtained with the Metropolis-Hastings algorithm (Chib and Greenberg, 1995), as in Arora et al. (1998) and Sawtooth Software, Inc. (1999). Then  $k(\cdot)$  is calculated for each draw and the results averaged.

The above formulas take the population parameters  $\theta$  as given,<sup>6</sup> but the researcher utilizes an estimate of  $\theta$  that varies from the true value due to sampling and simulation error. When the number of draws increases faster than the square root of the number of observations, the maximum simulated likelihood estimator of the mixed logit parameters is consistent, asymptotically normal, and has a covariance given by McFadden and Train (1999). If  $f(\theta | m, W)$  denotes the multivariate normal density of  $\theta$  with mean  $m$  and covariance  $W$ , the expectation of  $k(\beta)$  conditional on  $m$ ,  $W$ , and the customer's choices  $y$ , is:

$$E(k | y, m, W) = \int E(k | y, \theta) f(\theta | m, W) d\theta \quad .$$

A simulated approximation to this expectation is obtained by taking draws of  $\theta$  from the distribution of the estimator, and, for each draw, simulating the expectation of  $k$  conditional on that value of  $\theta$ :

$$(5) \quad \tilde{E}(k | y, m, W) = \frac{1}{S} \sum_s \tilde{E}(k | y, \theta^s)$$

where  $\theta^s$  is the  $s$ -th draw from  $f(\theta | m, W)$ .

In section III, we explore differences between various calculations of the conditional expectations, namely:  $\tilde{E}(k | y, \hat{m}, \hat{W})$ , which incorporates sampling variance in the estimates of the population parameters;  $\tilde{E}(k | y, \hat{\theta})$ , which ignores this sampling variance;  $k(\tilde{E}(\beta | y, \hat{m}, \hat{W}))$ , which calculates  $k$  based on the expected tastes and incorporates the sampling variance of the population estimates; and  $k(\tilde{E}(\beta | y, \hat{\theta}))$  which ignores the sampling variance. The second and fourth methods are faster to calculate than the first and third, and the last two can be calculated by retaining only each customer's conditional mean of  $\beta$  rather than their entire conditional distribution.

---

<sup>6</sup> Allenby and Rossi (1999) use the term “approximate Bayesian” for COIT based on point estimates of the population parameters.

### C. Monte Carlo Exercise

We illustrate the COIT procedure on a simulated data set where the true population parameters  $\theta$  are known as well as the true  $\beta$  for each customer. This simulation allows us to compare the predicted tastes,  $E(\beta)$ , with the true  $\beta$  for each customer as well as investigate the impact of increasing the number of choice situations and random draws. For this experiment, we constructed fifty data sets consisting of 300 “customers” each with 1, 10, 20, and 50 choice situations. There are three alternatives and four variables in each data set. The coefficients for the first two variables are held fixed for the entire population at 1.0, and the coefficients for the last two variables are distributed normal with a mean and variance of one. We use the true population parameters for  $\theta$ , and 100, 1000, and 10,000 draws from  $g(\beta | \theta)$  to calculate  $E(\beta)$  for each customer. To increase the simulation accuracy, we use Halton draws instead of random draws.<sup>7</sup>

Figure 1 depicts the conditional distribution for one “respondent” using different numbers of observed choice situations (and 10,000 draws in the simulation.) The first panel gives the distribution without conditioning which is the population distribution. As the number of observed choice situations rises, the conditional distribution becomes more concentrated. With ten observed choices, the conditional distribution is quite tight, with a variance of 0.33 for each coefficient compared to variances of 1 for the unconditional distribution.

The mean of the conditional distribution is a summary measure that can be used to differentiate customers. Table 1 presents the average standard deviation of  $E(\beta)$  as well as the average absolute deviation between the  $E(\beta)$  and the true  $\beta$ . The standard deviation of  $E(\beta)$  would be zero if there were no conditioning and one if each customer’s coefficients were known exactly. Similarly, the average absolute deviation of  $E(\beta)$  from the true  $\beta$  is 0.8 without conditioning and zero if each customer’s coefficients were known exactly.

---

<sup>7</sup> See Bhat, 1999, and Train, 1999, for a discussion of Halton draws with mixed logit.

With only one choice situation, a considerable amount of the variation in tastes is captured through conditioning. The average standard deviation in  $E(\beta)$  rises about 40% of the way to the standard deviation that occurs with perfect knowledge of each customer's tastes. The average absolute deviation between expected tastes and actual tastes drops from 0.8 (without conditioning) to about 0.72. With ten choice situations, which is similar to our data set, the average standard deviation in  $E(\beta)$  is over 80% of the value that it would have with perfect knowledge, and the absolute deviation is less than half as high as would be attained without conditioning. Strongly decreasing returns are evident. Going from 10 to 20 choice situations only raises the standard deviation from around 0.82 to around 0.88, and the absolute deviation between expected and actual coefficients drops from around 0.45 to 0.35. With fifty choice situations, the expected coefficients for each customer are only 0.25 different from their actual values, on average.

The number of Halton draws that are used in simulation does not seem to materially affect the results if more than 100 draws are used. There is essentially no difference between 1,000 and 10,000 draws suggesting that the use of 100,000 draws, as we used in our empirical analysis in section III, is more than enough. With 100 draws, the standard deviations are slightly lower than with 1,000 or 10,000 draws, and the absolute deviations are slightly higher, with these differences increasing with the number of choice situations. This result suggests that the number of draws should rise with the number of choice situations that are used in conditioning.

#### **D. Comparison of Sample Average of Conditional Distributions with Population Distribution**

For a correctly specified model at the true population parameters, the conditional distribution of tastes, aggregated over all customers, equals the population distribution of tastes. Denote the true frequency of  $y$ , conditional on the explanatory variables, as  $f(y/\theta^*)$ , which depends on the true parameters  $\theta^*$ . If the model is correctly specified and

consistently estimated  $P(y|\hat{\theta})$  approaches  $f(y|\theta^*)$  asymptotically. Conditional on the explanatory variables, the expected value of  $h(\beta|y, \hat{\theta})$  is then:

$$E_y h(\beta|y, \hat{\theta}) = \sum_y \frac{P(y|\beta) \cdot g(\beta|\hat{\theta})}{P(y|\hat{\theta})} f(y|\theta^*) \rightarrow \sum_y P(y|\beta) \cdot g(\beta|\hat{\theta}) = g(\beta|\hat{\theta}).$$

This relation provides a diagnostic tool, similar to that described by Rossi and Allenby (1999): If the average of the moments of the sampled customers' conditional taste distributions are similar to the moments of the estimated population distribution, the model is correctly specified and accurately estimated. If the moments are not similar, the difference could be due to: (1) specification error, (2) an insufficient number of draws in simulation, (3) an inadequate sample size, and/or (4) the maximum likelihood routine converging at a local rather than global maximum. Note that the distinction between a local and global maximum is difficult to diagnose in flexible discrete choice models without using COIT.

## **E. Hierarchical Bayes**

It is useful at this point to say a few more words about COIT in comparison to HB. In principle, both methods can be applied with any behavioral model of customer choice that allows for heterogeneous tastes. While we have described COIT in terms of mixed logit, other behavioral models, such as probit, could be used. And while most HB applications have used probit, several recent applications of HB have specified mixed logits (Arora, Allenby, and Ginter, 1998; Sawtooth Software, Inc., 1999). When the same behavioral model is used in each, the two methods give essentially the same results with sufficiently large sample size and number of draws. For illustration, we applied HB and COIT to a simulated data set where the true population parameters and each customer's true tastes were known. The model was specified to be mixed logit with tastes normally distributed in the population. The simulated data set consisted of 300 "customers," with each customer facing ten choice situations with three alternatives. Both methods returned the true population parameters (up to the fourth decimal point), and the correlation between

the expected tastes calculated by COIT and those using HB was 0.98 or higher.

There are two important differences between the methods. First, they differ in the way that uncertainty about the population parameters  $\theta$  is conceptualized and calculated. With COIT, a covariance matrix is obtained for the estimate of  $\theta$  as part of the maximum likelihood procedure. This covariance reflects sampling variance and simulation noise, and is calculated on the basis of asymptotic formulas. With HB, the researcher specifies a prior distribution on  $\theta$  which reflects the researcher's knowledge and uncertainty about  $\theta$  before examining the data. The covariance matrix of the posterior on  $\theta$  is calculated from Bayes' formula and reflects the researcher's remaining uncertainty after updating the prior distribution with the information that is contained in the data.

The second major difference was mentioned in the Introduction. COIT requires calculation of the likelihood function for  $\theta$ , and HB requires draws from the conditional posterior distributions of  $\theta$  and each customer's  $\beta$ . For a given specification of the behavioral model, one of these methods might be more convenient computationally than the other. Most applications of HB have specified tastes to be the normally distributed in the population because of the convenient conditioning and Bayesian updating properties of normals, such that drawing from the posterior distribution is easy.<sup>8</sup> There are many situations, however, in which a normal distribution for tastes is inappropriate. For example, any parameter that is known to have a particular sign, such as a price coefficient or the part-worth of a desirable attribute, is not meaningfully represented by the normal since some share of the population will necessarily be given the wrong sign for this parameter. The use of normal distributions in these situations has been found to provide counterintuitive forecasts (Brownstone and Train, 1999). HB can and has been applied with more flexible distributions (e.g., Boatwright, McCulloch, and Rossi, forthcoming.) However, the specification of the posterior distribution and the mechanics of taking

---

<sup>8</sup> McCulloch and Rossi (1994) and Allenby and Rossi (1999) specify normally distributed tastes within a pure probit model, while Arora, Allenby, and Ginter (1998) and Sawtooth Software, Inc. (1999) specify normally distributed tastes within a mixed logit.

draws from it are usually more difficult than with normals. COIT allows many alternative distributions with no additional burden.

On the other hand, HB provides an advantage over COIT if the researcher has prior information that is difficult to incorporate into the form of the population distribution. For example, if the researcher knows that the distribution is normal with parameters that are likely to lie within a certain range, the researcher can incorporate this information into HB by specifying a non-diffuse prior. In contrast, incorporating this information into COIT is difficult.

### **III. Application To Power Supply Data**

#### **A. Model Specification and Results**

We apply the method in Section II to stated-preference data on residential customers' choice of electricity supplier. Surveyed customers were presented with 8-12 conjoint-type choice experiments. In each experiment, the customer was presented with 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. We utilize a specification similar to Goett's, but eliminating or combining variables that he found to be insignificant. Details on the data and survey design are provided by these authors.

We use these data to estimate four mixed logit models using different assumptions for the mixing distributions. All choices except the last situation for each customer are used to

estimate the parameters of the population distribution, and the customer's last choice situation was retained for use in comparing the predictive ability of different models and methods.

Table 2 gives the estimated population parameters for the four mixed logit models.<sup>9</sup> The price coefficient in each of the four models is fixed across the population such that the distribution of willingness to pay for each non-price attribute (which is the ratio of the attribute's coefficient to the price coefficient) has the same distribution as the attribute's coefficient.<sup>10</sup> For model 1, all of the non-price coefficients are assumed to be normally distributed in the population. The mean  $m$  and standard deviation  $s$  of each coefficient are estimated. For model 2, the non-price coefficients have uniform distributions. The mean  $m$  and spread  $s$  are estimated, where the spread is the distance from the mean to the edge of the support, such that the uniform has support  $[m-s, m+s]$ . The uniform distribution has the advantage of having bounded support, which reduces the chance of unreasonably high or low coefficient values. For model 3, each non-price coefficient has a triangular distribution. Similar to the uniform, the mean  $m$  and spread  $s$  are estimated, giving a support of  $[m-s, m+s]$ . This distribution has both the bounded support of the uniform and the peaked density that is highest at the mean similar to the normal. For model 4, the first three non-price coefficients are specified to be normal, and the fourth and fifth are log-normal. The fourth and fifth variable are indicators of time-of-day and seasonal rates, and

---

<sup>9</sup> In addition to these four models, Revelt (1999) estimated models with truncated normal distributions for the random coefficients. These models fit about the same as models 2 and 3.

<sup>10</sup> 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.

their coefficients must logically be negative for all customers. The log-normal distribution (with the signs of the variable reversed) provides for this necessity. For the log-normal coefficients, the parameters  $m$  and  $s$  are estimated where the mean of the coefficient is  $e^{m+(s^2/2)}$  and the standard deviation is the mean multiplied by  $\sqrt{(e^{s^2/2} - 1)}$ . It is important to note that even though log-normals assure the correct sign many researchers have found them to be problematic in practice because their parameters can be difficult to estimate and because of their unbounded upper support (Revelt and Train, 1998; Algers et al., 1999; Brownstone and Train, 1999.) In all four models, the simulation of the choice probabilities in equation (2) was based on Halton draws to reduce simulation error (Bhat, 1998; Train, 1999.)

We discuss the estimation results for the population distribution before turning to the calculation of the customers' conditional taste densities. The estimated price coefficient and the estimated means of the non-price coefficients are very similar in all four models, and the estimates of  $s$  (which differs in interpretation over the models) provide qualitatively similar implications. The models differ, however, in the share of customers whose tastes are above or below certain levels, and these difference can have important implications for forecasting, particularly for suppliers offering niche products.

The estimates from our model produce the following qualitative results:

- The average customer is willing to pay about a fifth to a quarter cent per kWh in higher price, depending on the model, in order to have a contract that is shorter by one year. Stated conversely, a supplier that requires customers to sign a four to five-year contract must discount its price by one cent per kWh to attract the average customer.
- There is considerable variation in customers' attitudes towards contract length, with a sizeable share 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. Such contracts prevent the customer that arise during the term of the contract. Apparently,

many 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 contracts of different lengths with suppliers making profits by writing contracts that appeal to different 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. Only a small share of customers prefer an unknown supplier to their 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 1.7-1.8 cents per kWh for a known supplier relative to an unknown one. The estimated values of  $s$  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—may be successful in attracting customers for electricity supply relative to companies that were unknown prior to their entry as an energy supplier.
- The average customer evaluates the TOD rates in a way that is fairly consistent with time-of-day (TOD) usage patterns. In models 1-3, the mean coefficient of the dummy variable for the TOD rates implies that the average customer considers these rates to be equivalent to a fixed price of 9.4-9.7¢ per kWh. (In model 4, the estimated mean and standard deviation of the log of the coefficient imply a median willingness to pay of 8.4 cents and a mean of 10.4 cents, which span the means from the other models.) 9.5¢ is the average price that a customer would pay under the TOD rates if 75% of its consumption occurred during the day (between 8AM and 8PM) and the other 25%

occurred at night. These shares, while perhaps slightly high for the day, are not unreasonable. The estimated values of  $s$  are highly significant, reflecting heterogeneity in usage patterns and perhaps in customers' ability to shift consumption in response to TOD prices. These values are larger than reasonable 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.

- The average customer seems to avoid seasonal rates for reasons beyond the prices themselves. The average customer treats the seasonal rates as being equivalent to a fixed ten cents per kWh, which is the highest seasonal price. A 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, and 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 more difficult 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.

Model 4 attains the highest log-likelihood value. Model 1, with all non-price coefficients distributed normal, obtains the lowest log-likelihood presumably because of its unbounded support in both directions which implies unreasonable values for at least some strictly positive share of customers.

## **B. Expected Tastes**

We now use the estimated models to calculate customers' conditional taste densities and expected tastes. We calculate  $E(\beta)$  for each customer in two ways. First, we calculate  $\tilde{E}(\beta | y, \hat{\theta})$  using equation (3) with the point estimates of the population parameters  $\theta$  and the same 100,000 Halton draws for each customer. Second, we use formulas (3) and (5) integrating out the sampling distribution of the population parameters to calculate  $\tilde{E}(\beta | y, \hat{m}, \hat{W})$ . This method uses 1000 random draws of the population parameters and 100 Halton draws from the conditional taste densities for each customer given the population parameters (i.e.,  $S=1000$ ,  $R=100$ ), for a total of 100,000 evaluations.

The means and standard deviations of  $E(\beta)$  over the sampled customers using these two methods are given in Tables 3 and 4, respectively.<sup>11</sup> Consider Table 3 first. The means are very close to the estimated population means, as expected for a correctly specified and consistently estimated model. The standard deviation of  $E(\beta)$  would be zero if there were no conditioning and would equal the population standard deviation if each customer's coefficient were known exactly. The standard deviations in Tables 3 are considerably above zero and are fairly close to the estimated population standard deviations. For example, in model 1, the expected coefficient of contract length has a standard deviation of 0.318 over customers, and the point estimate of the standard deviation in the population is 0.379 (see Table 2.) Thus, variation in the expected coefficient captures more than 70% of the total estimated variation in this coefficient. Similar results are obtained for other coefficients. This result implies that expected tastes conditioned on customer's previous choices captures a fairly large share of the variation in tastes across customers and has the potential to be useful in distinguishing customers.

As discussed in section II.D, a diagnostic check on the specification and estimation of the model is obtained by comparing the sample average of the conditional distributions with the estimated population distribution. The means in Tables 3 represent the means of the

---

<sup>11</sup> The price coefficient is not listed in Table 3 since it is fixed across the population. Table 4 incorporates the sampling distribution of the population parameters, which includes variance in the price coefficient.

sample average of the conditional distributions. The standard deviations of the sample-average conditional distribution depend on the standard deviations of  $E(\beta)$ , which are given in Table 3, plus the standard deviation of  $\beta$  around its expectation for each customer. When this latter portion is added, the standard deviation of each coefficient matches very closely the estimated population standard deviation, which suggests that there is not significant specification error and that the estimated population parameters are fairly accurate. This suggestion is tempered, however, by the results in Table 4.

Figure 4 gives the sample mean and standard deviation of  $\tilde{E}(\beta | y, \hat{m}, \hat{W})$ . The means in Table 4 represent the means of the sample average of  $h(\beta | y, \theta)$  integrated over the sampling distribution  $\theta$ . For models 1-3, a discrepancy occurs that indicates possible misspecification. In particular, the means of the TOD and seasonal rates coefficients in Table 4 exceed their estimated population means in Table 2. Interestingly, the means for these coefficients in Table 4 for models 1-3 are closer to the analogous means for model 4 than to the estimated population means for model 1-3 in Table 2. Model 4 has the more reasonably shaped lognormal distribution for these coefficients and obtains a considerably better fit than the other models. The conditioning in models 1-3 appears to be moving the coefficients closer to the values in the better-specified model 4 and away from their own misspecified population distributions. This is an example of how a comparison of the estimated population distribution with the sample average of the conditional distribution can reveal information about specification and estimation. For reasons that we do not understand, this phenomenon does not occur (or is far less pronounced) when the point estimate of  $\theta$  is used (Table 3) rather than the sampling distribution of  $\theta$ .

The standard deviations in Table 4 are larger than those in Table 3. This difference is due to the fact that the variance in the estimated population parameters is included in the calculations for Table 4 but not for Table 3. The larger standard deviations do not mean that the portion of total variance in  $\beta$  that is captured by variation in expected tastes is larger when the sampling distribution is considered than when not. Rather, the total

variation in  $\beta$  is itself greater when the sampling distribution of the population parameters is considered than when not.

Revelt (1999) performed cluster analysis on the  $E(\beta)$ 's from each model. We summarize here the results for model 4. Five clusters were identified. The first cluster is characterized by a strong dislike for time-of-day and seasonal rates and comparatively little preference for the local utility. This cluster is a potential target for new suppliers that offer fixed rates. The fourth and fifth clusters tend to prefer seasonal and TOD rates. They differ from each other primarily in their attitude toward the local utility with the customers in the fourth cluster having a very strong preference for the local utility. These customers can be attracted by offering variable rates, with the local utility concentrating especially in the fourth cluster. It is harder to identify how to attract customers in the second and third clusters. The customers are about average in their attitudes about contract length, the local utility, and known suppliers. Customers in the second cluster are more open than the average customer to TOD rates (though not to seasonal rates) and can perhaps be attracted by suppliers offering these rate. None of the attitudes of customers in the third cluster lend themselves naturally to a marketing plan.

More specific marketing information can be obtained by examining the  $E(\beta)$  of each customer. The  $E(\beta)$ 's from model 4, using integration over the sampling distribution of  $\theta$ , are given below for the first three customers in the data set, along with the population means of  $\beta$ .

Variable	Expected coefficient			
	Population	Customer 1	Customer 2	Customer 3
Contract length	-0.213	0.198	-0.208	-0.401
Local utility	2.23	2.91	2.17	0.677
Known company	1.59	1.79	2.15	1.24
TOD rates	-9.19	-5.59	-8.92	-12.8
Seasonal rates	-9.02	-5.86	-11.1	-10.9

The first customer wants to enter a long-term contract, compared with the vast majority of customers who dislike long-term contracts. He is willing to pay a higher energy price if the price is guaranteed through a long-term contract. He evaluates TOD and seasonal rates very generously, as if all of his consumption were in the lowest-priced period (note that the lowest price under TOD rates is 5 cents per kWh and the lowest price under seasonal rates is 6 cents per kWh.) That is, the first customer is willing to pay to be on TOD or seasonal rates probably more than the rates are actually worth in terms of reduced energy bills. Finally, this customer is willing to pay more than the average customer to stay with the local utility. From a marketing perspective, the local utility can easily retain and make extra profits from this customer by offering a long-term contract under time-of-day or seasonal rates.

The third customer dislikes seasonal and TOD rates, evaluating them as if all of his consumption were in the highest-priced periods. He dislikes long-term contracts far more than the average customer, and yet, unlike most customers, prefers to receive service from a known company that is not his local utility. This customer is a prime target for capture by well-known company if the company offers him a fixed price without requiring a commitment.

The second customer is less clearly a marketing opportunity. A well-known company is on about an equal footing with the local utility in competing for this customer. This in itself might make the customer a target of well-known suppliers, since he is less tied to the local utility than most customers. However, beyond this information, there is little beyond low prices (which all customers value) that would seem to attract the customer. His evaluation of TOD and seasonal rates are sufficiently negative that it is unlikely that a supplier could attract and make a profit from the customer by offering these rates. The customer is willing to pay to avoid a long-term contract, and so a supplier could attract this customer by not requiring a contract if other suppliers were requiring contracts. However, if other suppliers were not requiring contracts either, there seems to be little

leverage that any supplier would have over its competitors. This customer will apparently be won by the supplier that offers the lowest fixed price.

The discussion of these three customers illustrates the type of information that can be obtained with COIT and how the information translates readily into characterizing each customer and identifying profitable marketing opportunities.

### **C. Conditional Expected Probability for the Last Choice**

In this section, we use the electricity supply data to calculate the probability of each customer's choice in their last choice situation. Specifically, we identify the alternative that was actually chosen by the customer and calculate the probability for that alternative in four ways. First, we calculate the probability using formulas (4) and (5), which takes into account the sampling variance of the estimates of the population parameters. Second, we calculate the probability using formula (4) with the point estimates of the population parameters. Third, we calculate the probability as  $L(i, T+1/E(\beta))$  where  $E(\beta)$  is from formulas (3) and (5) with  $k(\beta)=\beta$  and incorporates the simulation variance of the estimates of the population parameters. And fourth, we calculate the probability similar to the previous method except treating the population parameters as fixed. These last two procedures are equivalent to using the customers' expected tastes as representative of their true tastes.

Results are given in Table 5 for model 4 (the model with the best fit). The results for the other models are qualitatively similar. The most prominent result is that conditioning on each customer's previous choices improves the forecasts for the last choice situation considerably. The average probability of the chosen alternative increases from 0.35 without conditioning (i.e., under the mixed logit probabilities) to 0.50 with conditioning that takes the entire conditional distribution into consideration and accounts for sampling variance in the estimated population parameters. For 260 of the 361 sampled customers (nearly three-quarters), the prediction of their last choice situation is better with conditioning than without, with the average probability rising by 0.25. For the other

customers, the conditioning makes the prediction in the last choice situations less accurate, with the average probability for these customers dropping by 0.11.

There are several reasons why the predicted probability after conditioning is not always greater. First, the choice experiments in our data set were constructed such that each situation would be fairly different from the other situations so as to obtain as much variation as possible. If the last situation involves new tradeoffs, the previous choices will not be useful and may in fact be detrimental to predicting the last choice. A more appropriate test might be to design a series of choice situations that elicited information on the relevant tradeoffs and then design an extra “hold-out” situation that is within the range of tradeoffs of the previous ones. Second, we did not include in our model all of the attributes of the alternatives that were presented to customers. In particular, we omitted attributes that did not enter significantly in the estimation of the population parameters. Some customers might respond to these omitted attributes, even though they are insignificant for the population as a whole. Insofar as the last choice situation involves tradeoffs of these attributes, the conditional distributions of tastes would be misleading since the relevant tastes are excluded. This explanation might imply that, if a mixed logit is going to be used for obtaining conditional densities for each customer, the researcher might want include attributes that could be important for some individuals even though they are insignificant for the population as a whole. Third, regardless of how the survey and model are designed, some customers might respond to choice situations in a quixotic manner, such that the tastes that are evidenced in previous choices are not applied by the customer in the last choice situation. Last, random factors can cause the probability for some customers to drop with conditioning even when the first three reasons do not. While at least one of these factors may be contributing to the lower choice probabilities for some of the customers in our sample, the gain in predictive accuracy for the customers with an increase in probability after conditioning is over twice as great as the loss in accuracy for those with a decrease, and the number of customers with a gain is almost three times as great as the number with a loss.

The four methods of calculating the expected probability give consistent results. The fourth (and easiest) method, which uses the customers' expected tastes based on the point estimate of the population parameters, gives the highest probability. Accounting for variance (by using the conditional distribution of  $\beta$  rather than its expected value and/or by using the estimated distribution of  $\theta$  rather than the point estimate) decreases the average probability. This result can have two interpretations. First, it could mean that the variances over-estimate the amount of uncertainty that actually exists. The true conditional distribution of  $\beta$  and the true sampling distribution of theta are perhaps smaller than our formulas imply, such that the expected beta is a more reliable estimate than these distributions suggest. Second, the decrease in the average probability when uncertainty is incorporated is the natural outcome of concavity in the probability formula. Within the concave region of  $P(\beta)$ ,  $E(P(\beta)) < P(E(\beta))$ , which is the relation we observe.

#### **IV. Discussion**

This paper demonstrates how the distribution of each sampled customer's tastes conditioned on the customer's observed choices are obtained with mixed logit models. While these conditional distributions can be useful in several ways, it is important to recognize the limitations of the concept. First, the use of conditional distributions in forecasting is limited to those customers whose previous choices are observed. Second, while the expected tastes of each customer can be used in cluster analysis and for other identification purposes, the researcher will often need to relate the expected tastes to observable demographics of the customers. Yet, these observable demographics of the customers could be entered directly into the mixed logit model such that the population parameters vary with the observed characteristics of the customers in the population. In fact, entering demographics into the mixed logit itself is more direct and more accessible to hypothesis testing than estimating a mixed logit without these characteristics, calculating expected tastes, and then doing cluster and other analyses on the expected tastes.

Given these issues, there are three main reasons that a researcher might benefit from calculating customers' conditional distributions of tastes. First, information on the previous choices for customers might be available, as with scanner data. In this case, conditioning on these previous choices allows for targeted marketing, as described by Rossi *et al.* (1996). Second, the demographic characteristics that differentiate customers with different tastes might be more evident through cluster analysis on the expected tastes than through specification testing in a mixed logit, simply because cluster analysis has its own unique way of identifying patterns. Third, examination of customers' expected tastes can often identify patterns that cannot be related to observed characteristics of customers but are nevertheless useful to know. For instance, knowing that a product or marketing campaign will appeal to a group of people because of their particular tastes is often sufficient, without identifying the people on the basis of their demographics. The conditional taste densities from COIT can fruitfully enter analyses that have these goals.

## References

- Algers, S., P. Bergstrom, M. Dahlberg, and J. Dillen, 1999, "Mixed Logit Estimation of the Value of Travel Time," working paper, Department of Economics, Uppsala University.
- Allenby, G., and P. Rossi, 1999, Marketing Models of Consumer Heterogeneity," *Journal of Econometrics*, Vol. 89, Nos. 1-2, pp. 57-78.
- Arora, N., G. Allenby, and J. Ginter, 1998, "A Hierarchical Bayes Model of Primary and Secondary Demand," *Marketing Science*, Vol. 17, No. 1, pp. 29-44.
- Bhat, C., 1998, "Accommodating Variations in Responsiveness to Level-of-Service Measures in Travel Mode Choice Modeling," *Transportation Research*, Vol. 32A, pp. 495-507.
- Bhat, C., 1999, "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model," working paper, Department of Civil Engineering, University of Texas, Austin.
- Boatwright, P., R. McCulloch, and P. Rossi, forthcoming, "Account-Level Modeling for Trade Promotions: An Application of a Constrained Parameter Hierarchical Model," *Journal of the American Statistical Association*.
- Brownstone and Train, 1999, "Forecasting New Product Penetration with Flexible Substitution Patterns," *Journal of Econometrics*, Vol. 89, Nos. 1-2, pp. 109-129.
- Bolduc, D., B. Fortin, and S. Gordon, 1996, "Multinomial Probit Estimation of Spatially Interdependent Choices: An Empirical Comparison of Two New Techniques," working paper, Department of Economics, University of Laval.
- Chib, S. and E. Greenberg, 1995, "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, Vol. 49, pp. 327-335.
- 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., 1992, "On Efficiency of Methods of Simulated Moments and Maximum Simulated Likelihood Estimation of Discrete Choice Models," *Econometric Theory*, Vol. 8, pp. 518-552.

Liu, J., 1994, "The Collapsed Gibbs Sampler in Bayesian Computations with Applications to a Gene Regulation Problem," *Journal of the American Statistical Association*, Vol. 89, No. 427, pp. 958-966.

McCulloch, R., and P. Rossi, 1994, "An Exact Likelihood Analysis of the Multinomial Probit Model," *Journal of Econometrics*, Vol. 64, pp. 207-240.

McFadden, D., 1973, "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, ed., *Frontiers in Econometrics*, Academic Press: New York.

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

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

Revelt, D., 1999, "Three Discrete Choice Random-Coefficients Papers and One Police-Crime Study," Ph.D. dissertation, Department of Economics, University of California, Berkeley.

Revelt, D. and K. Train, 1998, "Mixed Logit with Repeated Choices: Households' Choice of Appliance Efficiency Level," *Review of Economics and Statistics*, Vol. LXXX, No. 4, pp. 647-657.

Rossi, P., R. McCulloch, and G. Allenby, 1996, "On the Value of Household Purchase History Information in Target Marketing," *Marketing Science*, Vol. 15, pp. 321-340.

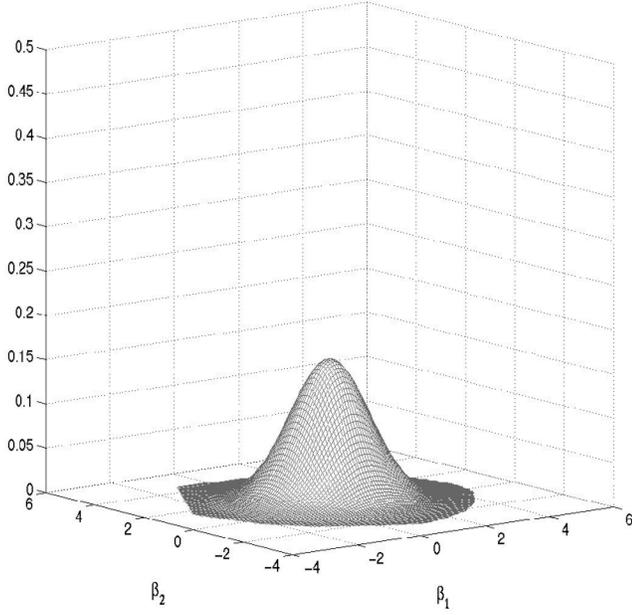
Ruud, P., 1996, "Approximation and Simulation of the Multinomial Probit Model: An Analysis of Covariance Matrix Estimation," working paper, Department of Economics, University of California, Berkeley.

Sawtooth Software, Inc., 1999, "The CBC/HB Module for Hierarchical Bayes Estimation," at Sawtooth Software's website, [www.sawtoothsoftware.com/Techabs.htm](http://www.sawtoothsoftware.com/Techabs.htm).

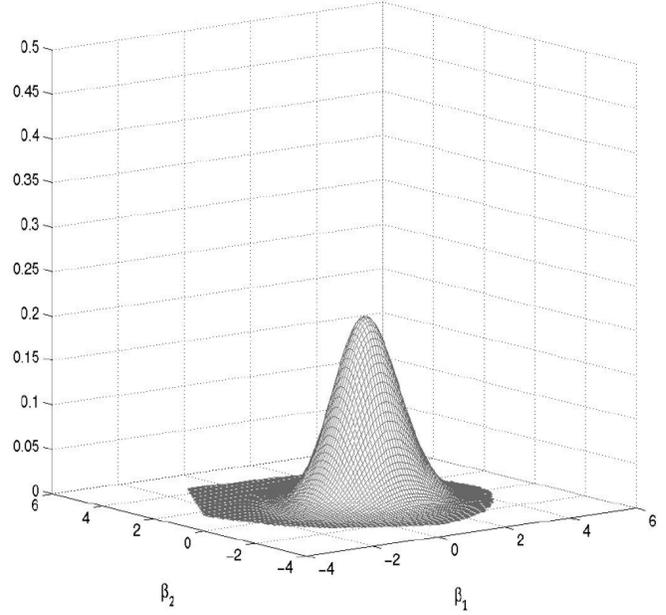
Train, K., 1999, "Halton Sequences for Mixed Logit," working paper, Department of Economics, University of California, Berkeley.

Figure 1: Conditional Distributions for One Respondent in Monte Carlo Experiment

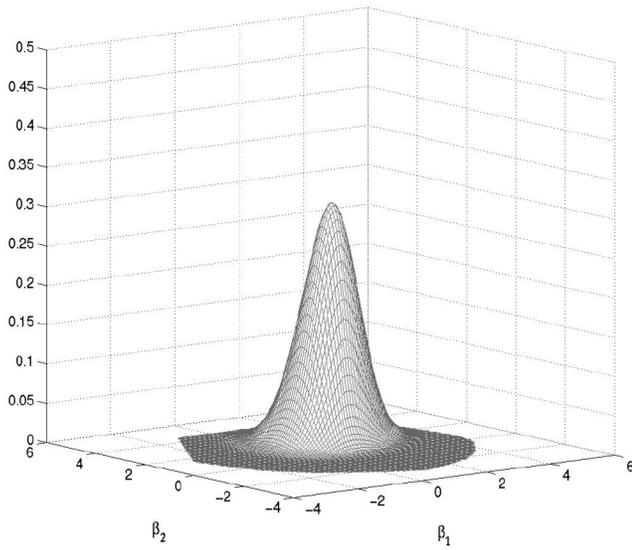
Unconditional Probability Distribution



$h(\beta|y)$  Probability Distribution With One Observed Choice



$h(\beta|y)$  Probability Distribution With Five Observed Choices



$h(\beta|y)$  Probability Distribution With Ten Observed Choices

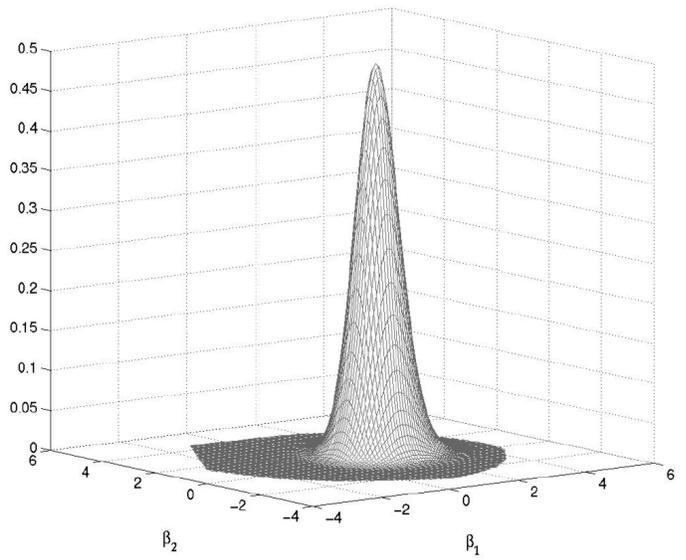


Table 1: Results of Monte Carlo Experiments  
(Averages over 50 constructed data sets.)

	100 Halton draws		1000 Halton draws		10,000 Halton draws	
	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$
1 choice situation						
Standard deviation of $E(\beta)$	0.398	0.418	0.411	0.416	0.413	0.416
Absolute diff. of $E(\beta)$ from $\beta$	0.726	0.718	0.726	0.718	0.726	0.718
10 choice situations						
Standard deviation of $E(\beta)$	0.802	0.812	0.822	0.824	0.826	0.826
Absolute diff. of $E(\beta)$ from $\beta$	0.444	0.450	0.442	0.448	0.422	0.448
20 choice situations						
Standard deviation of $E(\beta)$	0.866	0.867	0.890	0.883	0.894	0.886
Absolute diff. of $E(\beta)$ from $\beta$	0.358	0.353	0.354	0.350	0.354	0.350
50 choice situations						
Standard deviation of $E(\beta)$	0.917	0.927	0.946	0.950	0.951	0.953
Absolute diff. of $E(\beta)$ from $\beta$	0.255	0.254	0.244	0.243	0.243	0.243

Table 2: Mixed Logit Models of Customers' Choice Among Energy Supplier

Variable	Model 1 Normals	Model 2 Uniforms	Model 3 Triangulars	Model 4 Normals and log- normals
Price, in cents per kWh, for fixed rates (zero for seasonal and time-of-day rates.) Fixed coefficient	-0.8574 (0.0488)	-0.8861 (0.0493)	-0.8582 (0.0490)	-0.8827 (0.0497)
Length of contract, in years <i>m</i>	-0.1833 (0.0289)	-0.2251 (0.0310)	-0.1898 (0.0310)	-0.2125 (0.0261)
<i>s</i>	0.3786 (0.0291)	0.6127 (0.0416)	0.8951 (0.0661)	0.3865 (0.0278)
1 if supplier is local energy utility, 0 otherwise.* <i>m</i>	2.0977 (0.1370)	2.2455 (0.1435)	2.1384 (0.1411)	2.2297 (0.1266)
<i>s</i>	1.5585 (0.1264)	2.7011 (0.2184)	3.7303 (0.2965)	1.7514 (0.1371)
1 if supplier is a well-known company (other than local utility), 0 otherwise.* <i>m</i>	1.5247 (0.1018)	1.5122 (0.1013)	1.5214 (0.1015)	1.5906 (0.0999)
<i>s</i>	0.9520 (0.0998)	1.8318 (0.1794)	2.3856 (0.2424)	0.9621 (0.0977)
Dummy for time-of-day rates: 11c/kWh 8AM-8PM and 5c/kWh 8PM-8AM.** <i>m</i>	-8.2857 (0.4577)	-8.3759 (0.4473)	-8.2675 (0.4638)	2.1328 (0.0543)
<i>s</i>	2.5742 (0.1676)	4.6013 (0.3212)	6.2933 (0.4696)	0.4113 (0.0397)
Dummy for seasonal rates: 10c/kWh in summer, 8c/kWh in winter, and 6c/kWh in spring and fall.** <i>m</i>	-8.5303 (0.4468)	-8.6923 (0.4566)	-8.5590 (0.4496)	2.1577 (0.0509)
<i>s</i>	2.1259 (0.1604)	3.4803 (0.2478)	5.0523 (0.3734)	0.2812 (0.0217)
Log-likelihood at convergence	-3646.5187	-3638.1634	-3644.1433	-3618.9175

Standard errors of the estimates are in parentheses.

Number of customers: 361. Number of choice situations: 4308.

\* 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 3: Expected  $\beta$  Conditional on Customer's Choices and the Point Estimates of the Population Parameters

	Model 1	Model 2	Model 3	Model 4
Contract length				
Mean	-0.2028	-0.2079	-0.2011	-0.2149
Standard deviation	0.3175	0.2965	0.3058	0.3262
Local Utility				
Mean	2.1205	2.1477	2.1208	2.2146
Standard deviation	1.2472	1.2776	1.2288	1.3836
Known company				
Mean	1.5360	1.5357	1.5326	1.5997
Standard deviation	0.6676	0.7755	0.6888	0.6818
TOD rate				
Mean	-8.3194	-8.5664	-8.3367	9.2584
Standard deviation	2.2725	2.3010	2.2549	3.1051
Seasonal rate				
Mean	-8.6394	-8.8234	-8.6374	9.1344
Standard deviation	1.7072	1.7001	1.7065	2.0560

Table 4: Expected  $\beta$  Conditional on Customer's Choices and the Estimated Sampling Distribution of the Population Parameters

	Model 1	Model 2	Model 3	Model 4
Price				
Mean	-0.8753	-9.0237	-0.8744	-0.8836
Standard deviation	0.5461	0.5571	0.5584	0.0922
Contract length				
Mean	-0.2004	-0.2100	-0.1976	-0.2111
Standard deviation	0.3655	0.3513	0.3600	0.3720
Local Utility				
Mean	2.1121	2.1597	2.1210	2.1921
Standard deviation	1.5312	1.5611	1.5148	1.6815
Known company				
Mean	1.5413	1.5514	1.5402	1.5832
Standard deviation	0.9364	1.0520	0.9649	0.9527
TOD rate				
Mean	-9.1615	-9.4472	-9.2506	9.0216
Standard deviation	2.4309	2.3378	2.3666	3.8785
Seasonal rate				
Mean	-9.4528	-9.6605	-9.5213	8.9408
Standard deviation	1.9222	1.7841	1.8591	2.5615

Table 5: Prediction of Last Choice Situation for Each Customer for Model 4

	Using point estimate of $\theta$	Using estimated distribution of $\theta$
Average probability for chosen alternative		
E(P(last choice   $\beta$ ))	0.5213	0.5041
P(last choice   E( $\beta$ ))	0.5565	0.5487
Mixed logit (unconditional)	0.3530	0.3517
Number of customers whose probability rises with conditioning, relative to mixed logit probability.		
E(P(last choice   $\beta$ ))	266	260
P(last choice   E( $\beta$ ))	268	264
Average rise in probability for customers whose probability rises.		
E(P(last choice   $\beta$ ))	0.2725	0.2576
P(last choice   E( $\beta$ ))	0.3240	0.3204
Number of customers whose probability drops with conditioning, relative to mixed logit probability.		
E(P(last choice   $\beta$ ))	95	101
P(last choice   E( $\beta$ ))	93	97
Average drop in probability for customers whose probability drops.		
E(P(last choice   $\beta$ ))	0.1235	0.1182
P(last choice   E( $\beta$ ))	0.1436	0.1391