Mixed Logit with Bounded Distributions of Correlated Partworths*

Kenneth Train and Garrett Sonnier
University of California, Berkeley and Los Angeles

August 20, 2004

Abstract

The use of a joint normal distribution for partworths is computationally attractive, particularly with Bayesian MCMC procedures, and yet is unrealistic for any attribute whose partworth is logically bounded (e.g., is necessarily positive or cannot be unboundedly large). A mixed logit is specified with partworths that are transformations of normally distributed terms, where the transformation induces bounds; examples include censored normals, log-normals, and S_B distributions which are bounded on both sides. The model retains the computational advantages of joint normals while providing greater flexibility for the distributions of correlated partworths. The method is applied to data on customers' choice among vehicles in stated choice experiments. The flexibility that the transformations allow is found to greatly improve the model, both in terms of fit and plausibility, without appreciably increasing the computational burden.

1 Introduction

Mixed logit is a flexible discrete choice model that incorporates random variation in partworths.[1] McFadden and Train (2000) show that mixed

---

[1]The partworth of an attribute is the coefficient of the attribute in the utility function. The term is used extensively in marketing, and we adopt it here because it is more succinct than "utility coefficient" and more specific than "coefficient."

logit can approximate any random utility choice model to any degree of accuracy through appropriate specification of distributions of the partworths. Procedures for estimating mixed logits have been developed within both the classical (e.g., Revelt and Train, 1998, Brownstone and Train, 1999) and Bayesian (Allenby, 1997; Sawtooth Software, 1999) traditions.

Mixed logit models have been used for environmental analysis in numerous contexts, including: households' response to rebates on energy-efficient appliances (Revelt and Train, 1998); the impact of fish stock, which is affected by water quality, on anglers' choice of fishing site (Train, 1998); the demand for wind, hydro and other forms of renewable power generation (Goett *et al.*, 2000); and consumers' willingness to pay for water service improvements (Hensher *et al.*, 2004). Consumers' choice of vehicle, which is the application in the current paper, is particularly important for environmental analysis since energy consumption and emissions are largely dependent on this choice. Mixed logits of vehicle choice have been previously estimated by Brownstone and Train (1999), Brownstone *et al.* (2000), and Train and Winston (2004).

The distribution of partworths is critical in any application. Normal distributions are relatively easy to implement in both the classical and Bayesian methods. However, since the normal is unbounded on each side of zero, its use in many setting is inappropriate. A normal distribution for a price coefficient implies that some share of the population actually prefer higher prices. Also, since the normal distribution overlaps zero, its use for a price coefficient can preclude the calculation of willingness-to-pay statistics: The willingness-to-pay for an attribute is the partworth of that attribute divided by the price coefficient. Since division by zero is undefined, and division by a number arbitrarily close to zero gives an arbitrarily large result, the mean and variance of willingness-to-pay need not exist when the price coefficient is normal.[2] A normal distribution is also inappropriate for the partworth of a desirable attribute that is valued (or, at worst, ignored) by all customers or an undesirable attribute that is disliked (or ignored) by all customers. Similarly, when an attribute consists of various levels, the partworth for

---

[2]The applications cited in the previous paragraph specify the price coefficient to be fixed or log-normal in order to avoid this issue. See, e.g., Merrill (1928), Geary (1930), Fieller (1932), Marsaglia (1965), and Hinkley (1969) on the distribution of a ratio of two normally distributed terms.

each higher level must logically be no smaller than the partworth for each lower level; normal distributions do not embody this requirement.

Bounded distributions can and have been used in mixed logits estimated by both the classical and Bayesian procedures (e.g., Bhat, 1998, 2000; Revelt and Train, 1998; Train, 1998; Revelt, 1999; Boatwright *et al.*, 1999; Brownstone and Train, 2000; Johnson, 2000; Train 2001). However, each estimation procedure, while feasible with bounded distributions, entails numerical difficulties that are intrinsic to its form, as described and illustrated by Train (2001). In particular: Classical procedures handle triangular, truncated normal, and similarly bounded distributions easily while Bayesian procedures are relatively slow with these distributions. On the other hand, fully correlated partworths are difficult to handle in classical procedures due to the proliferation of parameters, while the Bayesian procedures accommodate these correlations readily. Obtaining partworths that are bounded *and* correlated has been relatively difficult with either procedure.

Bayesian procedures operate effectively with normals because of the convenient posteriors that arise with normals. In this paper, we build upon the observation in Train (2001) that the Bayesian procedures operate as effectively with log-normals as normals because the log-normal is simply a transformation of the normal that does not entail any other parameters. This concept is expanded by using other transformations that provide various types of bounded distributions. These transformations can operate on correlated normals to provide correlated partworths with bounded distributions. The numerical advantages of the Bayesian procedures with correlated normals are retained while having partworths whose distributions are bounded.

Many useful distributions can be obtained as transformations of normals. Let scalar β be normally distributed with mean b and variance ω. Bounded distributions are obtained through the following kinds of transformations. These transformations are weakly monotonic (non-decreasing in β) and depend only on β without utilizing b and ω.

- Log-normal. The transformation is $c = exp(\beta)$. The distribution is bounded below by zero. It is useful for the partworths of attributes that are liked by all customers. The sign is reversed for undesirable attributes, such as a price variable, such that the partworth is necessarily negative.

3

- Normal censored from below at zero. The transformation is $c = max(0, \beta)$. There is a mass at zero, with the density above zero being the same as the normal density of β. The share at zero is $\Phi(-b/\omega)$, where Φ is the standard normal cumulative distribution. This transformation is useful for partworths of an attribute that some customers do not care about (i.e., are indifferent to its presence and simply ignore) and other customers find desirable. The estimation of b and ω determines the share massed at zero and the share distributed above zero.

- Johnson's (1949) S_B distribution.[3] The transformation $c = exp(\beta)/(1 + exp(\beta))$ creates a partworth that is distributed between zero and one, with mean, variance and shape determined by the mean and variance of β.[4] For a distribution that has support from ℓ to u, the transformation is $c = \ell + (u - \ell) \cdot (exp(\beta)/(1 + exp(\beta)))$. The S_B distribution is useful for a variety of purposes. S_B densities can be shaped like log-normals but with an upper bound and with thinner tails below the bound. S_B densities are more flexible than log-normals: they can be shaped like a plateau with a fairly flat area between drop-offs on each side (as in Figure 2 for our application) and can even be bi-modal. When a lower bound other than zero is specified, the distribution is useful for an attribute that some people like and others dislike but for which there is a limit for how much the person values having or avoiding the attribute.

For multiple partworths, β is generalized to be a vector with length equal to the number of partworths, with mean vector b and variance matrix Ω. Each partworth is defined as a transformation of the corresponding element of β. The covariance among the elements of β induces covariance among the partworths. As such, the procedure allows for correlated partworths under any combination of the above distributions.

Numerous authors have implemented log-normal distributions within mixed logit, though usually without allowing full correlation; see, e.g., Bhat

---

[3]See also Johnson and Kotz, 1970, p. 23.

[4]As Johnson and Kotz note, the formulas for the moments are very complex. We calculate them through simulation as described section 4. The median is $1/(1 + exp(b/\sqrt{\omega}))$.

(1998, 2000), Train (1998), and Revelt and Train (1998).[5] R. Johnson (2000) examined censored normals and found that they provided more reasonable results and better fit than uncensored normals in his application. The use of the S_B distribution seems to be new. We will investigate its usefulness in the context of our application.

The computational advantage of the method rests on the simplicity of the posteriors on b and Ω that arise, as described in the next section, when the transformation of β does not depend on b and Ω. Transformations that depend on b and Ω can be useful in some settings, but do not provide the same simplicity. For example, truncated normals cannot be accommodated within our procedure because the necessary transformation entails b and Ω rather than depending only on β.[6] Since b and Ω affect the utility of each customer through this transformation, the posteriors for b and Ω conditional on the β's depend on the choices of the respondents and no longer have the convenient form that we utilize. Boatwright *et al.* (1999) provide MCMC methods for truncated normals. These methods, which can be generalized to essentially any bounded distribution, are considerably more difficult and slower computationally than those we utilize in this paper. The question for the researcher for a particular application is whether transformations of normals that do not depend on b and Ω can adequately represent the relevant distributions of partworths. If so, the simple methods in this paper can be exploited; if not, the methods of Boatwright *et al.* can be used.[7]

While the estimation procedure that we describe is Bayesian, the results can be interpreted from either a Bayesian or classical perspective. Bayesian interpretation is of course straightforward since the procedure itself is Bayesian. Classical interpretation is less well recognized. The Bernstein-von Mises theorem (see, e.g., Train, 2003, for a discussion with historical references) establishes that, under conditions that are maintained in our specification, the mean of the Bayesian posterior is a classical estimator that

---

[5]Experience indicates that the parameters of log-normal distributions are hard to estimate with classical procedures, due to the fact that the log-likelihood surface is highly non-quadratic. The Bayesian procedure avoids this difficulty.

[6]E.g., a one-dimensional normal truncated below at zero is created as $c = \Phi^{-1}(m(1 - z) + z) \cdot \omega + b$ where $z = \Phi((\beta - b)/\omega)$ and $m = \Phi(-b/\omega)$.

[7]Classical estimation procedures accommodate truncated normals as readily as normals; see, e.g., Revelt (1999.) However, as stated above, classical procedure have difficulty dealing with correlated partworths due to the proliferation of parameters.

is asymptotically equivalent to the maximum likelihood estimator. The theorem also establishes that the covariance of the posterior is the asymptotic sampling covariance of this estimator. The results from the Bayesian procedures can therefore be interpreted by a classical researcher in the same way that the researcher would interpret estimates obtained by maximum likelihood. To facilitate this interpretation, we present our results in the format that is standard for classically estimated models, namely by reporting the parameter estimates (which are the posterior means) and their standard errors (the posterior standard deviations).

In section 2, we describe Bayesian estimation of a mixed logit with normally distributed partworths. We then show in section 3 how this procedure is changed to accommodate transformations of the normal. We apply the method in section 4 to data on customers' choice among vehicles.

2 Mixed logit with normally distributed partworths

The behavioral derivation of mixed logit with repeated choices is given by Revelt and Train (1998) and Train (1998) for general distributions of partworths. The Bayesian procedure for estimating the model with normally distributed partworths was developed by Allenby (1997) and implemented by Sawtooth Software (1999).[8] We give the derivation and Bayesian procedure under normally distributed partworths in this section. The generalization to transformation of normals is described in the following section.

Person n faces a choice among J alternatives in each of T time periods. J can be as small as 2, and T can be as small as 1. The person's utility from alternative j in period t is

$$U_{njt} = \beta'_n x_{njt} + \varepsilon_{njt},$$

where $\varepsilon_{njt} \sim$ iid extreme value and $\beta_n \sim N(b, \Omega)$. The vectors of variables x_{njt} and partworths β_n have length K. Person n chooses alternative i in period t if $U_{nit} > U_{njt} \ \forall j \neq i$. Denote the person's chosen alternative in period t as y_{nt}, the person's sequence of choices over the T time periods as $y_n = \langle y_{n1}, \ldots, y_{nT} \rangle$, and the set of $y_n \forall n$ as Y. Conditional on β_n, the

---

[8]Related methods for probit models were developed by Albert and Chib (1993), McColluch and Rossi (1994), and Allenby and Rossi (1999). Bayesian procedures for non-mixed logits are discussed by Koop and Poirier (1993, 1996) and Poirier (1994, 1996).

probability of person n's sequence of choices is the product of standard logit formulas:

$$L(y_n \mid \beta_n) = \prod_t \frac{e^{\beta_n' x_{n y_{nt} t}}}{\sum_j e^{\beta_n' x_{njt}}}.$$

The unconditional probability is the integral of $L(y_n \mid \beta_n)$ over all values of β_n weighted by the density of β_n:

$$P_n(y_n \mid b, \Omega) = \int L(y_n \mid \beta_n) g(\beta_n \mid b, \Omega) d\beta_n. \tag{1}$$

where $g(\cdot)$ is the multivariate normal density. This unconditional probability is called the mixed logit choice probability, since it is a product of logits mixed over a density of partworths.

For Bayesian analysis, it is necessary to specify the priors on the model parameters b, Ω, and $\beta_n \forall n$. Since we have already specified β_n to be normal with mean b and variance Ω, the prior on each β_n is proportional to this density times the prior on b and Ω.[9] We specify the prior on b to be a diffuse normal, denoted $N(b \mid 0, \Theta)$, which has zero mean and variance Θ sufficiently large that the density is effectively flat from a numerical perspective. The advantage of a normal prior on b is that it provides a conditional posterior on b (*i.e.*, conditional on $\beta_n \forall n$ and Ω) that is normal and hence easy to draw from, while the large variance assures that the prior has minimal (effectively no) influence on the posterior. The standard diffuse prior on Ω is inverted Wishart with K degrees of freedom and parameter KI where I is the K-dimensional identity matrix. This density is denoted $IW(\Omega \mid K, KI)$. It provides a conditional posterior on Ω that is IW and hence easy to draw from. The joint posterior on $\beta_n \forall n$, b and Ω is

$$\Lambda(\beta_n \forall n, b, \Omega \mid Y) \propto \prod_n L(y_n \mid \beta_n) \cdot g(\beta_n \mid b, \Omega) \cdot N(b \mid 0, \Theta) \cdot IW(\Omega \mid K, KI).$$

---

[9]Several terms have been used for these parameters. (1) Often, b and Ω are called population parameters that describe the distribution of customer-level β_n's in the population. With this usage, the distribution $g(\beta_n \mid b, \Omega)$ is interpretaed as the actual distribution of partworths in the population. (2) In Bayesian analyses especially, b and Ω are often called hyper-parameters, since the prior on each β_n depends on b and Ω which themselves have priors. Under this usage, $g(\beta_n \mid b, \Omega)$ is interpreted as an aspect of the researcher's prior information about β_n. (3) Sometimes, the β_n's are called nuisance parameters, to reflect the concept that they are incorporated into the analysis simply (under this usage) to facilitate estimation of b and Ω.

Information about the posterior is obtained by taking draws from the posterior and calculating relevant statistics, such as moments, over these draws. Gibbs sampling is used to facilitate the taking of draws (see Casella and George, 1992, for a general explanation of Gibbs sampling.) In particular, draws are taken sequentially from the conditional posterior of each parameter given the previous draw of the other parameters. The sequence of draws from the conditional posteriors converges to draws from the joint posterior.

The conditional posterior distributions in this model are especially convenient. Given β and Ω, the posterior on b is $N(\bar{\beta}, \Omega/N)$ with $\bar{\beta} = (1/N)\sum \beta_n$.[10] This distribution is easy to draw from: A draw of b is created as $\tilde{b} = \bar{\beta} + L\eta$, where L is the lower-triangular Choleski factor of Ω/N and η is K-dimensional vector of independent draws from a standard normal density. A draw of the vector b requires only K draws from a random number generator, K means over N terms each, and a few arithmetic calculations. It takes a tiny fraction of a second.

Given b and β, the conditional posterior of Ω is $IW(\Omega \mid K+N, KI+N\bar{V})$, where $\bar{V} = (1/N)\sum(\beta_n - b)(\beta_n - b)'$. Draws from the inverted Wishart are easily obtained. Take $K+N$ draws of K-dimensional vectors of iid standard normal deviates. Calculate the Choleski factor, M, of $(KI+N\bar{V})^{-1}$. Create $S = \sum_r (M\eta_r)(M\eta_r)'$. Then $\tilde{\Omega} = S^{-1}$ is a draw. This calculation is also extremely fast.

The only computationally intensive part is drawing $\beta_n \forall n$. Given b and Ω, the conditional posterior for β_n is proportional to $L(y_n \mid \beta_n)g(\beta_n \mid b, \Omega)$. The Metropolis-Hasting (M-H) algorithm is used to take draws from this distribution. (See Chib and Greenberg, 1995, for a general explanation of the M-H algorithm.) The previous draw is labeled β_n^0 and the new one is β_n^1. The new draw is obtained as follows.

1. Calculate $d = \sigma L\eta$, where η is a draw of a K-dimensional vector of iid standard normal deviates, L is the Choleski factor of Ω, and σ is a scalar that the researcher sets in a way to be described below.

2. Create a "trial" value of β_n^1 as $\tilde{\beta}_n^1 = \beta_n^0 + d$.

3. Evaluate the posterior at this trial value and compare it with the

---

[10]More precisely, the posterior on b approaches $N(\bar{\beta}, \Omega/N)$ as the variance of the prior on b rises without bound. This variance is specified to be sufficiently high such that the posterior is numerically indistinguishable from $N(\bar{\beta}, \Omega/N)$.

posterior at the previous draw. That is, calculate the ratio

$$R = \frac{L(y_n \mid \tilde{\beta}_n^1)g(\tilde{\beta}_n^1 \mid b, \Omega)}{L(y_n \mid \beta_n^0)g(\beta_n^0 \mid b, \Omega)}.$$

4. Take a draw from a standard uniform and label the draw μ.

5. If $\mu < R$, accept the trial draw. Otherwise, reject the trial draw and use the previous draw as the current draw. That is, set $\beta_n^1 = \tilde{\beta}_n^1$ if $\mu < R$ and set $\beta_n^1 = \beta_n^0$ otherwise.

A sequence of draws taken by the M-H algorithm converges to draws from the target distribution, in this case the conditional posterior. One draw of β_n within the M-H algorithm for each person is taken in each iteration of the Gibbs sampling over b, Ω, and $\beta_n \forall n$. Movement to convergence in the M-H algorithm for each person and in the overall Gibbs sampling is thereby attained simultaneously. In our application we used 30,000 iterations for "burn-in" (i.e., movement toward convergence) followed by 20,000 iterations, of which the draws in every 10-th iteration were retained. (Run-times were only 1.5 hours, even with this large number of iterations.) The 2000 retained draws are used to conduct inference. For example, the average of these draws is the simulated mean of the posterior, which, from a classical perspective, is the estimate of the parameters. The standard deviation of the draws is the simulated standard deviation of the posterior and the classicists' standard error of the estimate.

The value of σ in step (1) affects the acceptance rate in the M-H algorithm. For smaller values of σ, the acceptance rate is generally higher but the jumps between draws is smaller so that more draws are needed for the algorithm to reach convergence and, once at convergence, to traverse the conditional posterior. Gelman *et al.* (1995) found that the optimal acceptance rate is .4 for $K = 1$ and decreases to .23 for higher dimensions. They recommend an adaptive acceptance rate to achieve optimality. This adaptation is implemented by changing σ in each iteration of the Gibbs sampling based on the acceptance rate among the N trial draws of $\beta_n \forall n$ in the previous iteration. Following Sawtooth Software (1999) and in accordance with the optimal rate found by Gelman *et al.*, we lower σ if the acceptance rate is below .3 and raise it if the rate is above .3.

3   Transformation of normals

Denote the partworths of person n as c_n, which is a vector with the same length as β_n. The partworths are defined by $c_n = T(\beta_n)$, where T is a transformation that depends only on β_n and is weakly monotonic (such that $\partial c_n^k / \partial \beta_n^k \geq 0$ for each element k of c_n and β_n). The distribution of c_n is determined by the transformation.

Little is changed in the estimation procedure by this transformation. Normally distributed β_n's are drawn as before but then transformed to c_n's when they enter utility. Utility is specified as

$$U_{njt} = T(\beta_n)' x_{njt} + \varepsilon_{njt}.$$

The probability of the person's choice sequence given β_n is

$$L(y_n \mid \beta_n) = \prod_t \frac{e^{T(\beta_n)' x_{ny_{nt}t}}}{\sum_j e^{T(\beta_n)' x_{njt}}}.$$

This probability is used in step 3 of the M-H algorithm instead of the probability based on untransformed β_n. The rest of the procedure is same.

In this set-up, β_n can be considered a latent value that determines the person's partworths. This latent value is normally distributed, with mean b and covariance Ω. The conditional posteriors for b and Ω are the same as before, and the conditional posterior of β_n changes only by the transformation that occurs in utility in the logit formula. The advantages of normal distributions within a Bayesian procedure are maintained while allowing the partworths to take other distributions. For any given value of β_n, the partworths c_n are calculated, and the distribution of β_n induces a distribution of c_n.

4   Application

We present an analysis of customers' choice among gas, electric, and hybrid gas-electric vehicles. We apply the methods described above to investigate the use of various bounded distributions of partworths, all of which involve transformations of normals.

Stated choice experiments were designed to elicit customers' choice among gas, electric, and hybrid vehicles under various prices, operating costs, and

other attributes. The experiments were conducted as part of a survey of vehicle owners in California. The state of California is particularly relevant for electric and hybrid vehicles because the state's Air Resources Board has implemented, and is continually revising, regulations that promote these vehicles. Survey respondents were contacted through random-digit dialing throughout the state. Respondents intending to purchase a new vehicle within the next three years were asked to participate in the study. Those who were willing to participate in the study were sent a packet of materials, including information sheets that described the new vehicles and the choice experiments. The respondents were later called to go over the information, obtain their choices in the experiments, and ask demographic and other questions. A total of 500 respondents were obtained.

4.1 Choice experiments

Each choice experiment consisted of three vehicles. For each vehicle, the following information was listed:

- Body type, such as midsize car

- Engine type: gasoline, electric, or hybrid

- Purchase price

- Operating cost in dollars per month

- Performance

- Range: miles between refueling/recharging.

The respondent was asked to consider the attributes of all three vehicles and state which one he/she would buy if making the purchase today. Each respondent was presented with 15 choice experiments, with each experiment containing different vehicles with different attributes.

The choice experiments were designed to provide as wide variation in each attribute, and as little covariance among attributes, as possible while maintaining plausibility. Ten body types were considered in the experiments: mini car, small car, midsize car, large car, small SUV, midsize SUV, large SUV, compact pick-up, large pick-up, mini-van. Respondents were given examples of vehicles with that body type.

Each vehicle in each experiment was listed as being gasoline, electric, or hybrid. In any one experiment, the respondent might face a choice among two electric and one hybrid vehicle, among three gasoline vehicles, or any other combination.

The purchase price and operating cost of each vehicle were chosen randomly from a range of possible prices and operating costs.

The performance of each vehicle was described in terms of top speed and the number of seconds required to go from zero to 60 mph. These two performance measures were not varied independently, since respondents know that they are linked. Rather, three performance levels were specified, and each vehicle was randomly assigned one of the three performance levels. The three levels were: (1) Top speed of 120 mph, and 8 seconds to reach 60 mph. This level is called "high" performance in the discussions below; however, the respondent did not see the word "high". The respondent saw the numbers for top speed and seconds to 60. (2) Top speed of 100 mph, and 12 seconds to reach 60 mph. This level is called "mid" in the discussions below. (3) Top speed of 80 mph, and 16 seconds to reach 60 mph. This level is called "low." The performance for gas and hybrid vehicles was randomly chosen from all three levels. The performance for electric vehicles was randomly chosen from the mid and low levels.

For the miles between refueling/recharging, a range of miles was given for each vehicle. The miles between refueling was given as "300-500" miles for gas vehicles and "400-700" miles for hybrid vehicles in all the experiments. A constant level was used for these vehicles because the study did not intend to estimate the value of increasing the range of vehicles that are refueled conventionally. The goal was to estimate the value to customers of increasing the range of electric vehicles. The range for gas and hybrid vehicles was stated so that the experiments would not place undue emphasis on the electric vehicle range in the eyes of the respondent. (If the range of electric vehicles was stated in the experiment but not the range of gas or hybrid vehicles, then respondents might be induced to place more importance on this aspect of electric vehicles than they otherwise would.) For electric vehicles, the possible ranges included every 10 mile increment starting with 60-70 and going up to 190-200. The range for each electric vehicles in the choice experiments was chosen randomly from these levels.

4.2 Models

Price, operating cost, and range are linearized, such that their partworths represent the value of a one-unit increment. The negative of price and operating cost are entered, such that their partworths are expected to be positive (so that log-normal distributions, which have positive support, can be used.) For performance, the low level is taken as the base and the medium and high levels are represented in increments. That is, two variables are entered for performance: a dummy indicating that the vehicle has either medium or high performance, and a dummy indicating that the vehicle has high performance. For engine types, gas is taken as the base, such that the partworths of the electric and hybrid vehicles are the value of these engine types relative to that of a gas engine. Similarly, the large car is taken as the base body type, with the partworths for the others representing value relative to the large car.

We start with a model in which all the partworths are distributed jointly normal $N(b, \Omega)$. As stated above, 2000 draws of b, Ω and β_n $\forall n$ are obtained from their posterior distribution. The means of the 2000 draws of b and of the diagonal elements of Ω are given in Table 1. (The partworths for body types are omitted from this and subsequent tables to save space and because they contain relatively less interpretable content.) From a classical perspective, these figures represent the estimated mean and variance of the β_n's in the population. And since the β_n's are untransformed, the figures also represent the mean and variance of partworths in the population. The standard deviations of the draws of b and the diagonal elements of Ω are given in parentheses. From a classical perspective, these are the standard errors of the estimated mean and variance of β_n's in the population. For example, for our first model, the partworth associated with range is normally distributed in the population with an estimated mean of 1.213 and estimated variance of 4.050. These estimates imply that 73 percent of the population have positive partworth for range while the other 27 percent have a negative partworth. (These negative partworths for range are of course implausible and the basis for our exploration of other distributions below.) The standard error on the estimated mean is .2442, which gives a t-statistic of 4.97, implying that the mean is significantly different from zero. Similarly, the standard error on the estimated variance is .7190, for a t-statistic

of 5.63, implying that the variance is also significant (that is, the hypothesis of no variance can be rejected.) The classical log-likelihood of the model is given at the bottom of the Table. This value is the log-likelihood of the observed choices calculated at the estimated values of b and Ω (that is, at the mean of the draws of b and Ω.) We give this statistic, which is not used in Bayesian inference but is standard in classical analysis, to emphasize the ability of the Bayesian procedures to provide results that are interpretable classically.

The mean partworth associated with moving from low to medium or high performance (0.8056) is greater than that for moving from medium to high performance (0.3092), which is consistent with decreasing marginal utility of performance. The estimates for the other partworths are self-explanatory.

The model is implausible in several regards. The estimates imply that 22 percent of the population prefer higher prices. The existence of price coefficients with the wrong sign renders the model un-useable for calculation of willingness to pay and other welfare measures. The estimates also imply that 37 percent of people prefer higher operating costs, 27 percent prefer electric vehicles with shorter ranges over those that can be driven further between recharging, 24 percent prefer low performance over medium or high performance, and 40 percent prefer medium performance over high performance. Also, for any arbitrarily large value in either direction, the model implies that some people have a partworth in excess of this value. The model therefore implies that some people would buy a vehicle that is worse in all features than any other currently-offered vehicle. These implications are the result of using normal distributions for partworths when actual partworths have known signs and limited magnitudes.

Table 2 gives the correlation among the partworths implied by the estimate of Ω. The largest correlation (in magnitude) is between the partworths for range and electric vehicle: the correlation of -.64 implies that people who are concerned about the range of an electric vehicle tend not to like electric vehicles at any range. This result is questionable. It is probably true that people who are concerned about range tend not to like electric vehicles, since electric vehicles generally have short ranges. However, the range of the electric vehicle is explicitly included in the experiments and the model. The negative correlation in partworths therefore implies that people who care about range tend to not like electric vehicles for reasons beyond range. It is

not clear what these reasons might be or why they relate to concern about range.

We estimate two models under other distributional assumptions that are easy to accommodate within our estimation procedure. For the first of these two models, the two performance variables are given normal distributions that are censored from below at zero. With this distribution, a share of the population is completely unconcerned about performance (i.e., have a partworth of zero) while the other share of the population places a positive value of higher performance with this value varying over people. The coefficients of price, operating cost, and range are given log-normal distributions (with the negative of price and operating cost entering the model.)

The mean and standard deviation of b and the diagonal elements of Ω are given in Table 3. Note that the log-likelihood of this model is considerably higher than that for the model with all normals: -6171.5 compared to -6835.5. As stated above, b and Ω are the mean and variance of the β_n in the population, which are transformed to obtain the partworths. The distribution of partworths is obtained through simulation on the estimated values of b and Ω. In particular, draws of β_n are taken from a normal distribution with mean equal to the estimated value of b and variance equal to the estimated value of Ω. Each draw of β_n is then transformed to obtain a draw of partworths.[11] The mean and variance of these partworths are given in the latter columns of Table 3. The specification of the distributions assures that no one in the population dislikes (i.e., has a strictly negative partworth for) price reductions, operating cost reductions, range improvements, or either of the two performance improvements.[12] The mean partworths are smaller in this model than in the model with all normal distribution.[13] This difference

---

[11]An alternative procedure, which is more consistent with Bayesian concepts and less consistent with classical concepts, is to retain the draw of the partworths (i.e., the transformation of the draw of β_n) for each person in each iteration of the MCMC procedure after convergence and calculate statistics such as means and variances over these draws.

[12]The mean β_n is negative for many of these attributes, even though the partworths themselves are positive. For log-normal distributions, β_n is exponentiated such that the partworth is positive even if β_n is negative. In this case, a negative mean for β_n implies that the median partworth is between zero and one. Similarly, if the partworth is a normal censored at zero, a negative mean for β_n implies than more than half of the population does not care about the attribute.

[13]Even though the means drop, the ratios of means move in both directions relative to those with all normal distributions.

15

is evidence of how the use of normal distribution can distort the estimated mean partworths. In particular: For a desirable attribute, the normal distribution gives an implausible negative sign for some share of customers; in estimation, the distribution is moved "up" to avoid the poor fit that these implausible values imply. With distributions that do not contain implausible values, the estimation procedure is not distorted to avoid implausible values.

The estimates imply that 51 percent of the population do not care about an improvement from low to mid-level performance, and 69 percent of the population do not care about an improvement from mid-level to high performance. These shares seem larger than expected (at least what we expected.) However, this result might simply indicate that the other attributes that were included in the choice experiments are more important to a large share of respondents, such that the partworth for performance appears to be zero for these respondents. If attributes that were considered less important than performance had been included in the experiments, with variation only over performance and these less important attributes, then a positive partworth for performance might have been evidenced.

Table 4 gives the correlation among partworths implied by the estimated model. The implications are generally more reasonable that for the model with all normal distributions. People who are concerned about price are also concerned about operating cost. People who like electric vehicles also tend to like hybrid vehicles. This result suggests that customers have a willingness or unwillingness to consider new technologies that transcends the particular technology. The questionable correlation between the partworths of range and electric vehicles that arose in the model with all normal distributions is not evidenced in the model with transformations of normals.

As specified, the partworths for price, operating cost and range have lognormal distributions, which allow unboundedly large partworths and have fairly thick tails. It might be more appropriate to give these partworths a S_B distribution. To investigate this question, and to illustrate how various distributions can be tested, we estimated a model that is the same as the one just described except that the partworths for these three variables are specified as S_B with a lower bound of zero and a high upper bound for each. The upper bounds are 1.0 for price and operating cost, and 2.0 for range, which are high enough to accommodate nearly all of the cumulative

16

distribution under the respective log-normals and yet allow a different shape of the distribution within the relevant range. The log-likelihood for this model is higher than for the previous one: -6159.7 compared to -6171.5. For price and operating cost, the mean and variance of the partworths are about the same with the S_B distribution as the log-normal; however, the shape differed, with the tail of the S_B distribution being considerably smaller even within the support of the S_B. Figure 1 illustrates the difference for the price coefficient, with the solid line representing the S_B distribution and the dashed line representing the log-normal.

For the partworths associated with range, using the S_B distribution instead of the log-normal had a substantial effect. Figure 2 shows the estimated densities under the S_B distribution (solid line) and log-normal (dashed line). The S_B distribution provides a plateau shape that cannot be accommodated with a log-normal. The question arises of whether this shape is the result of placing a maximum of 2.0 on the S_B distribution when the estimated log-normal distribution gives a non-negligible share above 2. We re-estimated the model with the maximum of the range coefficient set at 5.0 instead of 2.0. The estimated S_B density with this maximum takes a shape that is similar to the log-normal. However, the log-likelihood dropped considerably, from -6159.7 with the maximum set at 2.0 to -6163.1 with the maximum set at 5.0. Apparently the improvement in fit that arises from using S_B distribution instead of the log-normal is due to the plateau shape that the S_B distribution takes when its maximum is set at 2.0 for this partworth.

The bounds of the S_B distribution can be estimated as parameters, rather than specified by the researcher. Doing so requires an extra layer of Gibbs sampling, with the bounds drawn from their posterior distribution conditional on β_n $\forall n$. The conditional posterior is proportional to the logit likelihood for the entire sample, $\prod_n L(y_n \mid \beta_n)$, times the prior on the bounds, where the utility that is used to calculate the logit formulas in $L(\cdot)$ depends on the bounds of the S_B distributions. A M-H algorithm is used to take draws from this conditional posterior, similar to that used by Train (2001) for fixed coefficients.

We estimated a model with the upper bound of the S_B distribution for the range coefficient treated as a parameter. Using a flat prior, the estimated value was 2.86 with a standard error of 0.42. The log-likelihood of the model
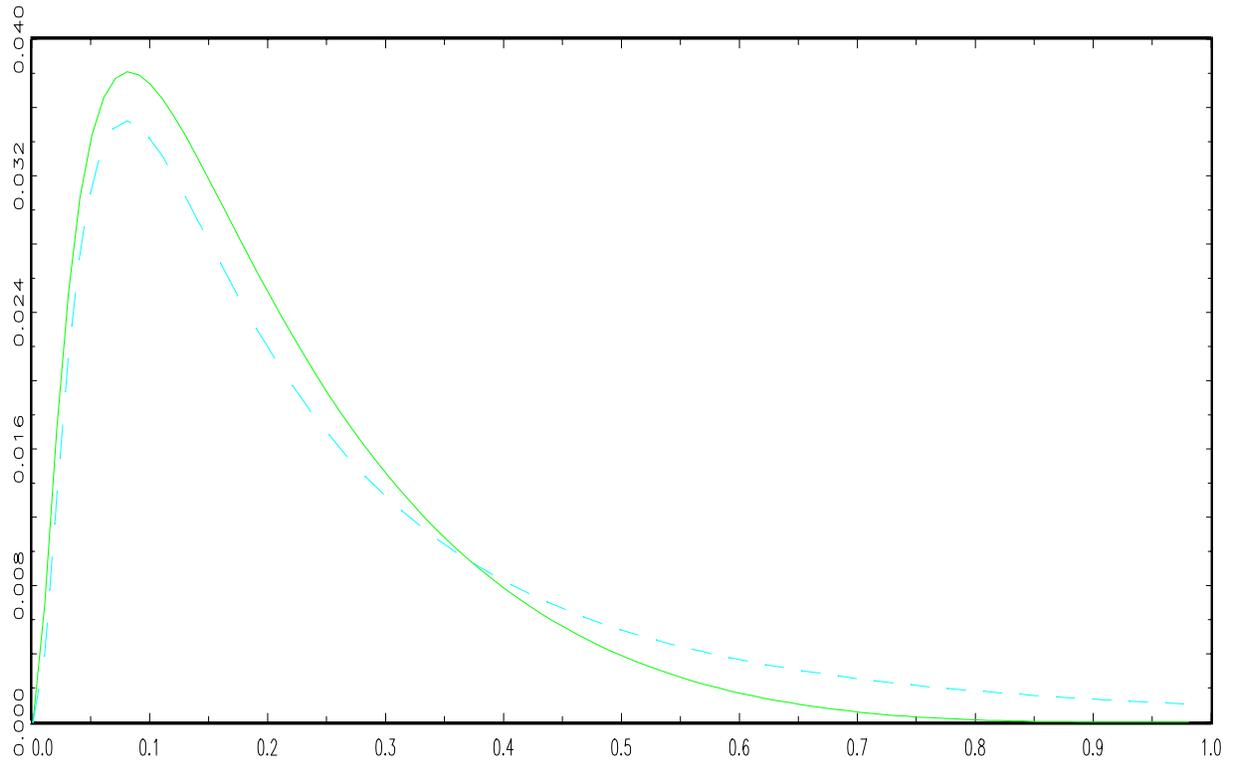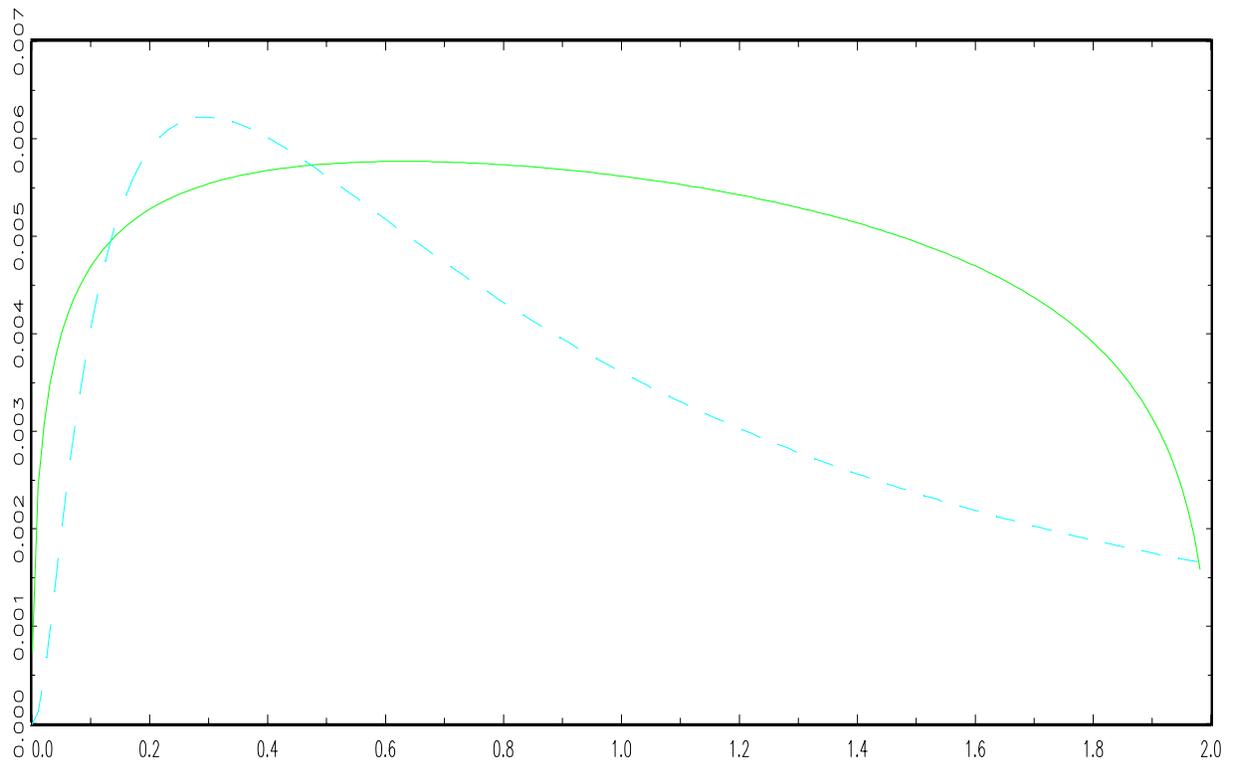
Figure 1: Density of Price Coefficient

Figure 2: Density of Range Coefficient

dropped slightly from -6159.7 with the upper bound set at 2.0 to -6160.56 with the estimated bound. Run time approximately doubled, since the M-H algorithm for the bounds of the S_B distribution requires about the same amount of calculation as the M-H algorithm for $\beta_n \ \forall n$. As noted above, run times are fairly short with the procedure such that doubling them is not a burden. However, identification becomes an issue when the bounds are treated as parameters, since the difference between the upper and lower bounds, $u - \ell$, is closely related to the variance ω of the latent normal term. An important area for further work is whether the S_B distributions can be re-parameterized in a way that improves identification of each parameter when the researcher does not specify the bounds.

References

Albert, J. and S. Chib (1993), 'Bayesian analysis of binary and polychotomous response data', *Journal of the American Statistical Association* **88**, 669–679.

Allenby, G. (1997), 'An introduction to hierarchical bayesian modeling', Tutorial notes, Advanced Research Techniques Forum, American Marketing Association.

Allenby, G. and P. Rossi (1999), 'Marketing models of consumer heterogeneity', *Journal of Econometrics* **89**, 57–78.

Bhat, C. (1998), 'Accommodating variations in responsiveness to level-of-service variables in travel mode choice models', *Transportation Research A* **32**, 455–507.

Bhat, C. (2000), 'Incorporating observed and unobserved heterogeneity in urban work mode choice modeling', *Transportation Science* **34**, 228–238.

Boatwright, P., R. McCulloch and P. Rossi (1999), 'Account-level modeling for trade promotion: An application of a constrained parameter hierarchical model', *Journal of the American Statistical Association* **94**, 1063–1073.

Brownstone, D., D. Bunch and K. Train (2000), 'Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles', *Transportation Research B* **34**, 315–338.

Brownstone, D. and K. Train (1999), 'Forecasting new product penetration with flexible substitution patterns', *Journal of Econometrics* **89**, 109–129.

Casella, G. and E. George (1992), 'Explaining the gibbs sampler', *The American Statistician* **46**, 167–174.

Fieller, E. (1932), 'The distribution of the index in a normal bivariate population', *Biometrika* **24**, 428–440.

Geary, R. (1930), 'The frequency distribution of the quotient of two normal variates', *Journal of the Royal Statistical Society* **93**, 442–446.

Gelman, A., J. Carlin, H. Stern and D. Rubin (1995), *Bayesian Data Analysis*, Chapman and Hall, Suffolk.

Goett, A., K. Hudson and K. Train (2000), 'Consumers' choice among retail energy suppliers: The willingnes-to-pay for service attributes', *The Energy Journal* **21**, 1–28.

Hensher, D., N. Shore and K. Train (2004), 'Households' willingness to pay for water service attributes', working paper, Department of Economics, University of California, Berkeley.

Hinkley, D. (1969), 'On the ratio of two normal random variables', *Biometrika* **56**, 635–639.

Johnson, N. (1949), 'Systems of frequency curves generated by methods of translation', *Biometrika* **36**, 149–176.

Johnson, N. and S. Kotz (1970), *Continuous Multivariate Distributions I*, John Wiley, New York.

Johnson, R. M. (2000), 'Monotonicity constraints in choice-based conjoint with hierarchical bayes', Research paper, Sawtooth Software.

Koop, G. and D. Poirier (1993), 'Bayesian analysis of logit models using natural conjugate priors', *Journal of Econometrics* **56**, 323–340.

Koop, G. and D. Poirier (1996), 'Rank ordered logit models: An empirical analysis of voter preferences before the 1988 canadian federal election', *Journal of Applied Econometrics* **9**, 369–388.

Marsaglia, G. (1965), 'Ratios of normals variables and ratios of sums of uniform variables', *Journal of the American Statistical Association* **60**, 193–204.

McCulloch, R. and P. Rossi (1994), 'An exact likelihood analysis of the multinomial probit model', *Journal of Econometrics* **64**, 207–240.

McFadden, D. and K. Train (2000), 'Mixed mnl models of discrete response', *Journal of Applied Econometrics* **15**, 447–470.

Merrill, A. (1928), 'Frequency distribution of an index when both the components follow the normal law', *Biometrika* **20A**, 53–63.

Poirier, D. (1994), 'Jeffrey's prior for logit models', *Journal of Econometrics* **63**, 327–329.

Poirier, D. (1996), 'A bayesian analysis of nested logit models', *Journal of Econometrics* **75**, 163–181.

Revelt, D. (1999), Three Discrete Choice Random Coefficients Papers and One Police Crime Study, PhD thesis, University of California, Berkeley.

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

SawtoothSoftware (1999), 'The cbc/hb module for hierarchical bayes', at www.sawtoothsoftware.com.

Train, K. (1998), 'Recreation demand models with taste variation', *Land Economics* **74**, 230–239.

Train, K. (2001), 'A comparison of hierarchical bayes and maximum simulated likelihood for mixed logit', Working Paper, Department of Economics, University of California, Berkeley.

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

Train, K. and C. Winston (2004), 'Vehicle choice behavior and the declining market of us automakers', working paper, Department of Economics, University of California, Berkeley.

Table 1: Model of vehicle choice with all normal distributions

β_n's and partworths for:	Mean	Variance	Share>0
Price (negative):	.1900	.0632	.78
	(.0127)	(.0048)	
Operating cost (negative):	.0716	.0467	.63
	(.0127)	(.0032)	
Range:	1.213	4.050	.73
	(.2442)	(.7190)	
Electric vehicle:	-3.554	16.95	.19
	(.4120)	(3.096)	
Hybrid vehicle:	1.498	6.483	.72
	(.1584)	(.9729)	
High performance:	.3092	1.425	.60
	(.1004)	(.2545)	
Mid and high performance:	.8056	1.298	.76
	(.1030)	(..2384)	
Log-likehood	-6835.5.		

Table 2: Correlations among partworths with all normal distributions

Price	1	0.11	-.10	0.05	-.18	-.07	-.01
Operating cost		1	-.05	0.15	0.01	0.01	-.01
Range			1	-.64	0.36	-.01	0.15
Electric vehicle				1	0.12	0.02	-.19
Hybrid vehicle					1	0.19	0.06
High performance						1	0.17
Med and high performance							1

Table 3: Model of vehicle choice with transformations of normals

| | β_n | | Partworths | |
	Mean	Variance	Mean	Variance
Price (negative):	-2.531	0.9012	0.1204	0.0170
	(.0614)	(.1045)		
Operating cost (negative):	-3.572	1.015	0.0455	0.0031
	(.1100)	(.1600)		
Range:	-1.222	1.370	0.5658	0.8965
	(.2761)	(.3368)		
Electric vehicle:	-1.940	2.651	-1.9006	2.6735
	(.1916)	(.4965)		
Hybrid vehicle:	0.9994	2.870	1.0003	2.8803
	(.1267)	(.4174)		
High performance:	-.7400	2.358	0.3111	0.3877
	(.2953)	(.7324)		
Mid and high performance:	-.0263	1.859	0.5089	0.5849
	(.1538)	(.3781)		
Log-likehood	-6171.5			

Table 4: Correlations among partworths with transformations of normals

Price	1	0.25	0.14	0.00	0.35	0.12	0.05
Operating cost		1	0.08	-.10	0.17	0.02	-.04
Range			1	-.05	0.27	0.03	0.02
Electric vehicle				1	0.38	0.04	-.11
Hybrid vehicle					1	0.22	0.09
High performance						1	0.14
Med and high performance							1