# Forecasting New Product Penetration with Flexible Substitution Patterns

by

David Brownstone
Department of Economics
University of California, Irvine

and

Kenneth Train
Department of Economics
University of California, Berkeley

October 1996

ABSTRACT: We describe and apply choice models, including generalizations of logit called "mixed logits," that do not exhibit the restrictive "independence from irrelevant alternatives" property and can approximate any substitution pattern. The models are estimated on data from a stated-preference survey that elicited customers' preferences among gas, electric, methanol, and CNG vehicles with various attributes.

ACKNOWLEDGEMENTS: David Bunch and Tom Gollob collected the data and conducted preliminary analyses upon which our analysis relies. We are grateful to them for allowing us to use the data. They are not, of course, responsible for any errors or representations that we make in this paper.

# Forecasting New Product Penetration with Flexible Substitution Patterns

by

David Brownstone and Kenneth Train

## 1. Introduction

By far the most popular econometric models for forecasting demand for new products are logit and nested logit (McFadden, 1973, 1978; Ben-Akiva and Lerman, 1985). While computationally convenient, these models exhibit the well-known and restrictive "independence from irrelevant alternatives," or iia, property. Logit exhibits the property over all alternatives while nested logit exhibits it over alternatives within each nest. This property states that the ratio of the probabilities for any two alternatives is independent of the existence and attributes of any other alternative. As a result of this property, the models necessarily predict that a change in the attributes of one alternative (or the introduction of a new alternative, or the elimination of an existing alternative) changes the probabilities of the other alternatives proportionately, such that the ratios of probabilities remain the same. This substitution pattern can be unrealistic in many settings. For example, consider the introduction of electric cars, as examined, e.g., by Train (1980, 1986) and Brownstone. et al. (1996). The logit model predicts that, among households with the same observed characteristics, the electric vehicle will draw the same proportion of households from large luxury gas cars as from small gas cars. Howevever, if the electric car is similar in size to a subcompact gas car, one might expect the electric car to draw disproportionately from different classes of vehicles, with, for example, households who would have chosen a subcompact gas car switching more readily to the electric car than households who would have chosen a large gas car. More fundamentally, identification of the correct substitution pattern is an empirical issue, and the iia property of logit and nested logit imposes a particular substitution pattern rather than allowing the data analysis to find and reflect whatever substitution pattern actually occurs.

In this paper, we describe and estimate models for new product forecasting that can represent very general patterns of substitution. We first provide a general specification that distinguishes several types of models, particularly mixed logits with various structures and probits. These specifications have been known (citations given below), though perhaps not described in the same manner. More importantly, there have been few applications, particularly of mixed logits. We apply the models to data from a stated-preference survey on households' choices among gas, methanol, CNG, and electric vehicles. We compare the forecasts from models that allow flexible substitution patterns with those from a standard logit model.

## II. Specification

A person faces a choice among J alternatives. Without loss of generality, the person's utility from any alternative can be decomposed into a nonstochastic, linear-in-parameters part that depends on observed data, a stochastic part that is perhaps correlated over alternatives and heteroskedastic, and another stochastic part that is independently, identically distributed (iid) over alternatives and people. In particular, the utility from alternative i is denoted $U_i = \beta' x_i + [\eta_i + \varepsilon_i]$ where $x_i$ is a vector of observed variables relating to alternative i and the person; $\beta$ is a vector of parameters to be estimated which are fixed over people and alternatives; $\eta_i$ is a random term with zero mean whose distribution over people and alternatives depends in general on underlying parameters and observed data relating to alternative i; and $\varepsilon_i$ is a random term with zero mean that is iid over alternatives, does not depend on underlying parameters or data, and is normalized to set the scale of utility. Stacking the utilities, we have: $U=\beta' X+[\eta+\varepsilon]$ where $V(\varepsilon)=\alpha I$ with known (i.e., normalized) $\alpha$ and $V(\eta)$ is general and can depend on underlying parameters and data. For standard logit, each element of $\varepsilon$ is iid extreme value, and, more importantly, $\eta$ is zero, such that the unobserved portion of utility (i.e., the term in brackets) is independent over alternatives. This independence gives rise to the iia property and its restrictive substitution patterns. We consider below models that allow correlation and heteroskedasticity.

### 1. Mixed logit: General distribution for $\eta$ and extreme value for $\varepsilon$.

Let each element of $\varepsilon$ be iid extreme value as for standard logit; however, allow any distribution for $\eta$. Denote the density of $\eta$ as $f(\eta|\Omega)$ where $\Omega$ are the fixed parameters of the distribution. Given the value of $\eta$, the conditional choice probability is simply logit, since the remaining error term is iid extreme value:

$$L_i(\eta) = \exp(\beta' x_i + \eta_i) / \sum_j \exp(\beta' x_j + \eta_j).$$

Since $\eta$ is not given, the (unconditional) choice probability is this logit formula integrated over all values of $\eta$ weighted by the density of $\eta$:

$$P_i = \int L_i(\eta)\ f(\eta|\Omega) d\eta$$

Models of this form are called "mixed logit" since the choice probability is a mixture of logits with f as the mixing distribution. The probabilities do not exhibit iia and different substitution patterns are attained by appropriate specification of f.

The choice probability cannot be calculated exactly because the integral does not have a closed form in general. The integral is approximated through simulation. For a given value of the parameters $\Omega$, a value of $\eta$ is drawn from its distribution. Using this draw, the logit formula $L_i(\eta)$ is calculated. This process is repeated for many draws, and the average of the resulting $L_i(\eta)$'s is taken as the approximate choice probability:

$$SP_i = (1/R) \sum_{r=1,\ldots,R} L_i(\eta^r)$$

where R is the number of replications (i.e., draws of $\eta$), $\eta$ is the r-th draw, and $SP_i$ is the simulated probability that the person chooses alternative i. By construction, $SP_i$ is an unbiased estimate of $P_i$ for any R; its variance decreases as R increases. It is strictly positive for any R, such that $\ln(SP_i)$ is always defined, which is important when using $SP_i$ in a log-likelihood function (as below). It is smooth (i.e., twice differentiable) in parameters and variables, which helps in the calculation of elasticities and especially in the numererical search for the maximum of the likelihood function. The simulated probabilities sum to one over alternatives, which is useful in forecasting.

The choice probabilities depend on parameters $\beta$ and $\Omega$, which are to be estimated. Adding subscript n to index sampled individuals and denoting the chosen alternative for each person by i, the log-likelihood function $\sum_n \ln(P_{ni})$ is approximated by the simulated log-likelihood function $\sum_n \ln(SP_{ni})$ and the estimated parameters are those that maximize the simulated log-likelihood function.[1] Lee (1992) and Hajivassiliou and Ruud (1994) derive the asymptotic distribution of the maximum simulated likelihood estimator based on smooth probability simulators with the number of replications increasing with sample size. Under regularity conditions, the estimator is consistent and asymptotically normal. When the number of replications rises faster than the square root of the number of observations, the estimator is asymptotically equivalent to the maximum likelihood estimator.

The gradient of the simulated log-likelihood function is simple to calculate, which speeds the search for the maximum:

$$G(\beta) \equiv \delta \sum_n \ln(SP_{ni}) \,/\, \delta\beta = \sum_n [1/SP_{ni}](1/R) \sum_r L_{ni}(\eta_n^r)[\sum_j (d_{nj} - L_{nj}(\eta_n^r))x_{nj}]$$

$$G(\Omega) \equiv \delta \sum_n \ln(SP_{ni}) \,/\, \delta\Omega = \sum_n [1/SP_{ni}](1/R) \sum_r L_{ni}(\eta_n^r)[\sum_j (d_{nj} - L_{nj}(\eta_n^r)(\delta\eta_n^r/\delta\Omega)]$$

where $d_{nj} = 1$ for $j=i$ and zero otherwise. The derivative $\delta\eta_n^r/\delta\Omega$ depends on the specification of $\eta$ and f. Also, if the same parameters enter $\beta$ and $\Omega$ (as in the third model in section III), the gradient is adjusted accordingly. Analytic second derivatives can also be calculated. However, Revelt and Train (1996) found that calculating the Hessian from formulas for the second derivatives resulted in computationally slower estimation than using the bhhh or other approximate-Hessian procedures.

Different types of mixed logit models have been used in empirical work; they differ in the type of structure that is placed on the model, or, more precisely, in the specification of f. In section III below, as in Train (1995) and Ben-Akiva and Bolduc (1996), we specify an error-components structure: $U_i = \beta' x_i + \mu' z_i + \varepsilon_i$ where $\mu$ is a random vector with zero mean that does not vary over alternatives and has density $g(\mu|\Omega)$ with parameters $\Omega$; $z_i$ is a vector of observed data related to alternative i; and $\varepsilon_i$ is iid extreme value. This is a mixed logit with a particular structure for $\eta$, namely, $\eta_i=\mu' z_i$. The terms in $\mu' z_i$ are interpreted as error components that induce heteroskedasticity and correlation over alternatives in the unobserved portion of utility: $E([\mu' z_i +\varepsilon_i]\ [\mu' z_j +\varepsilon_j]) = z_i' V(\mu) z_j$. Even if the elements of $\mu$ are uncorrelated such that $V(\mu)$ is diagonal, the unobserved portion of utility is still correlated over alternatives.

In this specification, the choice probabilities are simulated by drawing values of $\mu$ from its distribution and calculating $\eta_i=\mu' x_i$. Insofar as the number of error components (i.e., the dimension of $\mu$) is smaller than the number of alternatives (the dimension of $\eta$), placing an error-components structure on a mixed logit reduces the dimension of integration and hence simulation that is required for calculating the choice probabilities.

Different patterns of correlation, and hence different substitution patterns, are obtained through appropriate specification of $z_i$ and g. For example, an analog to nested logit is obtained by specifying $z_i$ as a vector of dummy variables -- one for each nest taking the value of 1 if i is in the nest and zero otherwise -- with $V(\mu)$ being diagonal (thereby providing an independent error component associated with each nest, such that there is correlation in unobserved utility within each nest but not across nests). Restricting $V(\mu)=\sigma I$ is analogous to restricting the log-sum coefficients in a nested logit model to be the same for all nests. Importantly, McFadden (1996) has shown that any random utility model can be approximated by a mixed logit with an error-components structure and appropriate choice of the $z_i$'s and g.[2]

Most recent empirical work with mixed logits has been motivated by a random-parameters, or random-coefficients, specification (Bhat, 1996a and b; Mehndiratti, 1996; Revelt and Train, 1996; Train 1996).[3] The difference between a random-parameters and an error-components specification is entirely interpretation. In the random-parameters specification, the utility from alternative i is $U_i = b' x_i + \varepsilon_i$ where coefficients b are random with mean $\beta$ and deviations $\mu$. Then $U_i = \beta' x_i + [\mu' x_i +\varepsilon_i]$, which is an error-components structure with z=x. Elements of x that do not enter z can be considered variables whose coefficients do not vary in the population. And elements of z that do not enter x can be considered variables whose coefficients vary in the population but with zero means. In different contexts one or the other interpretation will seem more natural.

Other types of mixed logits have also been used. Elrod (1988) and Erdem (1995) provide a factor-analytic structure to $\eta$. This specification is the same as the error components described above with the important difference that the $z_i$'s are estimated (subject to normalization) rather than being observed variables. Ben-Akiva and Buldoc (1996) specify a "general autoregressive" error structure, under which each $\eta_i$ is correlated with each other $\eta_j$ for all $j \neq i$ with the covariance being proportional to weights associated with each i-j pair. This set-up is particularly useful for spatial choice models (such as destination choice for travel and shipping), with the weight for any two locations reflecting the distance between the them.

Ben-Akiva and Bolduc (1996) use the term "probit with a logit kernel" to describe any model where $\eta$ is normally distributed but elements of $\varepsilon$ are iid extreme value. This term is instructive since it points out that the distinction between pure probit models (with all components of the error being normal) and mixed logits with a normal mixing distribution (where all error components are normal except the final component which is iid extreme value) is conceptually minor and might be empirically indistinguishable. Unlike pure probits, however, mixed logits can also represent situations where $\eta$ is not normal, as might arise when random coefficients must take a particular sign (such as a price coefficient that must be negative for all people) such that the unrestricted range of the normal is inappropriate.

Mixed probits: General distribution for $\eta$ and standard normal for $\varepsilon$

When the elements of $\varepsilon$ are iid standard normal, then a family of models arises that is analogous to that discussed above for extreme value $\varepsilon$ and with estimation performed by the same type of simulation (i.e., averaging the conditional choice probability over numerous draws of $\eta$). However, given $\eta$, the conditional choice probabilities do not have a closed form as in the mixed logit case. Conditional on $\eta$ <u>and</u> $\varepsilon_i$ (that is, the iid term for the chosen alternative), the conditional choice probability is a product of univariate cumulative normals, which is easy to calculate; this is essentially the Stern (1992) simulator generalized to allow for non-normal distributions of $\eta$. However, conditioning on $\varepsilon_i$ adds an extra dimension of simulation when calculating the unconditional probability (Train, 1995). Therefore, unless there is a reason to expect $\varepsilon$ to be normal instead of extreme value, assuming $\varepsilon$ to be extreme value seems preferable on pragmatic grounds for general distributions of $\eta$.

Pure probits: Normal $\eta$ and standard normal for $\varepsilon$

When $\eta$ is normal as well as $\varepsilon$, then the model is a pure probit and simulation methods that have been developed for probits can be utilized. The model can be characterized as $U_i = \beta' x_i + \xi_j$ where the vector of unobserved utility components $\xi=(\xi_1, ..., \xi_J)'$ is distributed normal with zero mean and covariance matrix $\Lambda$. Given our notation, $\xi_i = \eta_i + \varepsilon_i$, and any of the structures given above can be placed on $\eta$, which gives a structure to $\Lambda$. Hajivassiliou, McFadden, and Ruud (1992) describe and compare, using Monte Carlo methods, several probit simulators. These simulators differ structurally from the simulator described above for mixed logits. In particular, the probit simulators are based on draws of $\xi$ (or more precisely, on draws of the difference between $\xi_j$ for each non-chosen alternative and $\xi_i$ for the chosen alternative) while the mixed logit simulator is based on draws of the random terms that compose $\eta$. The probit simulators draw from a J-1 dimensional distribution of utility differences, while the mixed logit simulator draws from a mixing distribution whose dimension is determined by the specification of the model.

Hajivassiliou et al. (1992) found the GHK simulator (due to Geweke, 1991, Hajivassiliou and McFadden, 1990, and Keane, 1990) performed better than other probit simulators for the specifications that they examined. To our knowledge, there has been no comparison of the mixed logit simulator with probit simulators. We provide a comparison on our data set below. However, the advantages and limitations of each depend on the specific situation. For example, in situations where the the dimension of the mixing distribution is less than the number of alternatives (as in Train, 1996, which had 59 alternatives and seven error components), the mixed logit simulator might have an advantage simply because the simulation is over fewer dimensions. The opposite occurs when the dimension of the mixing distribution exceeds the number of alternatives (as in

Revelt and Train, 1996).

In the following section we estimate: (1) a mixed logit with an error-components structure and normally distributed $\eta$, (2) a pure probit with the same structure, and (3) a mixed logit with some elements of $\eta$ being non-normal. Before describing the models, we describe the data that were used in estimation.

## III. Estimated Models

We utilize the data collected and described by Brownstone et al. (1996) for their models of vehicle choice. A sample of California households were interviewed regarding their attitudes and preferences for alternative-fueled vehicles. Completed questionaires were obtained from 4654 respondents. A central component of the survey was a conjoint-type experiment in which vehicles with different fuels and different attributes (such as price and operating cost) were described to the respondent and the respondent was asked to state which he/she would choose.[4] Figure 1 shows the information and question that was presented to one of the respondents; different respondents were presented with different vehicle characteristics. Note that each respondent was given detailed information about three vehicles, each with a different fuel and each of a particular size class, and was told that each of these three vehicles came in either of two specified body types. Thus the respondent faced a choice among six alternatives. However, the total number of distinct alternatives was 120, consisiting of each combination of 4 fuels (gas, methanol, compressed natural gas, electricity), 5 size classes (mini, subcompact, compact, mid-size, large), and 6 body types (regular car, sports car, truck, van, station wagon, sports utility vehicle).

The variables that enter the model are defined in Table 1. The choice of variables to enter the nonstochastic portion of utility was determined though exploration and testing with a standard logit model. Most of the variables are self-explanatory; however, a few notes are required. (i) Dividing price by the natural log of income provides a higher log-likelihood for the logit model than: not dividing by income, dividing by untransformed income, or dividing by the square root of income (with price and income always measured in thousands of dollars.) The price coefficient for a respondent with median income is essentially the same under any of these specifications. (ii) The questionaire described to the respondent the cost of recharging/refueling the vehicle at a station, as well as, for electric and CNG vehicles, the cost of refueing at home. We found that station refueling cost is not significant for electric vehicles and home refueling cost is not significant for CNG vehicles. Consistent with these findings, the variable that enters the model is defined as home refueling cost for electric vehicles and station refueling cost for non-electric vehicles. (iii) When a separate constant is included for each size class, with the constant for the mini class normalized to zero, the estimated coefficients obtain the following pattern nearly exactly: the coefficient for compacts is twice that of subcompacts, and the coefficients for mid-sized and large vehicles are equal to each other and three time that of subcompacts. The size variable that enters the model is a parsimonious representation of this result: 0 for mini, 1 for subcompact, 2 for compact, and 3 for mid-size and large (multimplied by 0.1 for scaling.)[5].

Column 1 of Table 2 gives the estimated parameters and standard errors for the logit model. Column 2 presents a mixed logit with the same specification for the nonstochastic portion of utility plus four error components. The first and second error components are iid normal deviates that enter the utility for each non-electric vehicle and each non-CNG vehicle, respectively. These error components are motivated by the nested logit specification of Bunch and Bradley (1995), which includes two overlapping nests: one consisting of non-electric vehicles and the other consisting of non-CNG vehicles. The third and fourth error components relate to the dimensions of the vehicle. In particular, the third error component is a normal deviate multiplied by the size variable described above, and the fourth error component is a normal deviate multiplied by the luggage space variable. To be precise, the stochastic portion of utility for alternative i is defined as $[\sum_{k=1-4}\sigma_k(\varsigma_k z_{ki})] + \varepsilon_i$ where $\varsigma_k$ is iid standard normal, $z_{ki}$ are the four variables described above, and $\varepsilon_i$ is iid extreme value. The parameters $\sigma_k$ for k=1-4 are estimated; each denotes the standard deviation of the normal deviate that generates that error component. In simulating the choice probability for a respondent, four numbers are drawn from a random-number generator for the standard normal distribution; the four "variables" $\varsigma_1 z_{1i}$ - $\varsigma_4 z_{4i}$ are created; and the conditional probability is evaluated with coefficients $\sigma_k$ k=1-4 for the four "variables." This process is repeated for numerous draws and the conditional probabilities are averaged to obtain the simulated probability. We used 250 draws in estimation of the mixed logit model.[6]

The error compoents enter significantly. Gas and methanol vehicles enter both the non-EV and non-CNG error components, unlike electric and CNG vehicles which enter only one. The covariance in the stochastic portion of utility is therefore greater for gas and methanol vehicles than other pairs of fuels. CNG vehicles enter the non-EV component, which has a larger coefficient than the non-CNG component; the covariance between the stochastic portion of utility for CNG vehicles with that for gas and methanol vehicles is larger than for electric vehicles with gas and methanol vehicles. Stated succinctly, the following pairs are given in order of decreasing covariance: a gas vehicle paired with a methanol vehicle, gas or methanol paired with CNG, gas or methanol paired with electric, CNG paired with electric.

The error component associated with the variable "size" induces covariance across size classes. Since the variable is largest for mid-size and large vehicles, the covariance is largest for these. The covariance decreases for mid-size or large vehicles paired with either compact, subcompact, and mini vehicles, respectively. Similarly, the error component associated with luggage space induces greater covariance for pairs of vehicles with greater luggage space. It seems doubtful (a least to us) that households actually care about luggage space per se sufficiently to justify an error component based upon it. However, the luggage space error component entered significantly in all the specifications that we tried during preliminary analysis. We interpret, for the following reason, the luggage space variable as a second indication of vehicles' overall dimensions. Respondents were told the size class of each vehicle as well as the luggage space relative to a comparable gas vehicle. Respondents could easily consider the luggage space information as an indication of relative dimensions of the vehicle within the fairly broad size classes. For example, if a respondent is told that an electric vehicle is a mini with 75% of the luggage space of a mini gas car, the respondent could logically think that the electric vehicle is smaller than a mini gas car - a mini-mini, so to speak.[7]

The estimated parameters that enter the nonstochastic portion of utility are generally larger in magnitude in the mixed logit than the standard logit. This phenomenon is expected. The scale of utility is determined by the normalization of the iid term ε. In a standard logit, all stochastic terms are absorbed (as well as possible, given that they are not, in reality, all iid) into this one error term. The variance of this error term is larger in the standard logit model than in a mixed logit since, in the mixed logit, some of the variance in the stochastic portion of utility is captured in η rather than ε. Utility is scaled so that ε has the variance of an extreme value. Since the variance before scaling is larger in the standard logit than the mixed logit, utility (and hence the parameters) are scaled down in the standard logit relative to the mixed logit. This is the same result as obtained by Revelt and Train (1996).

The ratios of estimated parameters, which are the economically meaningful statistic, are very similar in the standard logit and mixed logit models. For example, the ratio of the first two coefficeints (for price and range) is 0.529 in the standard logit and 0.511 in the mixed logit. The ratio of the second to the third coefficients (range and acceleration) is .489 in both models. Bhat (1996a) and Train (1996) also found the ratios of coefficients not to differ significantly between a standard and mixed logit, while Bhat (1996b) found fairly substantial differences.

The third column of Table 2 presents the estimated parameters of a pure probit model. This model has the same specification as the mixed logit in column 2 except that the final term in the stochastic portion of utility, ε, consists of iid standard normal terms rather than iid extreme value. The choice probabilities are simulated with the GHK simulator. This simulator requires more computer time per replication than the mixed logit simulator; to keep the computer time manageable we reduced the number of replications to 50. The relative accuracy and speed of the mixed logit and GHK simulators on our data are examined in section V below.

The ratios of estimated parameters are similar in the pure probit model to those in the standard and mixed logits. The scale of the estimated parameters is about the same as in the standard logit. This is due to two counteracting factors. First, as described above for the mixed logit, the incorporation of part of the stochastic portion of utility into η rather than ε causes the parameters to rise in magnitude, since the parameters are scaled by the variance of ε. Second, the variance of a standard normal is smaller than that of an extreme value; therefore, utility is scaled down further in a probit model where ε is standard normal than in a logit model where ε is extreme value. In our application, it is simply a coincidence that these two factors have approximately the same impact, though in opposite directions, such that the probit parameters are similar in magnitude to the standard logit parameters. It is interesting to note that we did not use the logit parameters directly as starting values for the probit model but instead used the scaled logit parameters (i.e., the logit parameters divided by 1.6 to account for the difference in the variance between standard normal and extreme value). The iteration process moved the parameters back to the approximate scale of the original logit parameters.

The fourth column of Table 2 presents a mixed logit with more error components than the mixed logit in column 2 and the pure probit. These extra error components do not have normal distributions; as result, there is no pure probit analog to this model. The motivation and specification of this model require some discussion. We wanted to try a fairly complete random-parameters specification, in which households' tastes regarding each attribute of vehicles vary in the population. We first estimated a model in which each of the following variables was assumed to have a coefficient that is distributed independently normally in the population with mean and standard deviation being estimated: price/ln(income), range, acceleration, pollution, size, big enough, luggage space, operating cost, station availability, the EV constant, and the CND constant. (Note that the mixed logit in column 2 could be interpreted as having random parameters for size, luggage space, EV constant, CNG constant. See footnote 7.) In this specification, only one extra coefficient, beyond those four in the mixed logit model of column 2, obtained a statistically significant standard deviation. More importantly, the model produced counter-intuitive forecasts under some scenarios. For example, the predicted share of households choosing a large gas car was predicted to rise in response to a 20% rise in the price of large gas cars. This phenomenon is a natural (though undesirable) consequence of having the coefficients of attributes take a normal distribution when in reality all households can be expected to have the same sign for their coefficients. For example, the price coefficient is necessarily negative for all households; and yet the normal distribution for this coefficient necessarily implies that some households having positive coefficients. In forecasting, the households with positive price coefficients prefer a vehicle more when its price rises. If the share of households with positive price coefficients is large compared to the share of households choosing a given vehicle, then the model can predict that a price increase raises demand. This phenomenon occurs not just for price but for any attribute whose coefficient is given a normal distribution and yet has an expected sign for all households. The phenomenon was evidenced in one of the models of Revelt and Train (1996) in forecasting the impact of rising interest rates on households' decisions to take loans. Pure probits with a random-parameters specification are, by their nature, susceptible to it.

The solution is to specify a density that is strictly positive only on one side of zero. For the model in column 4, we assume log-normal distributions for the coefficients of price/ln(income), range, acceleration, top speed, pollution, big enough, operating cost and station availability. Since the log-normal distribution gives positive coefficients, variables whose coefficients are necessarily negative (price/ln(income), acceleration, pollution, and operating cost) are entered as the negative of the variable. The four variables that enter the error components of the mixed logit in column 2, could logically take different signs by different households; these are therefore assumed to have normal distributions, as in the mixed logit model in column 2.

The k-th coefficient with a log-normal distribution is specified as $\exp(b_k+s_k\nu_k)$, where $\nu_k$ is iid standard normal and $s_k$ and $b_k$ are parameters. The mean coefficient is $\exp(b_k+(s_k^2/2))$ and the standard deviation of the coefficient is the mean multiplied by $\sqrt{(\exp(s_k^2)-1)}$. In preliminary analysis, models with unrestricted $b_k$'s and $s_k$'s failed to converge, with one or more of the $s_k$'s becoming so large that $\exp(b_k+s_k\nu_k)$ exceeded the numerical limit of the software. We therefore constrained the parameters such that the standard deviation of each log-normally distributed coefficient was equal to its mean. (Mechanically, we constrained each $s_k$ to be .8326.) This constraint is appealing, since it results in a model with no more parameters than in the mixed logit in column 2 or the pure probit, and yet contains variation in the stochastic portion of utility over more attributes than in these models. The implications of this variation for substitution patterns is explored below.

Table 2 gives two sets of estimates for the variables with log-normally distributed coefficients.

On the right is the estimate of $b_k$ for each coefficient; the standard error is for the estimate of $b_k$. On the left is the mean coefficient implied by the point estimate of $b_k$ and the constrained value of $s_k$ (multiplied by -1 if the negative of the variable was entered.). The left-hand number is comparable to the estimates in the previous columns for other models. The ratios of these estimates are similar to those in the other three models; the scale is similar to that of the mixed logit in column 2. If indeed the ratios of coefficients are adequately captured by a standard logit model, as our results and those of Bhat (1996a) and Train (1996) indicate, then the extra difficulty of estimating a mixed logit or a probit need not be incurred when the goal is simply estimation of willingness to pay, without using the model for forecasting.

## IV. Substitution Patterns

Table 4 provides an illustration of the substitution patterns that are implied by each of the models. In particular, the table gives the probabilities from each of the four models under various scenarios compared with a base situation. The probabilities are for a typical household; specifically, they are for a person with a college education, who commutes less than 5 miles to work, and whose household has three members and an annual income of $50,000. In the base situation, the household can choose among five classes of gas cars (mini, sub-compact, compact, mid-sized, and large); no alternative-fueled vehicles are offered. Four scenarios are considered, three in which new alternative-fueled vehicles are offered and a fourth in which the price of gas vehicles is changed. The table gives the change in probabilities that results from the introduction of new vehicles or the change in price.

It is important to note that the probabilities do not represent forecasts; rather, they illustrate the substitution patterns that are implied by each model. To use the models for forecasting, it is necessary to (i) calibrate the model against real-world choices, (ii) forecast the characteristics of households and develop a set of households for each forecast year that represents the distribution of characteristics in that year, and (iii) develop forecasts of vehicle characteristics that realistically represent future offerings or the range of possible future offerings. These tasks are currently being undertaken as a continuation of the work reported in Brownstone et al (1996).

While the figures in Table 4 are not forecasts, the differences in the substitution patterns that arise under the different models will also occur in forecasting since these differences are intrinsic to the model specifications.

We first describe the differences between the standard logit and the mixed logit in column 2 of Table 2, called mixed logit A. We then describe differences with the pure probit and the mixed logit in column 4, called mixed logit B.

In part 1 of Table 4, a mini electric car is introduced to a base situation consisting of five gas cars. The logit model, because of the iia property, implies that the new electric car will draw proportionately from all five of the gas cars. The electric car is predicted to be chosen by 7.2% of households with the same observed characteristics as our typical household. As the figures show, 7.2% of the households who chose large gas cars in the base situation are predicted to switch to the electric mini car - the same as for households who chose gas mini cars in the base situation. The predictions from mixed logit A differ from those of the standard logit model in two ways. First, the percent of households who switch to the new electric car is higher among households who would have bought the gas mini car in the base than among households that would have bought larger gas cars. This more realistic substitution pattern is the consequence of the error component relating to size. The difference in substitution patterns is particularly important for policy analysis. For example, electric cars are seen as a way of reducing gas consumption and tailpipe emissions. The predicted reductions are lower when households are predicted, realistically, to switch from small gas cars more readily than from large gas cars.

Second, the mixed logit model predicts that more households in total choose the new electric car (11.2% as opposed to 7.2% in the standard logit model.) This higher prediction is due to the non-EV error component: the random term associated with this component is negative for some households, reflecting the fact that some households prefer electric vehicles to non-electric vehicles independent of the measured attributes (such as price and operating cost) which enter the model. The mixed logit probabilities incorporate the fact that some households actually like electric vehicles, thereby raising the share who are predicted to buy one.[8] This feature of mixed logits, namely, the ability to reflect greater distribution of preferences regarding new products relative to existing products, is potentially important in forecasting penetration rates for any new product, but especially for products that are expected to satisfy niche markets.

For part 2 of Table 4, a second electric car in introduced, comparable in size to a gas subcompact. The previous scenario (five gas cars and a mini electric car) is taken as the base. The standard logit model predicts, again because of the iia property, that the subcompact electric car will attract the same share of households who had bought any of the gas cars as it does of households that had chosen the mini electric car. The mixed logit model provides more realistic predictions: the new electric car, which is larger than the original electric car, draws more proportionately from the original electric car than from the gas cars.

The total share obtained by the new sub-compact electric car is the same from both the standard and mixed logit. This result, while it might at first appear to contradict the finding in part 1 that the mixed logit predicts a larger total share of electric cars, actually confirms that finding and explanation. In part 1, the mixed logit gives a larger share for the electric car because the mixed logit incorporates the fact that some people actually prefer electric cars. In part 2, the base situation contains an electric car, and households who prefer electric cars are predicted to buy this car in the base situation. The introduction of a second, larger electric car does not meet previously unmet preferences of households who like electric cars (as in part 1) since, in part 2, their preferences have been met in the base situation. The second electric car induces some of these households to switch from the smaller electric car from the larger one: this is one of the reasons for the large percent draw from the mini electric that is predicted by the mixed logit.

For part 3, the scenario in part 2 is taken as the base (five gas cars and two electric cars) and two methanol cars, a compact and a mid-sized, are introduced. The logit model of course predicts proportionate switching. The mixed logit predicts disproportionate switching, with greater switching from the larger gas cars than the mini and subcompact gas cars, and with greater switching from gas cars than electric cars. The later result is due to the fact that the same error components enter for methanol and gas, indicating a similarity in households' views of these two types of fuel (relative to electric). This finding is reasonable, since refueling with methanol is essentially the same as refueling with gas, whereas the procedures for recharging an electric car are quite different.

Finally, for part 4, the base is the five gas cars, as in part 1, and the scenario is a 20% rise in the price of large gas cars. The standard logit predicts that households switch proportionately to each of the other size classes. The mixed logit model predicts that households switch more readily to mid-size cars than to smaller cars, as one would expect in reality.

The predictions from the pure probit model are qualitatively the same as those from mixed logit A. Mixed logit B also has similar predictions, with one exception. In part 4, mixed logit B predicts a considerably smaller share of households switching away from large cars in reaction to a 20% price increase. Mixed logit B includes variation over households in the price coefficient, which neither mixed logit A and the pure probit do. Households who place relatively litte importance on price have a greater tendency to buy the expensive large car. And these households, since their price coefficient is relatively small, do not react to price increases as readily as a household with average price coefficient. As a result, mixed logit B predicts a smaller share of customers switching away from the large car when the price is raised.

In conclusion, the substitution patterns that are obtained with the mixed logit and probit models are more reasonable and provide more realistic policy implications than the patterns implied by standard logit. It is important to note that the improved substitution patterns were obtained with the addition of only four extra parameters.

## V. Simulator Speed and Variance

To our knowledge, there have been no previous analyses of the variance of the mixed logit simulator. To examine this issue, we calculated the variance in the simulated values of the average probability, log-likelihood function, and gradient over different draws for mixed logit A. Using the parameters values in Table 2, we calcuated the average probability, log-likelihood, and gradient repeatedly using different seeds for the random-number generator (that is, using different draws for the random terms that constitute the error components). We set the number of replications at 50, 125, and 250, and for each number of replications, performed the calculations with ten different seeds. The mean and variance over the ten seeds for each number of replications are given in Table 4.

The "average probability" is the simulated probability for each respondent, averaged over respondents. The mean (over the ten seeds) of the average probability is the same under each number of replications; this is expected since the simulated probability is unbiased for the true probability for any finite number of replications. The variance of the average probability decreases as the number of replications increases, also as expected. The variance decreases faster than 1/R, where R is the number of replications; in fact, it decreases at about twice this rate. With accept/reject simulation of a choice probability, variance is expected to decrease at the rate 1/R (McFadden, 1989). The faster rate for the mixed logit simulator indicates that additional draws of the mixed logit simulator provide more extra information than an additional draw of an accept/reject simulator.

The mean over the ten seeds of the simulated log-likelihood function decreases as the number of replications rises. The log of the simulated probability is not unbiased for the log of the true probability; rather, given the log transformation, it is biased downward for a finite number of replications, with the bias decreasing as the number of replications increases. The figures in Table 4 are consistent with these facts. Whether the bias can be considered large depends on the perspective that one takes. The mean of the simulated log-likelihood increases by 28.5 points when the number of replications rises from 50 to 250, and by 6.7 points when the number of replications rises from 125 to 250. However, these absolute changes represent only 0.4% and 0.09% changes. The issue can be considered in the context of interating to the maximum. In our estimation, the iteration process came within 6.2 points of the final maximized value after the third iteration; the remaining increase took another seven iterations. During the final seven iterations the point estimates changed in the second and sometimes the first digit (though the ratios changed less than the absolute values). However, the hypothesis that the true log-likelihood was maximized after three iterations (or more precisely, that the point estimates obtained after three iterations were the true ones) could not be rejected, even at low confidence levels.[9] Unfortunately, as for most empirical work, the range of point estimates for which the null hypothesis that they are true cannot be rejected is fairly wide. However, within this context, the bias in the log-likelihood function that is introduced by simulation seems to be minor.

The variance in the simulated log-likelihood function decreases, like the average probability, faster than 1/R.

The variance in the gradient decreases as the number of replications rises, as expected. It is difficult to determine the rate since there are 25 elements to the gradient and some inconsistency in the change in the variances. More than ten seeds are apparently needed to obtain a reliable sense of the rate of decrease in the variance of each element of the gradient.

The calculations were performed on a PC with a pentium 166 MHz chip and 32 MB of RAM (which was sufficient memory to hold all the data and intermediate results, thus speeding the process considerably) using GAUSS for Windows95. The computer time for both the log-likelihood and gradient was nearly proportional to the number of replications. Our model with 250 replication required ten iterations to converge, and each iteration entailed several calculations of the log-likelihood function and one calculation of the gradient; the total time for estimation of this model was somewhat more than two-and-a-half hours. By comparison, the standard logit model took less than a minute.

We performed similar calculations for the pure probit. The results are given in the last column of Table 4. The variance in the average probability and the log-likelihood function is somewhat smaller for the GHK simulator with 50 replications than the mixed logit with the same number of replications. However, the calculation of the GHK simulator is considerably slower. The mixed logit with 125 replications takes about the same time to calculate the log-likelihood as the GHK simulator with 50 replications, and yet the variance in the average probability and log-likelihood is less than half as large[10]. Stated succinctly: for a given number of replications, the GHK simulator has somewhat less variance than the mixed logit simulator, while for a fixed amount of computer time, the mixed logit simulator has considerably lower variance. The variance in the gradient also follows this patterns generally; however, there are sufficient differences that any conclusion about gradients cannot be considered very strong.[11]

These comparisons should not be taken as general. The relative performance, both in speed and variance, depend on the model specification, data, and other factors. For example, the GHK simulator requires a Choleski decomposition for the covariance matrix of the differences in the stochastic portion of utility. With our data (where each respondent faced a different set of alternatives) this decomposition had to be performed separately for each respondent. In other situations -- for example, when each person faces the same alternatives and the covariance matrix does not depend on data that varies over people -- the Choleski decomposition is not needed for each person; one decomposition for each alternative is sufficient. Also, in our application this covariance matrix has large off-diagonal elements (i.e., the correlation across alternatives in the stochastic portion of utility is high). The GHK simulator is known to have larger simulation variance when the correlation in utility differences is high (e.g., Geweke, Keane, Runkle, 1996). Issues also arise regarding the computer program. GAUSS is particularly fast for calculations that can be "vectorized." The mixed logit simulator and its gradient can easily be vectorized, while only certain parts of the GHK simulator can be. A comparison of speeds using FORTRAN, which does not provide an advantage to vectorization, could have different results. Nevertheless, the recursive nature of the GHK simulator (where the range for the random draw for one alternative depends on the value of previous draws for other alternatives) is inherently slow compared to simulators, like the mixed logit simulator, which draw simultaneously from unrestricted ranges. In light of these issues, the results in Table 4 are perhaps best interprested as simply an indication that the mixed logit simulator is reasonably accurate compared to the GHK simulator, particularly for given computer time.

## References

Ben-Akiva, M. and D. Buldoc, 1996, "Multinomial Probit with a Logit Kernel and a General Parametetric Specification of the Covariance Structure," working paper, Department of Economics, MIT.

Ben-Akiva, M., and S. Lerman, 1985, Discrete Choice Analysis, MIT Press.

Bhat, C., 1996a, "Accommodating Variations in Responsiveness to Level-of-Service Measures in Travel Model Choice Modeling," working paper, Department of Civil Engineering, University of Massachusetts at Amherst.

Bhat, C., 1996b, "Incorporating Observed and Unobserved Heterogeneity in Urban Work Travel Mode Choice Modeling," working paper, Department of Civil Engineering, University of Massachusetts at Amherst.

Boyd, J., and R. Melman, 1980, "The Effect of Fuel Economy Standards on the U.S. Automotive Market: An Hedonic Demand Analysis," Transportation Research, Vol. 14A, No. 5-6, pp. 367-378.

Brownstone, D., D. Bunch, T. Golob, and W. Ren, 1996, "Transactions Choice Model for Forecasting Demand for Alternative-Fuel Vehicles," in S. McMullen, ed., Research in Transportation Economics, Vol. 4, pp. 87-129, JAI Press, Inc.

Bunch, D., and M. Bradley, 1995, "Personal Vehicle Survey Model Results: Modeling Reactions to Hypothetical Future Alternative and Conventional Fuel Vehicles," paper presented at

Cardell, N. and F. Dunbar, 1980, "Measuring the Societal Impacts of Automobile Downsizing," Transportation Research, Vol. 14A, No. 5-6, pp. 423-434.

Erdem, T., 1995, "A Dynamic Analysis of Market Structure Based on Panel Data," working paper, Haas School of Business, University of California, Berkeley.

Elrod, T., 1988, "Choice Map: Inferring a Product Market Map from Observed Choice Behavior," Marketing Letters, Vol. 2, No. 3, pp. 253-266.

Geweke, J., 1991, "Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints," Computer Science and Statistics: Proceedings of the Twenty-ThirdSymposium on the Interface, p. 571-78, American Statistical Association, Alexandria, VA.

Geweke, J., M. Keane, and D. Runkle, 1996, "Statistical Inference in the Multinomial Multiperiod Probit Model," Journal of Econometrics, forthcoming.

Griliches, Z., 1961, "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Changes," in The Price Statistics of the Federal Government, National Bureau of Economic Research, New York.

Hajivassiliou, V. and D. McFadden, 1990, "A Method of Simulated Svires for the Estimation of LDV Models, Colwes Foundation Discussion Paper #967, Yale University.

Hajivassiliou, V., D. McFadden, and P. Ruud, 1992, "Simulation of Multivariate Normal Rectangle Probabilities and their Derivatives: Theoretical and Computational Results," Cowles Foundation Discussion Paper 1021, Yale University.

Keane, M., 1990, Four Essays in Empirical Macro and Labor Economics, Ph.D. dissertation, Brown University.

Lee, L., 1992, "On Efficiency of Methods of Simulated Moments and Maximum Simulated Likelihood Estimation of Discrete Response Models," Econometrica, Vol. 8, pp. 518-552.

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

McFadden, D., 1975, "On Independence, Structure and Simultaneity in Transportation Demand Aanlysis," Working Paper No. 7511, Urban Travel Demand Forecasting Project, Institute of Transportation and Traffic Engineering, University of California, Berkeley.

McFadden, D., 1978, "Modeling the Choice of Residential Location," in A. Karquist, et al., eds, Spatial Interaction Theory and Planning Models, North-Holland Publishing Company, Amsterdam.

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

McFadden, D., 1996, "Economic Choice Behavior: Psychological Foundations and the Contributions of Amos Tversky," paper presented at the 1996 Econometric Society European Meetings, Istanbul.

McFadden, D., and P. Ruud, 1994, "Estimation by Simulation," Review of Economics and Statistics, Vol. LXXVI, No. 4, pp. 591-608.

Mehndiratta, S., 1996, "Time-of-Day Effects in Inter-City Business Travel," Ph.D. thesis, Department of Civil Engineering, University of California, Berkeley.

Revelt, D. and K. Train, 1996, "Incentives for Appliance Efficiency in a Competitive Energy Environment: Random-Parameters Logit Models of Households' Choices," working paper, Department of Economics, University of California, Berkeley.

Stern, S., 1992, "A Method for Smoothing Simulated Moments of Discrete Probabilities in Multinomial Probit Models," Econometrica, Vol. 60, pp. 943-952.

Train, K., 1980, "The Potential Market for Non-Gasoline-Powered Automobiles," Transportation Research, Vol. 14A, pp. 405-414.

Train, K., 1986, Qualitative Choice Analysis, MIT Press.

Train, K., 1995, "Simulation Methods for Probit and Related Models Based on Convenient Error Partitioning," Working Paper No. 95-237, Department of Economics, University of California, Berkeley.

Train, K., 1996, "Unobserved Taste Variation in Recreation Demand Models," working paper, Department of Economics, University of California, Berkeley.

Figure 1: Vehicle Choice Survey Question

Suppose that you were considering purchasing a vehicle and the following three vehicles were available: (assume that gasoline costs $1.20 per gallon)

| | Vehilce A | Vehilce B | Vehicle C |
|---|---|---|---|
| Fuel Type | Electric<br>Runs on electricity only | Natural Gas (CNG)<br>Runs on CNG only | Methanol<br>Can also run on gasoline |
| Vehicle Range | 80 miles | 120 miles | 300 miles on methanol |
| Purchase Price | $21,000 (includes home charge unit) | $19,000 (includes home refueling unit) | $23,000 |
| Home refueling time | 8 hrs for full charge (80 miles) | 2 hrs to fill empty tank (120 miles) | Not available |
| Home refueling cost | 2 cents per mile (50 mpg gasoline equivalent) | 4 cents per mile (25 mpg gasoline equivalent) | |
| Service station refueling time | 10 min. for full charge (80 mi.) | 10 min. to fill empty CNG tank (120 mi.) | 6 min. to fill empty tank (300 mi.) |
| Service station fuel cost | 10 cents per mile (10 mpg gasoline equivalent) | 4 cents per mile (25 mpg gasoline equivalent) | 4 cents per mile (25 mpg gasoline equivalent) |
| Service station availability | 1 recgarge station for every 10 gasoline stations | 1 CNG station for every 10 gasoline stations | Gasoline available at current stations |
| Acceleration Time to 30 mph | 6 seconds | 2.5 seconds | 4 seconds |
| Top speed | 65 miles per hour | 80 miles per hour | 80 miles per hour |
| Tailpipe emissions | 'Zero' tailpipe emissions | 25% of new 1993 gasoline car emissions when run on CNG | Like new 1993 gasoline cars when run on methnaol |
| Vehicle size | Like a compact car | like a sub-compact car | Like a mid-size car |
| Body types | Car or truck | Car or van | Car or truck |
| Luggage space | Like a comparable gasoline vehcile | Like a comparable gasoline vehicle | Like a comparable gasoline vehicle |

Given these choices, which vehicle would you purchase? (please circle one choice)

1) Vehicle "A" (car)
2) Vehicle "A" (truck)
3) Vehicle "B" (car)
4) Vehicle "B" (van)
5) Vehicle "C" (car)
6) Vehicle "C" (truck)

Table 1: Variable Definitions

| Variable names: | Definitions: |
|---|---|
| Price / ln(income) | Purchase price in thousands of dollars, divided by the natural log of household income in thousands. |
| Range | Hundreds of miles that the vehicle can travel between refuelings/rechargings. |
| Acceleration | Seconds required to reach 30mph from stop, in tens of seconds (e.g., 3 seconds is entered as .3.) |
| Top speed | Highest speed that the vehicle can attain, in hundreds of miles per hour (e.g., 80mph is entered as .80). |
| Pollution | Tailpipe emissions as fraction of comparable new gas vehicle. |
| Size | 0=mini, 0.1=subcompact, 0.2=compact, 0.3=mid-size or large. |
| "Big enough" | 1 if household size is over 2 and vehicle size is 3; 0 otherwise. |
| Luggage space | Luggage space as fraction of comparable new gas vehicle. |
| Operating cost | Cost per mile of travel, in tens of cents per mile (e.g. 5 cents per mile is entered as .5.) For electric vehicles, cost is for home recharging. For other vehicles, cost is for station refueling. |
| Station availability | Fraction of stations that have capability to refuel/recharge the vehicle. |
| Sports utility vehicle | 1 for sports utility vehicle, zero otherwise. |
| Sports car | 1 for sports car, zero otherwise. |
| Station wagon | 1 for station wagon, zero otherwise. |
| Truck | 1 for truck, zero otherwise. |
| Van | 1 for van, zero otherwise. |
| Constant for EV | 1 for electric vehicle, zero otherwise. |
| Commute < 5 x EV | 1 if respondent commutes less than five miles each day and vehicle is electric; zero otherwise. |
| College x EV | 1 if respondent had some college education and vehicle is electric; zero otherwise. |
| Constant for CNG | 1 for compressed natural gas vehicle, zero otherwise. |
| Constant for methanol | 1 for methanol vehicle, zero otherwise. |
| College x methanol | 1 if respondent had some college education and vehicle is methanol; zero otherwise. |
| Non-EV | 1 if vehicle is not electric; zero if electric. |
| Non-CNG | 1 if vehicle is not CNG; zero if CNG. |

Table 2: Models of Vehicle Choice

| | Standard Logit | | Mixed Logit A | |
|---|---|---|---|---|
| | Estimate | Std. Error | Estimate | Std. Error |
| Variables: | | | | |
| Price / ln(income) | -0.185 | 0.027 | -0.264 | 0.043 |
| Range | 0.350 | 0.027 | 0.517 | 0.058 |
| Acceleration | -0.716 | 0.111 | -1.062 | 0.186 |
| Top speed | 0.261 | 0.080 | 0.307 | 0.115 |
| Pollution | -0.444 | 0.100 | -0.608 | 0.139 |
| Size | 0.935 | 0.311 | 1.435 | 0.508 |
| "Big enough" | 0.143 | 0.076 | 0.224 | 0.113 |
| Luggage space | 0.501 | 0.188 | 1.702 | 0.482 |
| Operating cost | -0.768 | 0.073 | -1.224 | 0.159 |
| Station availability | 0.413 | 0.097 | 0.616 | 0.145 |
| Sports utility vehicle | 0.820 | 0.144 | 0.901 | 0.148 |
| Sports car | 0.637 | 0.156 | 0.700 | 0.162 |
| Station wagon | -1.437 | 0.065 | -1.500 | 0.067 |
| Truck | -1.017 | 0.055 | -1.086 | 0.056 |
| Van | -0.799 | 0.053 | -0.816 | 0.056 |
| Constant for EV | -0.179 | 0.169 | -1.032 | 0.425 |
| Commute < 5 x EV | 0.198 | 0.082 | 0.372 | 0.166 |
| College x EV | 0.443 | 0.108 | 0.766 | 0.218 |
| Constant for CNG | 0.345 | 0.091 | 0.626 | 0.148 |
| Constant for methanol | 0.313 | 0.103 | 0.415 | 0.146 |
| College x methanol | 0.228 | 0.089 | 0.313 | 0.124 |
| Error components: | | | | |
| Non-EV | | | 2.464 | 0.541 |
| Non-CNG | | | 1.072 | 0.377 |
| Size | | | 7.455 | 1.819 |
| Luggage space | | | 5.994 | 1.248 |
| Log-likelihood | -7391.83 | | -7375.34 | |

Table 2 continued: Models of Vehicle Choice

| | Probit | | Mixed Logit B | | |
|---|---|---|---|---|---|
| | Estimate | Std. Error | Estimate | | Std. Error |
| Variables: | | | | | |
| Price / ln(income) | -0.184 | 0.031 | -0.286 | -1.599 | 0.172 |
| Range | 0.371 | 0.044 | 0.588 | -0.877 | 0.126 |
| Acceleration | -0.761 | 0.136 | -1.046 | -0.302 | 0.190 |
| Top speed | 0.207 | 0.082 | 0.361 | -1.364 | 0.335 |
| Pollution | -0.414 | 0.098 | -0.695 | -0.711 | 0.234 |
| Size | 0.983 | 0.363 | 1.541 | | 0.533 |
| "Big enough" | 0.164 | 0.080 | 0.246 | -1.748 | 0.495 |
| Luggage space | 1.333 | 0.371 | 1.563 | | 0.463 |
| Operating cost | -0.894 | 0.121 | -1.318 | -0.071 | 0.135 |
| Station availability | 0.434 | 0.103 | 0.674 | -0.741 | 0.236 |
| Sports utility vehicle | 0.594 | 0.103 | 0.897 | | 0.149 |
| Sports car | 0.448 | 0.114 | 0.698 | | 0.163 |
| Station wagon | -1.085 | 0.049 | -1.508 | | 0.067 |
| Truck | -0.798 | 0.041 | -1.094 | | 0.056 |
| Van | -0.614 | 0.042 | -0.819 | | 0.056 |
| Constant for EV | -1.058 | 0.359 | -0.905 | | 0.418 |
| Commute < 5 x EV | 0.294 | 0.136 | 0.359 | | 0.163 |
| College x EV | 0.615 | 0.181 | 0.770 | | 0.218 |
| Constant for CNG | 0.465 | 0.108 | 0.621 | | 0.152 |
| Constant for methanol | 0.315 | 0.101 | 0.476 | | 0.154 |
| College x methanol | 0.203 | 0.086 | 0.335 | | 0.128 |
| Error components: | | | | | |
| Non-EV | 2.232 | 0.435 | 2.289 | | 0.553 |
| Non-CNG | 0.707 | 0.300 | 0.971 | | 0.412 |
| Size | 5.187 | 1.434 | 6.808 | | 2.072 |
| Luggage space | 4.823 | 0.935 | 5.380 | | 1.293 |
| Log-likelihood | -7368.74 | | -7375.19 | | |

Table 3: Probabilities for a Typical Household under Different Scenarios

| Base: Five gas cars. Scenario: Add mini EV. | | | | |
|---|---|---|---|---|
| | Logit | Mixed Logit A | Probit | Mixed Logit B |
| Percent change in probability: | | | | |
| Mini gas | -7.2 | -17.5 | -17.9 | -19.3 |
| Subcomp. gas | -7.2 | -13.9 | -15.9 | -15.7 |
| Compact gas | -7.2 | -10.5 | -13.1 | -13.1 |
| Mid-size gas | -7.2 | -7.4 | -9.0 | -8.6 |
| Large gas | -7.2 | -7.4 | -9.4 | -9.0 |
| Probability for mini EV | 7.2% | 11.2% | 12.9% | 12.9% |

| Base: Five gas cars and mini EV. Scenario: Add subcompact EV. | | | | |
|---|---|---|---|---|
| | Logit | Mixed Logit A | Probit | Mixed Logit B |
| Percent change in probability: | | | | |
| Mini gas | -6.7 | -4.6 | -4.3 | -5.0 |
| Subcomp. gas | -6.7 | -4.9 | -4.7 | -5.3 |
| Compact gas | -6.7 | -5.0 | -5.0 | -5.6 |
| Mid-size gas | -6.7 | -4.6 | -4.5 | -4.7 |
| Large gas | -6.7 | -4.6 | -4.8 | -5.0 |
| Mini EV | -6.7 | -22.7 | -27.4 | -26.5 |
| Probability for Subcomp EV | 6.7% | 6.7% | 7.6% | 7.8% |
| Total EV Prob. | 13.4% | 15.4% | 17.0% | 17.3% |

| Base: Five gas car classes, two EV classes. Scenario: Add compact and mid-size methanol. | | | | |
|---|---|---|---|---|
| | Logit | Mixed Logit A | Probit | Mixed Logit B |
| Percent change in probability: | | | | |
| Mini gas | -42.3 | -33.3 | -34.2 | -34.5 |
| Subcomp. gas | -42.3 | -41.7 | -44.2 | -42.4 |
| Compact gas | -42.3 | -48.6 | -53.7 | -48.6 |
| Mid-size gas | -42.3 | -53.6 | -52.3 | -54.0 |
| Large gas | -42.3 | -53.6 | -59.2 | -53.2 |
| Mini EV | -42.3 | -23.2 | -18.9 | -26.1 |
| Subcomp. EV | -42.3 | -30.6 | -25.2 | -32.5 |
| Total prob. for methanol | 42.3% | 43.6% | 43.9% | 44.0% |

| Base: Five gas cars. Scenario: Raise price of large gas car by 20%. | | | | |
|---|---|---|---|---|
| | Logit | Mixed Logit A | Probit | Mixed Logit B |
| Percent change in probability: | | | | |
| Mini gas | 4.2 | 2.5 | 2.8 | 2.0 |
| Subcomp. gas | 4.2 | 4.0 | 4.4 | 3.1 |
| Compact gas | 4.2 | 5.6 | 6.3 | 4.2 |
| Mid-size gas | 4.2 | 7.2 | 6.6 | 5.8 |
| Large gas | -22.9 | -30.2 | -31.5 | -21.8 |

Table 4: Simulation Speed and Variance

| | Mixed logit | | | Pure Probit |
|---|---|---|---|---|
| | 50 replications | 125 replications | 250 replications | 50 replications |
| Seconds for | | | | |
| Loglikelihood | 19.8 | 48.1 | 95.5 | 51.0 |
| Gradient | 132. | 328. | 654. | 1310.* |
| Average prob: | | | | |
| Mean | .2388 | .2388 | .2388 | .2394 |
| Variance | 8.37xE-8 | 2.72E-8 | 0.751E-8 | 7.50E-8 |
| StdDev/Mean | .001212 | .000690 | .000363 | .00114 |
| Log-likelihood: | | | | |
| Mean | -7406.1 | -7384.3 | -7377.6 | -7356.3 |
| Variance | 35.3 | 11.6 | 2.19 | 26.9 |
| StdDev/Mean | .000802 | .000461 | .000200 | .000705 |
| Gradient Var: | | | | |
| Price / ln(inc) | 5.59 | 0.571 | 0.577 | 5.58 |
| Range | 16.5 | 6.59 | 3.90 | 12.3 |
| Acceleration | 0.555 | 0.105 | 0.0925 | 0.332 |
| Top speed | 1.12 | 0.102 | 0.0925 | 0.749 |
| Pollution | 1.75 | 0.699 | 0.286 | 1.15 |
| Size | 0.147 | 0.0554 | 0.0215 | 0.114 |
| "Big enough" | 1.08 | 0.376 | 0.232 | 0.628 |
| Lugg space | 0.414 | 0.0532 | 0.0242 | 0.0803 |
| Op cost | 1.16 | 0.0706 | 0.134 | 1.57 |
| Station avail. | 3.2 | 0.939 | 0.394 | 1.87 |
| Sports utility | 0.0944 | 0.0298 | 0.0128 | 0.192 |
| Sports car | 0.0848 | 0.0399 | 0.0140 | 0.471 |
| Station wagon | 0.115 | 0.0465 | 0.0370 | 2.27 |
| Truck | 0.310 | 0.177 | 0.0883 | 2.33 |
| Van | 0.464 | 0.197 | 0.0730 | 1.35 |
| EV Const | 5.68 | 1.58 | 0.957 | 0.382 |
| Cm<5xEV | 2.09 | 0.238 | 0.125 | 0.296 |
| College x EV | 2.20 | 1.22 | 0.852 | 0.178 |
| CNG Const | 3.09 | 0.886 | 0.273 | 1.92 |
| Meth Const | 3.65 | 1.39 | 0.448 | 2.80 |
| Collegexmeth | 5.65 | 1.82 | 0.514 | 3.42 |
| Error Comps: | | | | |
| Non-EV | 4.97 | 1.55 | 0.908 | 0.590 |
| Non-CNG | 5.11 | 1.90 | 0.682 | 1.05 |
| Size | 0.102 | 0.0730 | 0.0325 | 0.0227 |
| Lugg space | 0.464 | 0.133 | 0.0927 | 0.0417 |

* The gradient for the pure probit model is calculated numerically rather than analytically.

ENDNOTES:

1. Note that even though $SP_i$ is unbiased for $P_i$, $\ln(SP_i)$ is a biased estimator of $\ln(P_i)$ for finite R, such that simulator induces bias in the log-likelihood function. This bias decreases as R increases and, as stated, when R increases faster than the square root of the number of observations, disappears asymptotically. While we utilize maximum simulated likelihood (MSL) estimation, as do all the empirical studies cited below, other forms of parameter estimation could be applied, such as method of simulated moments (MSM), method of simulated scores, or Gibbs resampling. See, e.g., McFadden and Ruud (1994).

2. This results differs critically, and is stronger than the "mother logit" theorem, which states that any choice model can be approximated by a model that takes the form of a standard (i.e., non-mixed) logit (McFadden, 1975; Train, 1986, pp. 21-24.) In the mother logit theorem, any choice model can be expressed as a standard logit if attributes of one alternative are allowed to be entered in the "representative utility" other alternatives. However, when cross-alternative attributes are entered, the logit model is no longer a random utility model (i.e., is not consistent with utility maximizing behavior and cannot be used for welfare analysis) since the utility of one alternative depends on the attributes of other alternatives. In the theorem regarding mixed logits, any random utility model can be approximated by a mixed logit with an error-components structure without entering cross-alternative variables, or, more precisely, while still maintaining the mixed logit as a random utility model.

3. The earliest estimated versions of random-parameter logits were called hedonic models (Boyd and Mellman, 1980; Cardell and Dunbar, 1980); however, this term can cause confusion with respect to the hedonic price methods made popular by of Griliches (1961). These early models were estimated on aggregate data such that the computationally intensive simulation was required for only one "decision-maker" rather than, as in the recent models, for each sampled decision-maker.

4. Respondents were also asked to state their second choice. We did not utilize the second-choice data in our analysis. Logit models estimated on the first and second choice responses yielded essentially the same parameter estimates with only slightly smaller standard errors. Also, the inclusion of two choices for each respondent makes the estimation of mixed logits and pure probits somewhat more complicated (so as to incorporate the correlation in unobserved factors over the two choices) and increases the computer time required for estimation by considerably more than twice. See Revelt and Train (1996) and Train (1996) for estimation of mixed logits with repeated choices for each decisionmaker.

5. Interating to the maximum of the objective function with quasi-Newton Raphson procedures, as we use, is faster when variables are scalesd such that the elements of the diagonal of the Hessian have approximately the same magnitude.

6. Different draws are taken for different respondents. McFadden (1992) indicates the importance, in the context of method of simulated moments estimation, of taking different draws for each observation: with differerent draws for each observation, the simulation error in the simulated probability is uncorrelated over observations and tends to cancel out when the simulated probability is averaged over observations. With maximum simulated likelihood (MSL) estimation, the average of the log of the simulated probability is utilized rather than the average

of the untransformed probabilities. Since the log of the simulated probability is not an unbiased estimate of the log of the true probability, the cancelling-out can only be shown to occur in MSL when the number of replications is sufficiently large as to effectively eliminate the bias. However, some type of cancelling probably occurs even with fewer replications. Lee (1992) describes the properties of MSL estimators when the same draws are used for all observations.

7. As described in section II, a random parameters interpretation can be placed on the error components if such an interpretation seems reasonable or natural in the situation. For example, the error component related to the size variable can conceiveably be considered to represent the deviation in respondents' tastes for large vehicles around the mean tastes (where the mean tastes are captured by the coefficient of size in the nonstochastic portion of utility.) Under this interpretation, the coefficient of size has an estimated mean of 1.435 and an estimated standard deviation of 7.455. Since the distribution is normal, some share of the population has positive coefficients and some have negative, reflecting the fact that some household like bigger vehicles while some prefer smaller vehicles. The point estimates for the mean and standard deviation imply that 57% of the population prefers larger cars while the other 43% prefer smaller cars. Similarly, the point estimates for the other error compoents and their corresponding coefficients in the nonstochastic portion of utility imply: 61% of the population prefer larger luggage compartments while the other 39% prefer smaller luggage compartments (or, probably more accurately, the kinds of design that allow for little luggage space); 33% of the population who do not have college education and commute further than 5 miles prefer electric vehicles for reasons beyond the attributes that enter the model, while the other 67% prefer non-electric vehicles; and 72% of the population prefer CNG vehicles for reasons beyond the entered attributes, while the other 28% prefer non-CNG vehicles. (These last two statements are made by noting that the non-EV and non-CNG error components are equivalent to EV and CNG error components, such that the EV and CNG constants entering the nonstochastic portion of utility represent the mean tastes.) These shares do not seem unreasonable, though this impression is largely due to the fact that our expectations regarding such shares are largely unformed. From our perspective, we see the error components simply as ways of inducing correlation over fuel types and vehicle dimensions in the nonstochastic part of utility, and thereby of providing disproportional substitution patterns over vehicles with different fuels and dimensions. The fact that the error components enter highly significantly suggests that there is indeed such correlation and dispropotional substitution.

8. Actually, it is not the variance in the electric vehicle preference per se that drives the difference between the logit and mixed logit, it is the difference in the variance in the preference for electric vehicles relative to that for gas vehicles. In a standard logit, the iid error terms have the same variance, implying that preferences for an electric vehicle relative to any of the five as vehicles has the same variance as that for any of the gas vehicles relative to any of the other gas vehicles. The mixed logit, by including the non-EV component, implies that the preference for electric vehicles relative to any gas vehicle has greater variance than that of any gas vehicle relative to any other gas vehicle. The greater variance implies that a larger share of households actually like electric cars, which gives rise to the larger prediction. See Ben-Akiva and Bolduc (1996) for a discussion of the implications of error components associated with alternative-specific constants in mixed logits.

9. The test statistics is the gradient'(hessian-inverse)gradient, which is distributed under the null hypothesis as chi-squared with degrees of freedom equal to the number of parameters. At the third iteration, the test statistic took the value of 8.17, which is lower than the critical value of 37.6 for 25 degrees of freedom and 5% probability of rejecting the null hypothesis when true. The critical value using a 99% probability of rejecting when true (and hence a lower probability of accepting the null hypothesis when false) is 11.5, which also results in acceptance of the null hypothesis. (Our convergence criterion was that this test statistic be less than 0.0001.)

10. The time required for calculation of the gradient is not completely comparable between the GHK and mixed logit simulators, since we calculated the gradient numerically for the GHK simulator and analytically for the mixed logit. The numerical gradient with GHK simulator using 50 replications took twice as long to calculate as the analytic gradient with the mixed logit simulator using 250 replications. The variances indicate that the mixed logit with analytic gradient strongly dominates, for a given amount of computer time, the GHK probit simulator with numerical gradient. GHK with analytic gradient would undounbtedly be faster, though how much so is unclear.

11. Two general-purpose packages are available to estimate probit models with the GHK simulator: GAUSSX by Jon Breslaw, which uses numerical gradients, and the FORTRAN program written by Axel Borsch-Supan with modifications by Vassilis Hajivassiliou and Ann Royalty, which uses analytic derivatives. Both programs estimate each element of the covariance matrix of utilities and would require modifications to place structure, such as an errors-component structure, on the covariance matrix. Information on GAUSSX is available at web site http://artsci-ccwin.concordia.ca/Economics/Breslaw/gaussx/gaussx1.html. The FORTRAN programs can be downloaded from Hajivassiliou's home page at http://ariadne.econ.columbia.edu/~vassilis/. Our GAUSS programs for mixed logits with error-component structures are available at http://elsa.berkeley.edu/~train. Our GAUSS program for probit with error-component structure using the GHK simulator is not as user-friendly as the mixed logit routines, but is available upon request.