ELSEVIER

Journal of Econometrics 89 (1999) 109–129

JOURNAL OF Econometrics

Forecasting new product penetration with flexible substitution patterns

David Brownstone[a], Kenneth Train[b,*]

[a] *Department of Economics, University of California, Irvine, USA*

[b] *Department of Economics, University of California, Berkeley, 549 Evans Hall # 3880, Berkeley, CA 94720-3880, USA*

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. © 1999 Elsevier Science S.A. All rights reserved.

JEL classification: C15; C35; L62; R41

Keywords: Mixed logit; Random-coefficients logit; Probit; Simulation

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

* Corresponding author. E-mail: train@econ.berkeley.edu; web site: http://else.berkeley.edu/~train

PII: S0304-4076(98)00057-8

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. However, 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.

2. 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 over people and alternatives, 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; β is a vector of parameters to be estimated which are fixed over people and alternatives; η_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 ε_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) α and $V(\eta)$ is general and can depend on underlying parameters and data. For standard logit, each element of ε is iid extreme value, and, more importantly, η 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.

2.1. Mixed logit: General distribution for η and extreme value for ε

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

$$L_i(\eta) = \exp(\beta' x_i + \eta_i) \Big/ \sum_j \exp(\beta' x_j + \eta_j). \tag{1}$$

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

$$P_i = \int L_i(\eta) f(\eta|\Omega)\, d\eta. \tag{2}$$

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 Ω, a value of η 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:

$$\mathrm{SP}_i = (1/R) \sum_{r=1,\dots,R} L_i(\eta^r), \tag{3}$$

where R is the number of replications (i.e., draws of η), η^r is the rth 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 $\ln(\mathrm{SP}_i)$ 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 numerical 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 β and Ω, 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 $\sum_n \ln(\mathrm{SP}_{ni})$ 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 earliest applications of mixed logit were apparently the automobile demand models of Boyd and Melman (1980) and Cardell and Dunbar (1980). These researchers used aggregate, market-share data rather than customer-specific choice data. As a result, the computationally difficult integration for the choice probabililities (which in their case where shares) needed to be performed only once rather than for each customer in a sample. Recent applications (cited below) have used customer-specific data. These applications have placed various structures on the mixed logit specification, or, more precisely, on the specification of f. In Section 3 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 μ is a random vector with zero mean that does not vary over alternatives and has density $g(\mu|\Omega)$ with parameters Ω; z_i is a vector of observed data related to alternative i; and ε_i is iid extreme value. This is a mixed logit with a particular structure for η, 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 μ 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 μ from its distribution and calculating $\eta_i = \mu' z_i$. Insofar as the number of error components (i.e., the dimension of μ) is smaller than the number of alternatives

[1] Note that even though SP_i is unbiased for P_i, $\ln(\mathrm{SP}_i)$ 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).

(the dimension of η), 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[2] to nested logit is obtained as follows. The alternatives are grouped into K nests, labeled $k = 1, \dots, K$. For nest k, define a dummy variable d_i^k that equals 1 if alternative i is in nest k, and zero otherwise. Define z_i as the vector composed of these dummy variables: $z_i = \{d_i^1, \dots, d_i^K\}'$. Finally, specify μ to be a vector of K iid deviates, such that $V(\mu)$ is diagonal with elements σ_k, $k = 1, \dots, K$. Then the unobserved portion of utility is correlated for any two alternatives within a nest and uncorrelated for alternatives in different nests: for any $i \neq j$, $E[(\mu' z_i + \varepsilon_i)(\mu' z_j + \varepsilon_j)] = \sigma_k$ if i and j are in nest k, and $= 0$ if i and j are in different nests. The pattern of correlation is therefore the same as in a nested logit. Constraining the elements of μ to have the same variance, i.e., $\sigma_k = \sigma$ for all k, is analogous to restricting the correlations within a nested logit model to be the same for all nests.[3] In both cases, the correlation is the same within all nests. Overlapping nests can be handled by allowing alternatives to enter more than one nest in the definition of z_i.

Mixed logits with error-components are more general than their specification and our illustration with nested logit might immediately suggest. Importantly, McFadden and Train (1997) show that any random utility model can be approximated arbitrarily closely by a mixed logit with an error-components structure and appropriate choice of the z_i's and g. This result 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

[2] By 'analog' we mean that the mixed logit contains the same pattern of correlation as a nested logit, namely, equal correlation between all pairs of alternatives within a nest and no correlation over alternatives in different nests. The models are not exactly the same because the distributions are different: unobserved utility in a nested logit follows a generalized extreme value distribution which cannot be obtained exactly as the sum of several random variables. However, as mentioned below, the proof of McFadden and Train (1996) indicates that a mixed logit can approximate a nested logit arbitrarily closely.

[3] In the terminology of nested logit models, this is equivalent to restricting the 'log-sum coefficients' to be the same for all nests.

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.

Most recent empirical work with mixed logits has been motivated by a random-parameters, or random-coefficients, specification (Allenby and Lenk, 1994;[4] Bhat, 1996a, b; Mehndiratta, 1996; Revelt and Train, 1998; Train, 1998). 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 β and deviations μ. 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.

Other types of mixed logits have also been used. Elrod (1988) and Erdem (1995) provide a factor-analytic structure to η. 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 Bolduc (1996) specify a 'general autoregressive' error structure, under which each η_i is correlated with each other η_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 η is normally distributed but the elements of ε 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 represent situations where η 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.

2.2. Mixed probits: General distribution for η and standard normal for ε

When the elements of ε are iid standard normal, then a family of models arises that is analogous to that discussed above for extreme value ε and with estimation performed by the same type of simulation. However, a mixed probit

[4] Allenby and Lenk include autocorrelated additive error terms in addition to random parameters.

requires one additional dimension of simulation relative to a mixed logit (Train, 1995.) In particular, the probability of alternative i conditional on η is $\text{Prob}(\beta' x_i + \eta_i + \varepsilon_i > \beta' x_j + \eta_j + \varepsilon_j \forall j \neq i) = \text{Prob}(\varepsilon_j - \varepsilon_i < \beta'(x_i - x_j) + \eta_i - \eta_j \; \forall j \neq i)$. In a mixed logit, this probability is the logit formula; however, in a mixed probit, it is a multi-dimensional integral since $\varepsilon_j - \varepsilon_i$ is correlated over j's due to the common influence of ε_i. Conditional on η and ε_i, the conditional choice probability for a mixed probit, labeled $M_i(\eta, \varepsilon_i)$, is a product of univariate cumulative normals, which is easy to calculate: $M_i(\eta, \varepsilon_i) = \Pi_{j \neq i} \Phi(\beta'(x_i - x_j) + \eta_i - \eta_j + \varepsilon_i)$, where Φ is the standard normal distribution. The choice probabilities for a mixed probit are therefore simulated as $\text{SP}_i = (1/R) \sum_{r=1,\ldots,R} M_i(\eta^r, \varepsilon_i^r)$.[5] Conditioning on ε_i adds an extra dimension of simulation relative to mixed logit; therefore, unless there is a reason to expect ε to be normal instead of extreme value, assuming ε to be extreme value seems preferable on pragmatic grounds for general distributions of η.

2.3. *Pure probits: Normal η and standard normal for ε*

When η is normal as well as ε, 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 $\xi = (\xi_1, \ldots, \xi_J)'$ is distributed normal with zero mean and covariance matrix Λ. Given our notation, $\xi_i = \eta_i + \varepsilon_i$, and any of the structures given above can be placed on η (provided the distributions are normal), which gives a structure to Λ. Hajivassiliou et al. (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 ξ (or more precisely, on draws of the difference between ξ_j for each non-chosen alternative and ξ_i for the chosen alternative) while the mixed logit simulator is based on draws of the random terms that compose η. 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; Keane, 1990) performed better than other probit simulators for the specifications that they examined. To our knowledge, there has been no comparison of probit simulators with the simulator for mixed logit with normal η. (As stated above, these are essentially the same models, with the only difference being that the iid term is normal in the pure probit and extreme value in this version of the mixed logit.) We provide a comparison on our data set. However, the advantages and limitations of

---

[5] This is essentially the Stern (1992) simulator generalized to allow for non-normal distributions of η.

each depend on the specific situation. For example, in situations where the dimension of the mixing distribution is less than the number of alternatives (as in Train, 1998, 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, 1998).

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

3. Estimated models

3.1. Data

We utilize the survey data collected and described by Brownstone et al. (1996) on households' attitudes and preferences regarding alternative-fuel vehicles. Their sample was identified using pure random digit dialing, geographically stratified into 79 areas covering most of urbanized California. An initial computer-aided telephone interview (CATI) was completed for each of 7387 households. This initial CATI collected information on household structure, vehicle inventory, housing characteristics, basic employment and commuting for all adults, and the household's intended next vehicle transaction.

The data from the initial CATI were used to produce a customized mail-out questionnaire for each sampled household. This questionnaire contained a stated preference (SP) discrete-choice experiment. SP data have been used extensively in marketing and demand studies, especially for new products for which historic data on market choices are unavailable (see, e.g., Green and Srinivasan, 1990; Hensher, 1992; Louviere, 1994.) An example SP task from the questionnaire is given in Fig. 1. Three hypothetical vehicles with different fuel types were described, each of which was available in either of two different body types. The respondents was asked to choose among the six options.[6] There were four possible fuel types (gasoline, compressed natural gas, methanol, and electric), three of which appeared in each SP question. There were 10–12

[6] 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 (1998) and Train (1998) for estimation of mixed logits with repeated choices for each decisionmaker.

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 Runs on electricity only	Natural Gas (CNG) Runs on CNG only	Methanol 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)

Fig. 1. Vehicle choice survey question.

attributes per vehicle per choice set, depending upon the fuel type of the vehicle. Four levels were used to cover the range of most attributes. Variation in the attribute levels was obtained through an orthogonal main effects design. Respondents were specially instructed to treat all non-listed attributes (e.g., maintenance costs and safety) as identical for all vehicles in the choice set.

4747 households successfully completed the mail-out portion of the survey, which constitutes a 66% response rate among the households that completed the initial CATI. A comparison with Census data reveals that the sample is slightly biased toward home-owning larger households with higher incomes. Of these 4747 respondents, 4654 had sufficient non-missing information to use in model estimation.

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 questionnaire described to the respondent the cost of recharging/refueling the vehicle at a station, as well as, for electric and compressed natural gas (CNG) vehicles, the cost of refueling 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 are three times 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 (multiplied by 0.1 for scaling.)[7]

3.2. Standard logit and mixed logit estimates

Column 1 of Table 2 gives the estimated parameters and standard errors for a standard logit model. Column 2 presents a mixed logit with the same specification for the non-stochastic 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 contains nests for non-electric and 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

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

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 30 mph from stop, in tens of seconds (e.g., 3 s is entered as 0.3)
Top speed	Highest speed that the vehicle can attain, in hundreds of miles/h (e.g., 80 mph is entered as 0.80)
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/mile is entered as 0.5.) 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 $< 5 \times$ EV	1 if respondent commutes less than five miles each day and vehicle is electric; zero otherwise
College $\times$ 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 $\times$ 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

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 ς_k is iid standard normal, z_{ki} are the four variables described above, and ε_i is iid extreme value. The parameters $\sigma_1, \ldots, \sigma_4$ 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 $\sigma_1, \ldots, \sigma_4$ 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.[8]

The error components enter significantly. Gas and methanol vehicles enter both the no-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. As a referee pointed out for us, significant variation in the value of luggage space is expected, since

[8] Different draws are taken for different respondents. McFadden (1989) indicates the importance, in the context of method of simulated moments estimation, of taking different draws for each observation: with different 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 canceling-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 canceling probably occurs even with fewer replications. Lee (1992) describes the properties of MSL estimators when the same draws are used for all observations.

households that use the vehicle for commuting would generally place a lower value on luggage space than those who use the vehicle primarily for vacations and out-of-town trips. The luggage space variable probably also served 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

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 $< 5 \times$ EV	0.198	0.082	0.372	0.166
College $\times$ 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 $\times$ 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

	Probit		Mixed logit B		
	Estimate	Std. error	Estimate		Std. error
Variables					
Price/ln(income)	− 0.184	0.031	− 0.286	− 5.999	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 $< 5 \times$ EV	0.294	0.136	0.359		0.163
College $\times$ 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 $\times$ 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	

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.

The estimated parameters that enter the non-stochastic 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 (1998).

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 coefficients (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 0.489 in both models. Bhat (1996a) and Train (1998) also found the ratios of coefficients not to differ significantly between a standard and mixed logit, while Bhat (1996b) found fairly substantial differences.

3.3. *Pure probit estimates*

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.[9]

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

[9] We performed an experiment to compare the relative accuracy (or, more precisely, the simulation variance) of the GHK probit simulator and the mixed logit simulator. We calculated the simulated log-likelihood function at the estimated parameters in Table 2 separately ten times using ten different sets of random draws (i.e., different seeds for the random number generator.) We clocked the time required for each calculation and computed the variance in the simulated log-likelihood value over these ten sets of draws. The GHK probit simulator with 50 replications took an average of 51.0 seconds for each calculation of the simulated log-likelihood function and obtained a variance of 26.9 in the value of the simulated log-likelihood over the ten sets of draws. The mixed logit simulator with 50 replications took 19.8 seconds and obtained a variance of 35.3. The GHK probit simulator therefore had lower variance than the mixed logit simulator with the same number of replications. However, the mixed logit simulator took considerably less computer time. A mixed logit simulator with 125 replication took 48.1 seconds (which is about the same as the GHK probit simulator with 50 replications) and obtained a variance of 11.6 (which is less than half that of the GHK probit simulator with 50 replications.) These results suggest that: for a given number of replications the GHK probit simulator has somewhat less variance than the mixed logit simulator, while for a given amount of computer time, the mixed logit simulator has considerably lower variance.

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. At convergence, the probit obtained a higher simulated log-likelihood value than the mixed logit. However, the difference is within the range expected simply from simulation variance.[10]

3.4. Mixed logit with additional error components

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 a 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.) In this specification, only one extra coefficient, beyond those four in the mixed logit model of

[10] Using the method described in the previous footnote, the simulation variance in the log-likelihood value for the probit with 50 replications is estimated as 26.9, and that for the mixed logit with 250 replications as 2.19. The simulation variance for the difference is therefore 29.1, such that the standard deviation is 5.4. The difference in simulated log-likelihood values is 7375.34-7368.74 = 6.6, which is 1.2 times the standard deviation of 5.4.

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, expensive 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 have 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 (1998) 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 kth coefficient with a log-normal distribution is specified as $\exp(b_k + s_k v_k)$, where v_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 $\exp(b_k + s_k v_k)$ 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 0.8326.) 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 simulated log-likelihood value for this model is nearly the same as that for the mixed logit without the extra variation. While the extra variation does not improve the fit of the model, it changes the substitution patterns that are forecast by the model. The differences are 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 (1998) 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.

3.5. Substitution patterns

The substitution patterns that are implied by the models can be compared through several examples. Suppose a small electric car is introduced to a base situation consisting of gas cars that range in size from small to large. The standard logit model, because of the iia property, implies that the new electric car will draw proportionately from all sizes of the gas cars. In contrast, the mixed logits and probit predict that the electric car will draw more proportionately from smaller gas cars than from larger gas cars. This more realistic substitution pattern is the consequence of the error component relating to size. It 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.

Suppose now that a large methanol car is added to the small electric and various sizes of gas cars. The logit model of course predicts proportionate switching. The mixed logits and probit predict disproportionate switching, with greater switching from the larger gas cars than the smaller gas cars, and with greater switching from the gas cars than the electric car. The later difference 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 difference 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.

The distinction between the two mixed logits is illustrated by considering the predicted impact of a rise in the price of large gas cars. As always, the standard logit predicts that households switch proportionately to each of the other size classes. The two mixed logits and the probit predict that households switch more readily to mid-size cars than to smaller cars, as one would expect in reality. However, the two mixed logits imply a different amount of switching in total. In particular, the mixed logit in column 4 of Table 2 predicts considerably less switching away from large cars in reaction to a price increase, than the mixed logit in column 2 or the probit. Recall that this model, unlike the other mixed logit or the probit, includes variation in the price coefficient over households. Households who place relatively little importance on price have a greater tendency to buy an 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, the mixed logit with variation in the price coefficient would predict a smaller share of customers switching away from large cars when the price is raised. This example illustrates the flexibility of mixed logits to represent various substitution patterns.[11]

Acknowledgements

David Bunch and Tom Golob 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 made in this paper.

References

Allenby, G., Lenk, P., 1994. Modeling household purchase behavior with logistic normal regression. Journal of the American Statistical Association 89, 1218–1231.

Ben-Akiva, M., Bolduc, D., 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., Lerman, S., 1985. Discrete Choice Analysis. MIT Press, Cambridge, MA.

Bhat, C., 1996a. Accommodating variations in responsiveness to level-of-service measures in travel mode 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., Melman, R., 1980. The effect of fuel economy standards on the U.S. automotive market: An hedonic demand analysis. Transportation Research 14A, 367–378.

[11] GAUSS software to estimate mixed logits is available for downloading, along with a users' manual and sample runs, from Train's web site at http://elsa.berkeley.edu/ ~ train.

Brownstone, D., Bunch, D., Golob, T., Ren, W., 1996. Transactions choice model for forecasting demand for alternative-fuel vehicles. Research in Transportation Economics 4, 87–129.

Bunch, D., Bradley, M., 1995. Personal vehicle survey model results: Modeling reactions to hypothetical future alternative and conventional fuel vehicles. Working paper, Graduate School of Management, University of California, Davis.

Cardell, N., Dunbar, F., 1980. Measuring the societal impacts of automobile downsizing. Transportation Research 14A, 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 2, 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 23rd Symposium on the Interface. American Statistical Association, Alexandria, VA.

Green, P., Srinivasan, V., 1990. Conjoint analysis in marketing research: New developments and directions. Journal of Marketing 54, 3–19.

Hajivassiliou, V., McFadden, D., 1990. A method of simulated scores for the estimation of LDV models. Cowles Foundation Discussion paper #967, Yale University.

Hajivassiliou, V., McFadden, D., Ruud, P., 1992. Simulation of multivariate normal rectangle probabilities and their derivatives: theoretical and computational results. Cowles Foundation Discussion paper #1021, Yale University.

Hajivassiliou, V., Ruud, P., 1994. Classical estimation methods for LDV models using simulation. In: Engle, R., McFadden, D., (Eds.), Handbook of Econometrics, vol. IV. Elsevier, New York.

Hensher, D., 1992. Stated-preference analysis of travel choices: The state of the practice. Transportation 21, 107–133.

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 8, 518–552.

Louviere, J., 1994. Comjoint analysis. In: Bagozzi, R. (Ed.), Advanced Methods of Marketing Research. Blackwell Publishers, Cambridge, MA.

McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior. In Zarembka, P. (Ed.), Frontiers in econometrics. Academic Press, New York.

McFadden, D., 1975. On independence, structure and simultaneity in transportation demand analysis. Working paper # 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 Karquist, A. et al. (Eds.), Spatial Interaction Theory and Planning Models. North-Holland, Amsterdam.

McFadden, D., 1989. A method of simulated moments for estimation of the multinomial probit without numerical integration. Econometrica 57, 995–1026.

McFadden, D., Train, K., 1997. Mixed MNL models of discrete choice. Working paper, Department of Economics, University of California, Berkeley.

McFadden, D., Ruud, P., 1994. Estimation by simulation. Review of Economics and Statistics 76, 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., Train, K., 1998. Mixed logit with repeated choices: Households' choices of appliance efficiency level. Review of Economics and Statistics, 80 (4).

Stern, S., 1992. A method for smoothing simulated moments of discrete probabilities in multinomial probit models. Econometrica 60, 943–952.

Train, K., 1980. The potential market for non-gasoline-powered automobiles. Transportation Research 14A, 405–414.

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

Train, K., 1995. Simulation methods for probit and related models based on convenient error partitioning. Working paper, Department of Economics, University of California, Berkeley.

Train, K., 1998. Recreation demand models with taste differences over people. Land Economics, 74(2) 230–239.