PERGAMON

Transportation Research Part B 34 (2000) 315–338

TRANSPORTATION
RESEARCH
PART B

www.elsevier.com/locate/trb

Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles

David Brownstone [a,*], David S. Bunch [b], Kenneth Train [c]

[a] Department of Economics, University of California, Irvine, CA 92697-5100, USA
[b] Graduate School of Management, University of California, Davis, USA
[c] Department of Economics, University of California, Berkeley, USA

Accepted 2 June 1999

Abstract

We compare multinomial logit and mixed logit models for data on California households' revealed and stated preferences for automobiles. The stated preference (SP) data elicited households' preferences among gasoline, electric, methanol, and compressed natural gas vehicles with various attributes. The mixed logit models provide improved fits over logit that are highly significant, and show large heterogeneity in respondents' preferences for alternative-fuel vehicles. The effects of including this heterogeneity are demonstrated in forecasting exercises. The alternative-fuel vehicle models presented here also highlight the advantages of merging SP and revealed preference (RP) data. RP data appear to be critical for obtaining realistic body-type choice and scaling information, but they are plagued by multicollinearity and difficulties with measuring vehicle attributes. SP data are critical for obtaining information about attributes not available in the marketplace, but pure SP models with these data give implausible forecasts.

1. Introduction

Forecasting the demand for new products or transportation innovations requires information about consumers' preferences for products or services that do not exist in the current marketplace. Researchers have overcome this problem by designing stated preference (SP) experiments to measure consumers' preferences over hypothetical alternatives including new products. SP data have been subject to considerable criticism by economists and other researchers because of a belief that consumers react differently to hypothetical experiments than they would facing the same

* Corresponding author. Tel.: +1-949-824-6231; fax: +1-949-824-2182.
E-mail address: dbrownst@uci.edu (D. Brownstone).

PII: S0191-2615(99)00031-4

alternatives in a real market. One problem is that some attributes for "totally new products" might be novel enough that respondents do not completely understand them. This would introduce components related to both uncertainty and perceived risk that would affect the outcome of choice modeling efforts. Another problem that could be particularly severe arises when new products incorporate "politically correct" public good attributes such as "zero-pollution" electric vehicles. Respondents may misrepresent their choices in SP experiments to strategically signal their preference for provision of the public good (less pollution), although in reality they would not spend extra money on purchasing an electric vehicle (possibly because of the obvious free-rider problem).

However, many difficulties also arise in using revealed preference (RP) data to develop forecasting models. There are frequently high collinearity and limited variation among attributes in real markets. For the vehicle choices modeled in this paper there are additional problems with defining choice sets and the need to link physical attributes from external databases. The resulting data can then only approximate the actual choice situations faced by vehicle purchasers. Since the number of vehicle make/model/year combinations in the US vehicle market is huge, some sampling of alternatives is necessary to use discrete choice models. This sampling to produce choice sets introduces additional noise into the resulting models, and may bias estimates in more flexible alternatives to the standard multinomial logit model (MNL). Under these difficult conditions RP model estimates are often unstable, and can have theoretically incorrect signs.

One potential solution to these problems is to develop and estimate joint models to exploit the advantages of each type of data while mitigating the weaknesses. This paper describes models combining SP and RP vehicle choice data where the SP alternatives include electric, compressed natural gas (CNG), and methanol fueled vehicles that are not yet widely available in the marketplace. These data were collected as part of a larger project to build a microsimulation model of the California vehicle market. The SP data come from the first wave of a panel study initiated in mid-1993. The second wave occurred approximately 15 months later, at which time households were re-interviewed, allowing the collection of RP data on vehicle transaction behavior. The data set is discussed in more detail in Section 2.

The Wave 1 SP data used in this paper have already been used to build a large multinomial logit (MNL) model of alternative-fuel vehicle choice (Brownstone et al., 1996) which is incorporated in a microsimulation model of the vehicle market for the greater Los Angeles area (roughly 10% of the US vehicle market). For a discussion of this microsimulation forecasting system, see Bunch et al. (1996). More recently, Brownstone and Train (1998) used these SP data to compare MNL and "mixed logit" models where random error components are added to the MNL specification. They found strong evidence that the MNL specification is not appropriate for these data, and they demonstrated that there are large differences between forecasts based on the different specifications.

This paper extends the analysis in Brownstone and Train (1998) to jointly model SP and RP vehicle choices. Previous methodological work on combining SP and RP data have focused on the problems caused by scaling differences and the correlation in unobserved attributes across repeated choices by the same decision makers. We develop simple mixed logit specifications that easily incorporate unobserved correlation and scaling differences, although there is no evidence of unobserved correlation between SP and RP choices in our models. These mixed logit specifications are statistically superior to the "standard" joint scaled logit models previously used for these applications. The mixed logit models also yield very different forecasts for a policy experiment designed to simulate the early stages of alternative-fuel vehicle availability. These policy simulations show even larger differences between the pure SP and joint RP/SP models, which highlights the importance of jointly modeling SP and RP choices to exploit the strengths and avoid the weaknesses of each type of data.

Section 2 reviews the data sources. Section 3 reviews the general mixed logit model and joint RP/SP estimation. Section 4 gives estimation results for SP, RP and joint mixed logit models for vehicle choice. We then give results of some forecasting experiments in Section 5 that highlight the different substitution patterns between the MNL and mixed logit specifications.

2. Data

The SP and RP choice data used in the next sections were collected as part of a multi-wave panel survey carried out in California, starting in June 1993. The initial household sample was identified using pure random digit dialing and was 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. The survey also asked for information about the household's most-likely next vehicle transaction. If the next transaction were likely to involve a purchase, the survey asked for the body type, size, and approximate purchase price (including whether new or used). These data were used to produce a more detailed, customized mail-out questionnaire that was then sent by express delivery, along with an incentive (five dollars).

The customized mail-out questionnaire asked more detailed questions about each household member's commuting and vehicle usage, including information about sharing vehicles in multiple-vehicle and multiple-driver households. The information on the next intended vehicle transaction was used to create two customized SP vehicle-choice questions (discussed below) that contained hypothetical alternative-fuel and gasoline vehicles. After the households received the mail-out questionnaires, they were again contacted for a final CATI. This interview collected all the responses to the mail-out questions. Additional questions about the household's attitudes towards alternative-fuel vehicles were also included at the end of this interview. Taken together, questions from both CATIs comprise the Wave 1 survey of the panel study.

The 4747 households that successfully completed the mail-out portion of the Wave 1 survey in 1993 represent 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. Eighty percent of the households in the sample had exactly one driver per vehicle, showing that, in California, the number of drivers is the most important determinant of the vehicle ownership level. For two-vehicle households, a little over one-third of the vehicles are driven 10 000 miles per year or less, a third are driven 10 000–15 000 miles per year, and almost a third are driven more than 15 000 miles per year.

Models estimated in this paper use data from the Wave 1 SP vehicle-choice experiment, which we now describe. Each vehicle-choice question used the format given in Fig. 1. It is important to note that Fig. 1 gives a specific example that is only one of many possibilities: experimental design

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

	Vehicle A	Vehicle 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 recharge 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 methanol
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 vehicle	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. SP Vehicle choice survey question.

methods combined with household-specific customization ensured that, quite literally, no two vehicle choice questions in the survey were alike. Given the potential complexity of the choice task (and the length of the overall survey), each household was only asked to complete two questions of the type shown in Fig. 1.

The purpose of the experiment was to estimate preferences for vehicle attributes related to four possible fuel types: gasoline, compressed natural gas (CNG), methanol, and electric (EV). In the Fig. 1 format there are three vehicle columns available, each corresponding to a different fuel type. In our experiment three of the four fuel types appear in each SP question, giving six possible fuel-type format combinations (e.g., in Fig. 1 the combination is electric, CNG, and methanol). Each household was assigned two of the possible six combinations at random (ordering was also randomized). In addition, as part of the design process (described below) each column was assigned two possible body types, giving a total of six vehicle types (defined by the combination of fuel and body type).

Producing vehicle "profiles" requires assigning attribute descriptions to all the appropriate cells in the Fig. 1 format. However, note that attributes and their levels are clearly a function of fuel type, due to expected differences in technologies. Attributes may exist for some vehicles and not for others. For example, all electric vehicles were assumed to have home recharging whereas all gasoline vehicles were assumed to refuel exclusively at gas stations; hence, electric vehicles require home refueling times and costs, but these attributes do not exist for gasoline vehicles. In addition, attribute ranges might be expected to differ by fuel type. For example, refueling/recharging ranges are expected to be lower for electric vehicles than for gasoline vehicles.

To address these issues, we established "design translator" tables to define candidate attribute levels as a function of fuel type and also customization requirements (e.g., purchase price ranges, body type requirements). (The size of these tables precludes including them here.) In general, we used up to four attribute levels to cover the range of possibilities, allowing estimation of possible non-linear effects for quantitative attributes. The vehicle profiles for a specific question were constructed by combining the appropriate design translators with a randomly chosen row from an experimental design matrix. Respondents were specifically instructed to treat all non-listed attributes (e.g., maintenance costs and safety) as *identical* for all vehicles in the choice set.

In this paper we use only one SP choice per household, corresponding to the first SP question in each survey. The primary reason for this was that resource constraints precluded cleaning and coding the second SP choice question. However, if both SP choices were included in the data, the issue of unobserved error correlation across repeated choices would become relevant. We note that mixed logit specifications can easily accommodate repeated choices. See, e.g., Revelt and Train (1998).

Approximately 15 months after the Wave 1 survey, a geographically stratified sample of the approximately 7300 households who completed the first telephone interview was used for a second wave ("Wave 2") of interviewing. After excluding motor homes, motorcycles, and heavy trucks, 874 out of the 2857 households surveyed for this reinterview reported at least one vehicle purchase since the first interview. An RP data set was constructed using these purchases, as we now describe.

Households were asked for detailed information about each vehicle transaction that occurred between the Wave 1 and Wave 2 interviews. In this paper we focus on the choice of vehicle purchased to investigate aspects of using mixed logit models for SP/RP estimation. Models are developed using a classification scheme similar to that described in Brownstone et al. (1996). For each model year beginning usually in 1974, all vehicles are classified according to 13 body type/size categories (see Table 5 for definitions), and each of these categories are further subdivided into a high and low purchase price group and finally subdivided into a domestic and import group. We therefore have 689 categories approximating the universe of new and used vehicles from which respondents made their RP choices. For each of these categories we have: new and current used price, fuel economy, range, top speed, acceleration time (0–30 miles per hour), number of models in the class, luggage volume, emissions index (proportion relative to new 1996 gasoline vehicles of same body/size class), and maintenance costs. Due to missing and erroneous vehicle type data in our survey, we are able to match these attribute data for 607 of the 874 respondents who reported a vehicle transaction between the survey waves.

In addition to the data described above, additional SP tasks were given to the 2857 Wave 2 respondents. These tasks have more attributes than the Wave 1 SP design analyzed in this paper, and they have 17 vehicles per experiment instead of 6 in the Wave 1 design. Future work will add these data to the models described in the following sections.

The data used in this paper represent an extension and improvement over the more preliminary versions of the data used in Brownstone and Train (1998), which were limited models for the Wave 1 SP. The improvements come from implementing editing and consistency checks across the Wave 1 and Wave 2 data for, e.g., demographic variables, and the extensions are possible due to the availability of RP choices from the Wave 2 survey.

3. Mixed logit models and RP/SP joint estimation

A person faces a choice among J alternatives, which will be modeled using a random utility framework. For purposes of this paper we assume without loss of generality that 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 over alternatives and people. In particular, the utility to person n from alternative i is denoted

$$U_{in} = \beta' x_{in} + [\eta_{in} + \varepsilon_{in}],$$

where x_{in} is a vector of observed variables relating to alternative i and person n; β is a vector of structural parameters which characterizes choices by the overall population; η_{in} 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 person n; and ε_{in} is a random term with zero mean that is iid over alternatives and does not depend on underlying parameters or data. For any specific modeling context, the variance of ε_{in} may not be identified separately from β, so it 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. Taken together, these assumptions give rise to the Independence from Irrelevant Alternatives (IIA) property and its restrictive substitution patterns.

The Mixed Logit *class* of models assumes a general distribution for η and an iid extreme value distribution for ε. Denote the density of η by $f(\eta|\Omega)$ where Ω are the fixed parameters of the distribution. (The density f may also depend upon explanatory data for people and alternatives, but in what follows this is suppressed for notational convenience.) For a given value of η, the conditional choice probability is simply logit, since the remaining error term is iid extreme value

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

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

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

Models of this form are called "mixed logit" because 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

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

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, so 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. Using the subscript n to index sampled individuals, and denoting the chosen alternative for each person by i, the log-likelihood function $\sum_n \ln(P_{in})$ is approximated by the simulated log-likelihood function $\sum_n \ln(\mathrm{SP}_{in})$ and the estimated parameters are those that maximize the simulated log-likelihood function. Lee (1992) derives 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 is convenient for implementing the search for the maximum:

$$G(\beta) \equiv \partial \sum_n \ln(\mathrm{SP}_{ni})/\partial\beta = \sum_n [1/\mathrm{SP}_{ni}](1/R)\sum_r L_{ni}(\eta_n^r)\left[\sum_j (d_{nj} - L_{nj}(\eta_n^r))x_{nj}\right],$$

$$G(\Omega) \equiv \partial \sum_n \ln(\mathrm{SP}_{ni})/\partial\Omega = \sum_n [1/\mathrm{SP}_{ni}](1/R)\sum_r L_{ni}(\eta_n^r)\left[\sum_j (d_{nj} - L_{nj}(\eta_n^r)(\partial\eta_n^r/\partial\Omega)\right],$$

where $d_{nj} = 1$ for $j = i$ and zero otherwise. The derivative $\partial\eta_n^r/\partial\Omega$ depends on the specification of η and f. Also, if the same parameters enter β and Ω (as in the third model in Section 4), the gradient is adjusted accordingly.

Analytic second derivatives can also be calculated. However, in contrast to the standard MNL model with its globally concave log-likelihood function, the inclusion of the Ω structural parameters removes the guarantee of global concavity, and the Hessian matrix is not guaranteed to be positive definite. This creates a more complicated situation for the iterative search, e.g., Revelt and Train (1998) 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. To address this problem, we implemented specialized estimation code using the Bunch et al. (1993) optimization software. These methods are more robust, and generally converge in many fewer iterations than the more standard numerical procedures (see Bunch, 1988). Although the number of iterations makes little practical difference when estimating MNL models, this is not longer true when using computationally intensive simulation approaches for calculating choice probabilities and gradients.

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 4, 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 $\eta_i = \mu' x_i$. Insofar as the number of error components (i.e., the dimension of μ) is smaller than the number of alternatives (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 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 and Train

(1997) have 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. McFadden and Train (1997) also gives Lagrange Multiplier tests for the presence of significant random error components in MNL models. Our experience with these tests for the specifications in Section 4 shows that they are easy to calculate and appear to be quite powerful omnibus tests. However, they are not as good for identifying which error components to include in a more general mixed logit specification.

Most recent empirical work with mixed logits has been motivated by a random-parameters, or random-coefficients, specification (Bhat, 1996a,b; Mehndiratti, 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. In different contexts one or the other interpretation will seem more natural.

The random-coefficients interpretation is useful when considering models of repeated choices by the same decision maker. The most straightforward version is a model for which the same draws of the random coefficient vectors are used for all repeated choices. This specification does not lead to perfect error correlations because the independent extreme value term ε_i still enters the utilities for each choice. The error correlation across repeated choices therefore increases as the variance of the random coefficients increases. A feasible (but computationally more demanding) model that might be more appropriate for panel data would be to specify a first-order autoregressive process for the random coefficients. This more general model would permit the error correlation to decrease over time.

In our survey data we have two SP observations and one RP observation for some households, and the error correlation due to "repeated choices" and preference heterogeneity could be addressed as just described. However, an additional issue must be considered when jointly estimating a model containing both RP and SP choices. Although the error generation process for a collection of (repeated) SP choices in a controlled experiment might be expected to be the same, it is likely to be different from the process producing the RP choice data. In particular, the effect of unobserved variables is likely to produce different variances for the ε_{in} terms in the two data sets. In this case the variance of one data set must still be normalized to unity, but the relative variance (or "scale") for the remaining data set is identified and can be estimated. By convention, the RP data are assumed to reflect the "correct" scale associated with the "real market". An "SP scale" coefficient is then defined as the multiplicative factor applied to all of the SP data to equalize the variances of the stochastic portion of the utility functions. Because scale and variance have a reciprocal relationship, values less than one imply that the SP stochastic variance is larger than the RP stochastic variance component.

Various approaches to estimating the scale have been discussed in the literature. The "low-tech" solution is to simply rescale the SP data so that the magnitude of key coefficients is similar before fitting joint MNL models. With a bit more effort, the SP data could be iteratively rescaled until the joint likelihood is maximized (see, e.g., Swait and Louviere, 1993). More recent work (see

Ben-Akiva and Morikawa, 1997; Hensher and Bradley, 1993) estimates the scaling parameter jointly with the model coefficients. This may be done directly, or by using a specification "trick" in a nested MNL estimation routine. Our estimation code directly implements the case of multiple data sets with different scales so that all parameters are estimated simultaneously in the FIML search. [1]

Once scale differences are taken into account, the most ideal circumstances would yield a specification where the remaining structural parameters are the same for the two data sets. Unfortunately this is unlikely in a complex joint RP/SP estimation (see the discussion in Section 4), and analysis will generally be required to identify which parameters can be "pooled" across the two data sets, and which parameters must be estimated in a data-set-specific manner. We identified our specifications in the next section using standard likelihood ratio tests against a model with no pooled coefficients.

4. Model specifications

This section gives estimates for various MNL and mixed logit specifications of RP, SP and joint RP/SP models of vehicle choice. All of the specifications use subsets of the variables defined in Table 1. One notable feature of our problem is that preferences for certain attributes are only identified by one of the two data sets. Specifically, preferences for Station Availability, Station Wagon, EV, CNG, and Methanol are only identified in the SP data; preferences for Import, number of models, and Used/Vintage are only identified in the RP data. The remaining attributes (in various forms) appear in both data sets.

In addition to the models presented in this section, we examined a number of other specifications to find the most consistent framework for joint RP/SP modeling. One important issue was the level of detail at which to define vehicle body-types and classes. In the final specification we pool together certain combinations body-type-and-size classes (e.g., Van = Minivan + Standard Van, SmallCar = Mini + Subcompact + Compact). Final variable definitions are reflected in Table 1.

4.1. Stated preference models

The Multinomial Logit SP model in the first three columns of Table 2 was estimated using one SP response from each household that completed the 1993 (Wave 1) mail-out survey for which clean data were available, giving a total of 4656 responses. The starting point for this analysis was a model in a previous paper by Brownstone and Train (1998). The final specification in this paper requires a slightly different set of body type definitions to provide a consistent basis for joint

[1] For code that has been designed to estimate mixed logit models for a single data set, the scale for a second data set can be estimated through a computational "trick" if the code allows parameter restrictions to be imposed. A set of alternative-specific constants is added to each SP alternative, and the mean coefficients of these constants are constrained to equal zero while their standard deviations are constrained to be equal. Of course, this "trick" constrains the variance of the SP extreme value errors to be larger than the RP alternatives. If the RP variance is larger, then alternative-specific constants could be added to the RP alternatives instead of the SP alternatives. Our experience with this "trick" shows that it is computationally much slower than customized maximum likelihood code.

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. Mean household income is \$38 000. Range: 0.1–45, Mean: 4
Operating cost	Fuel cost per mile of travel, in cents per mile. For electric vehicles, cost is for home recharging. For other vehicles, cost is for station refueling. Range: 1–12, Mean: 5.3
Range	Hundreds of miles that the vehicle can travel between refuelings/rechargings. Range: 0.5–5.7, Mean: 3
Range squared	Range × Range
Acceleration	Seconds required to reach 30 mph from stop. Range: 2–6.2, Mean: 3.9
Top speed	Highest speed that the vehicle can attain, in hundreds of miles per hour (e.g., 80 mph is entered as 0.80). Range: 0.55–1.55, Mean: 1.0
Luxury	1 if vehicle is a "luxury" model, zero otherwise
Import	1 if vehicle has an import nameplate, zero otherwise
Log (models)	Natural logarithm of number of vehicles in class. Range 0–3.6, Mean 0.72
New	1 if vehicle is new; zero otherwise
Used 1	1 if vehicle is one year old, zero otherwise
Log (age)	Natural logarithm of vehicle age for used vehicles
Pollution	Tailpipe emissions as fraction of comparable 1995 new gas vehicle. Range: 0–6.1, Mean 1.5
Station availability	Fraction of stations capable of refueling/recharging the vehicle. Range: 0.1–1.0, Mean: 0.85
Small car	1 for compact, subcompact, and mini cars, zero otherwise
Sports utility vehicle	1 for compact and full size sports utility vehicle, zero otherwise
Mini sports utility	1 for mini sports utility vehicle, zero otherwise
Sports car	1 for sports car, zero otherwise
Sports car × HHG3	1 for sports car if household size is greater than or equal three, zero otherwise (23% of sample have household size greater than or equal to 3)
Station wagon	1 for station wagon, zero otherwise
Truck	1 for compact or standard pickup trucks, zero otherwise
Van	1 for mini or standard van, zero otherwise
Minivan × HHG3	1 for minivan if household size is greater than or equal three, zero otherwise
Constant for EV	1 for electric vehicle, zero otherwise
College × EV	1 if respondent had some college education and vehicle is electric; zero otherwise. 41% of sample have some college education
Electric truck	1 if electric powered truck, zero otherwise
Electric sports car	1 if electric powered sports car, zero otherwise
Constant for CNG	1 for compressed natural gas vehicle, zero otherwise
Constant for methanol	1 for methanol vehicle, zero otherwise

RP/SP modeling. The "base" vehicle class was "midsize/large" car, and gasoline was the "base" fuel type.

The MNL coefficients for the generic attributes (price, operating cost, range, acceleration, and top speed) are all significant with the expected signs. Range enters in a quadratic specification, showing that respondents value an increase in range more highly when starting from a lower base. The MNL fuel type coefficients show that respondents prefer CNG and Methanol to gasoline (all else equal), but only college-educated respondents prefer electric vehicles. However, respondents

Table 2
Stated preference models

Variable	Multimomial logit log likelihood = −7343.28			Normalized coefficients		Mixed logit log likelihood = −7302.24		
	Coefficient	Std. error	*t*-stat	MNL	ML	Coefficient	Std. error	*t*-stat
Price/ln(income)	−0.184	0.027	−6.9	−3.65	−3.65	−0.503	0.120	−4.2
Operating cost	−0.076	0.007	−10.4	−1.51	−1.71	−0.236	0.052	−4.5
Range	0.493	0.110	4.5	9.80	12.89	1.779	0.500	3.6
Range squared	−0.034	0.025	−1.4	−0.67	−1.29	−0.178	0.079	−2.2
Acceleration	−0.064	0.011	−5.8	−1.27	−1.10	−0.151	0.041	−3.7
Top speed	0.262	0.080	3.3	5.19	4.58	0.632	0.244	2.6
Pollution	−0.302	0.092	−3.3	−5.99	−4.99	−0.689	0.254	−2.7
Station availability	0.309	0.084	3.7	6.13	6.63	0.914	0.298	3.1
Small car	−0.084	0.044	−1.9	−1.67	−0.48	−0.066	0.073	−0.9
Sports utility vehicle	0.874	0.146	6.0	17.36	6.84	0.944	0.152	6.2
Mini sports utility	−0.037	0.353	−0.1	−0.73	2.66	0.367	0.417	0.9
Sports car	0.925	0.185	5.0	18.36	7.89	1.088	0.205	5.3
Sports car × HHG3	−0.845	0.378	−2.2	−16.77	−7.77	−1.072	0.389	−2.8
Station wagon	−1.430	0.066	−21.8	−28.40	−11.07	−1.527	0.068	−22.3
Truck	−0.999	0.061	−16.4	−19.85	−8.12	−1.120	0.068	−16.5
Van	−1.150	0.070	−16.5	−22.85	−8.76	−1.209	0.076	−15.9
Minivan × HHG3	0.994	0.107	9.3	19.74	8.57	1.183	0.120	9.9
Constant for EV	−0.007	0.116	−0.1	−0.14	−10.01	−1.382	0.660	−2.1
College × EV	0.272	0.083	3.3	5.41	6.65	0.917	0.350	2.6
Electric truck	−0.259	0.128	−2.0	−5.15	−2.18	−0.300	0.139	−2.2
Electric sports car	−0.461	0.234	−2.0	−9.15	−2.97	−0.409	0.383	−1.1
Constant for CNG	0.237	0.079	3.0	4.72	3.06	0.422	0.260	1.6
Constant for methanol	0.412	0.071	5.8	8.19	8.53	1.177	0.319	3.7
Std. dev. Gasoline						2.156	0.729	3.0
Std. dev. EV						5.157	1.294	4.0
Std. dev. CNG						3.663	0.982	3.7
Std. dev. methanol						1.333	0.918	1.5
Std. dev. fuelcost						0.579	0.145	4.0

did not like electric pickup trucks or sports cars. It is interesting to note that vehicle manufacturers are currently trying to sell these electric vehicle types.

The last three columns of Table 2 give the estimates for the best fitting SP mixed logit specification. The normally distributed random coefficients were initially detected using the Lagrange multiplier test from McFadden and Train (1997). This test indicated that there were significant random components for the fuel types, price, operating cost, and a few body types. After fitting the indicated mixed logit model, we only found significant error components for the operating cost, gasoline, EV, CNG, and Methanol variables.

To be precise, the stochastic portion of a household's utility for alternative i is defined as $[\sum_{k=1-5} \sigma_k(\zeta_k z_{ki})] + \varepsilon_i$ where ζ_k is iid standard normal, z_{ki} are the five variables described above, and ε_i is iid extreme value. The parameters σ_k for $k = 1$–5 are estimated (see the rows beginning with "Std. Dev." at the bottom of Table 2); each denotes the standard deviation of the normal deviate that generates that error component. In simulating the choice probability for a respondent, five numbers are drawn from a random-number generator for the standard normal distribution; the five "variables" $\zeta_1 z_{1i} - \zeta_5 z_{5i}$ are created; and the conditional probability is evaluated with coefficients σ_k k =1–5 for the five "variables". This process is repeated for numerous draws and the conditional probabilities are averaged to obtain the simulated probability. We used 1000 draws to estimate the mixed logit models in this paper. Experimentation with 250 and 500 draw showed that more draws were needed to obtain numerically reliable estimates and likelihood values with these data.

In previous unpublished work with these SP data, nested multinomial logit models were estimated in which significant nesting for EV, CNG, and Methanol fuel-types (versus gasoline) was observed. This illustrates how mixed logit models with variance components may model substitution patterns similar to those from nested logit models, as discussed in Section 3. Brownstone and Train (1998) used a different specification, with components for Size, Luggage Space, Non-EV and Non-CNG. The latter two components carry similar information to those captured by EV, CNG and Methanol, but the goodness of fit using the current specification is much better.

In addition to the more "traditional" fuel-type error components, the mixed logit specification can also capture the importance of preference heterogeneity on operating cost sensitivity: this would not be possible with standard nested logit models. Unfortunately, the relatively large error component for operating cost implies that the model will generate an (implausible) positive price effect for one third of the respondents. This problem might be circumvented by specifying a log-normal distribution for this random component, but such a restriction might also reduce the goodness of fit. Better approaches to dealing with these sorts of variance-component specification issues will no doubt be developed in the near future, as researchers start to gain experience using mixed logit models.

The mixed logit coefficient estimates in Table 2 show that the error components are both statistically and practically important. The standard deviations for the fuel type coefficients are quite large and indicate a wide range of negative and positive preferences for these alternative fuels. This large heterogeneity in taste for alternative-fuel vehicles suggests that models with more interactions between demographics and the alternative-fuel dummy variables might perform better. However, our preliminary investigations on those demographic variables that can be readily forecasted (e.g., income, age, household size) did not find additional significant interaction terms, which suggests that a substantial portion of the observed heterogeneity is due to other factors, such as behavioral differences in anticipated vehicle usage, respondents' uncertainty and different information about alternative-fuel vehicles.

A useful feature of the mixed logit specification is that MNL is a nested special case, allowing formal comparison of the models on the basis of likelihood ratio statistics. The likelihood ratio statistic for mixed logit versus MNL is 82.08 with five degrees of freedom, which is highly significant. Since the stochastic portion of utility has different variances in the MNL and mixed logit specifications, the coefficients must be normalized before they can be meaningfully compared. The "Normalized Coefficients" column normalizes the coefficients by dividing by the price coefficient divided by the natural log of median income in thousands (which is approximately \$38 000 in this sample). These normalized coefficients can be conveniently interpreted as the average amount that a respondent with median income would be willing to pay for an additional unit of a particular attribute. For example, the MNL estimates in Table 2 imply that the sample households with \$38 000 incomes are willing to pay \$600 to reduce tailpipe pollution by 10 percent, whereas the comparable figure for mixed logit is \$500. Note that some of the MNL body type coefficients are implausibly large, but mixed logit estimates give lower and more plausible body-type tradeoffs. The mixed logit estimates also show an average negative view of electric vehicles, which differs from the MNL results.

4.2. Revealed preference models

Table 3 gives estimates for the best MNL model using actual vehicle purchases reported by households that participated in the Wave 2 survey, i.e., observed vehicle purchases occurring between the first and second panel waves. For those households that made multiple purchases during this period, only the first purchase was used for modeling. Although the Lagrange multiplier test found significant error components for price and operating cost, we were unable to estimate any mixed logit models with log likelihood values significantly better than the MNL model in Table 3. It is likely that a larger sample size would reveal significant error components, but currently we are limited to the 607 observations with complete data.

The number of vehicle types potentially available for purchase in real markets is very large, containing thousands of make and models and many vintages. Even using a vehicle classification scheme produces a very large "universal choice set". In this application, we have adopted a 689-level classification scheme according to vintage, body type, size, import/domestic, and price level.

Table 3
Revealed preference models

Multinomial logit	Log likelihood = −1788.46 (607 observations)			
Variable	Coefficient	Std. error	t-stat	Normalized coefficients
Price/ln(income)	−0.337	0.064	−5.3	−3.90
Operating cost	−0.193	0.068	−2.8	−2.24
Range	2.482	1.605	1.5	28.76
Range squared	−0.254	0.180	−1.4	−2.94
Acceleration	−0.329	0.381	−0.9	−3.82
Top speed	1.249	2.629	0.5	14.47
Luxury	−0.280	0.207	−1.4	−3.24
Import	−0.261	0.129	−2.0	−3.02
Log (models)	0.694	0.081	8.6	8.04
New	1.073	0.251	4.3	12.43
Used 1	0.466	0.250	1.9	5.40
Log (age)	−0.261	0.167	−1.6	−3.02
Pollution	0.399	0.099	4.1	4.63
Small Car	−0.454	0.152	−3.0	−5.26
Sports utility vehicle	0.390	0.390	1.0	4.52
Mini sports utility	−1.184	0.792	−1.5	−13.71
Sports car	−0.733	0.271	−2.7	−8.49
Sports car × HHG3	0.845	0.306	2.8	9.78
Truck	−0.392	0.406	−1.0	−4.55
Van	−0.358	0.435	−0.8	−4.14
Minivan × HHG3	0.883	0.270	3.3	10.24

The specific vehicle purchased by each household was matched to this classification scheme to identify a chosen alternative. Therefore each respondent's RP choice is modeled as a discrete choice from among 689 alternatives. Unfortunately, estimating models with choice sets of this size creates a host of computational difficulties. One solution, which works well for the MNL model, is to randomly sample from the full choice set and treat the respondent's choice as having come from the reduced choice set. The IIA property of the MNL model allows consistent estimation using such a sampling approach. However, much less is understood about the effects of a sampling approach for non-IIA models, and this is an area requiring further study.

Despite the theoretical consistency of MNL estimates, we found serious problems with attempts to use simple random samples for this RP application. The problem is that 46% of the 607 respondents chose new vehicles, but new vehicles comprise only 52 of the 689 alternatives. It is therefore likely that any sample of size 30 would only contain one or two new vehicles, and this leads to implausibly high estimates for the new vehicle dummy variable. Our solution is to use a type of "importance sampling." We stratified the sample according to vintage so that each sampled choice set contains 7 new vehicles, 7 1–2 yr old vehicles, 7 3–10 yr old vehicles, and 7 more than 10-yr old vehicles. The resulting 28 alternative choice sets yields reasonable estimates for the vintage coefficients. For example, the "Normalized Coefficients" Table 3 show that a new car for households with $38 000 annual income is equivalent to an identical one-year old car with a purchase price reduced by $7000.

The MNL coefficient estimates in Table 3 give generally reasonable signs for the generic attributes, but only the price and operating cost coefficients are estimated with any accuracy due to high multicollinearity between range, top speed, and acceleration. The coefficients are larger in magnitude than the MNL estimates for the SP data given in Table 2. This indicates that the variance of the stochastic portion of utility is lower for the RP data. The normalized coefficients show that the different body types have lower values than the SP MNL model.

A comparison of the SP MNL coefficients in Table 2 with the corresponding RP MNL coefficients in Table 2 demonstrates some of the issues associated with attempts to combine discrete choice data from two data sources. First, we would expect there to be major agreement between the two models with respect to the signs of the coefficient estimates. There is indeed substantial agreement; however, there are some differences. The sign for SportsCar is negative in the RP model, whereas it is positive in the SP model. (And, both are statistically significant.) In addition, the interaction effects between SportsCar and Household-size-greater-than-three also have different signs. The SP model gives much more positive weight to sport utility vehicles. Finally, the sign for emissions is different between the RP and SP models.

The coefficients related to sports car are readily interpreted. Sports cars have a very small percentage of the actual vehicle market, even taking into account the objectively measured physical attributes and prices for these vehicles. This yields a negative coefficient for this body-type in the RP model. And, because the models in this paper are for vehicle purchases only, it would seem more likely for a larger household to purchase a sports car, *ceteris paribus*, since they are more likely to hold multiple vehicles.

With respect to the SP coefficients, it is possible to tell an "SP bias story" in which respondents are "tempted" to choose a sports car while in their "SP fantasy land", when in fact they might not do so in reality. Further, this effect is evidently mitigated for those respondents in larger households (a "guilt" effect?). This is a plausible interpretation due to the customization scheme described in Section 2, because only six vehicles are generated for each choice set. A relatively small number of households indicated in the telephone interview that their next purchase would be a sports car. Those households received choice sets containing sports cars. However, many other households also received choice sets that included sports cars, giving them a chance to consider and switch to such a vehicle in a manner that would perhaps be inconsistent with a more realistic choice process. We would expect this effect to potentially create bias for other body types as well, but not to the degree that might be expected for a specialized vehicle like a sports car. This discussion highlights the fact that, later on, we might expect to use body-type estimates derived from the RP choices to correct for these effects.

The sign difference for emissions is more problematic. The negative sign of the SP estimate is entirely expected, given the nature of the experiment. Even if one chooses to discount the result as due to some sort of public-good bias effect, the interpretation of the RP coefficient is equally problematic. Do people actually prefer *dirtier* vehicles to cleaner ones, all else equal? The high degree of collinearity between vehicle age and many of the other attributes (e.g., price, performance, size, emissions) creates a host of difficulties when estimating RP models. In particular, the emissions variable is almost completely correlated with vehicle age in the RP data, primarily due to the historical trend in government clean-air regulations.

For those remaining coefficients that are consistent in sign, it would be fortuitous if a very simple scaling effect would explain the differences between the RP and SP coefficients, so that the attribute trade-offs from the two models could be regarded as essentially the same (within the bounds of statistical error). For example, many of the attributes of major interest (price/log(income), fuel cost, range, acceleration, top speed) have smaller coefficients in the SP model than in the RP model. If this pattern were consistent throughout, it would correspond to greater error variability in the SP model. Unfortunately, it is clear that the coefficient differences are unlikely to be completely explained by a simple scaling effect, as will be addressed in more detail in the next section.

A final issue is the possibility of differences between the RP and SP models due to heterogeneity of the samples. The RP model is estimated using actual purchases from 607 households, whereas the SP model is based on hypothetical choices from a much larger sample. As part of our testing procedures, we divided the SP sample into two segments: One segment contained those households who made RP purchases, and the second contained those who did not. Based on likelihood ratio tests, there were no statistical differences between model estimates obtained from the two samples. In subsequent analyses we determined that it was inefficient to throw away a huge portion of our SP sample, especially for purposes of mixed logit model estimation.

4.3. Joint RP/SP models

The first block of columns in Table 4 contain estimates from the "best" joint RP/SP scaled MNL model. The coefficient estimates fall into one of three categories: (1) coefficients that are uniquely determined by either the RP or the SP data, (2) coefficients that are the same in the two data sets (except for a scale effect), (3) coefficients that different in the two data sets.

Even though joint estimation of RP/SP models is becoming increasingly common, there is still a limited amount of experience on the issue of which coefficients would be expected to "pool" and

Table 4
Joint revealed/stated preference models

Variable	Multinomial logit log likelihood = −9134.86				Mixed logit log likelihood = −9102.67				Normalized coefficients	
	Coefficient	*t*-stat	Coefficient	*t*-stat	Coefficient	*t*-stat	Coefficient	*t*-stat	MNL	ML
Price/ln (income)	−0.361	−6.8			−0.289	−5.8			−3.67	−3.67
Operating cost	−0.170	−5.3			−0.131	−4.8			−1.73	−1.66
Acceleration	−0.149	−4.2			−0.090	−3.8			−1.51	−1.14
Top speed	0.641	3.1			0.385	2.8			6.51	4.88
Range	1.268	4.5			0.998	4.1			12.88	12.66
Range squared	−0.116	−3.1			−0.099	−3.0			−1.18	−1.26
	RP-Specific		SP-Specific		RP-Specific		SP-Specific			
Pollution	0.376	4.5	−0.673	−2.8	0.317	3.8	−0.388	−2.7	−6.84	−4.93
Truck	−0.368	−2.3	−2.220	−5.2	−0.380	−2.4	−0.686	−3.7	−3.73	−4.82
Van	−0.339	−1.7	−2.554	−5.2	−0.445	−2.2	−0.742	−3.7	−3.44	−5.64
Sports utility vehicle	0.481	2.5	1.938	4.1	0.253	1.4	0.577	3.2	4.89	3.21
Mini sports utility	−1.276	−2.0	−0.090	−0.1	−1.404	−2.2	0.207	0.8	−12.95	−17.81
Small car	−0.488	−3.4	−0.188	−1.8	−0.471	−3.2	−0.046	−1.0	−4.95	−5.97
Sports car	−0.594	−2.4	2.050	3.7	−0.588	−2.4	0.661	3.1	−6.03	−7.46
Sports car × HHG3	0.852	2.8	−1.874	−2.1	0.870	2.9	−0.653	−2.2	8.65	11.03
Minivan × HHG3	0.937	3.5	2.208	4.7	1.051	3.9	0.721	3.6	9.52	13.34
	RP-Unique				RP-Unique					
Luxury	−0.097	−0.5			−0.242	−1.4			−0.98	−3.07
Import	−0.320	−2.9			−0.291	−2.7			−3.25	−3.69
Log (models)	0.730	9.6			0.718	9.4			7.42	9.11
Log (age)	−0.429	−3.3			−0.430	−3.3			−4.36	−5.46
New	0.957	4.1			0.768	3.4			9.72	9.75
Used 1	0.316	1.4			0.228	1.0			3.21	2.90
			SP-Unique				SP-Unique			
Station availability			0.682	3.0			0.526	3.1	6.93	6.68
Station wagon			−3.175	−5.4			−0.937	−3.7	−32.24	−11.89
Constant for EV			−0.051	−0.2			−0.784	−2.2	−0.52	−9.94
Constant for CNG			0.507	2.6			0.236	1.6	5.15	3.00
Constant for methanol			0.887	4.0			0.643	3.7	9.01	8.16
Electric truck			−0.577	−1.9			−0.187	−1.9	−5.86	−2.37
Electric sports car			−1.019	−1.9			−0.257	−1.1	−10.35	−3.26
College × EV			0.605	2.9			0.506	2.6	6.15	6.42
SP scale factor			0.45	5.6			1.628	3.8		
Std. dev. gasoline					1.097	2.9				
Std. dev. EV					2.937	4.3				
Std. dev. CNG					2.068	4.2				
Std. dev. methanol					0.836	1.9				
Std. dev. fuelcost					0.259	4.5				

which ones would not. However, it is generally accepted that coefficients acting as "alternative specific constants" ("ASC's") in the MNL model might be much less likely to "pool" than coefficients associated "generic attributes". An ASC represents the mean of a collection of random effects due to, e.g., unobserved variables, once effects associated with all the other variables have been taken into account. The body-type coefficients in our vehicle choice models would be expected to behave in this fashion, and the differences in the data generation processes producing the SP and RP choices in our survey could easily produce different coefficient values for these variables. Indeed we found that it was necessary to estimate data-set-specific coefficients for all the body-type variables in our models.

It would be ideal if all generic attribute coefficients could be estimated "in common" across both data sets (controlling for scale). We have already noted that the emissions coefficients for the RP and SP models had opposite signs. In our joint RP/SP model we found that we could pool all the remaining generic attributes. This was possible in part because the statistical significance for some attributes was relatively weak in the RP data. However, this serves to illustrate the manner in which SP and RP data can complement one another. Designed SP experiments have more statistical power for estimating trade-offs among generic attributes than do RP data with high levels of multicollinearity.

The middle block of columns in Table 4 gives the mixed logit model estimates with the best fit for the joint data. We use the same error components as in the SP mixed logit model in Table 2. Some experimentation with adding other error components to the model did not yield any significant improvement in the log likelihood. Similar to the SP case, the addition of the five error components given at the bottom of Table 4 significantly improves the log likelihood of the resulting model. The likelihood ratio statistic is 64.38. Furthermore, the presence of the EV, CNG, and Methanol error components (which are unique to the SP data) has a noticeable effect on the SP scale factor as compared to the corresponding scale factor from the joint MNL model. The MNL scale factor is smaller than one, indicating that the stochastic error term for the SP data set has a larger variance than the RP data set. However, the scale in the joint MNL specification is forced to capture all sources of random error, including preference heterogeneity. However, in the mixed logit model SP scale is larger than one (variance is less than one) once preference heterogeneity has been captured by the fuel-type error components.

One potential problem with these mixed logit estimates is that they are using the sampled RP choice sets described in Section 4.2. While sampling yields consistent estimates for the MNL model, the effect for non-IIA models such as mixed logit is unclear, as previously noted. We performed experiments to evaluate the effect of increasing the size of the sampled choice sets, and found no effects that would suggest systematic bias in the coefficient estimates; however, this is an issue that deserves further study.

Both the MNL and mixed logit models in Table 4 are estimated assuming that the unobserved error terms are independent across RP and SP choices made by the same households. We tried estimating some joint mixed logit models using the same random error components for both choices in the probability simulation calculations. This induces some correlation between choices for the same household, but it yields identical results to those in Table 4. It is likely that there would be more differences between these specifications in situations with many more repetitions per decision maker such as is commonly found in repeated SP tasks. Ben-Akiva and Morikawa (1997), and Morikawa (1994) specify models with state dependence (between the SP and RP choices) and serial correlation in the stochastic terms. Mixed logit versions of their models would be easy to specify and estimate, but they are not needed for our specifications.

The final set of columns in Table 4 give the "Normalized Coefficients" for the MNL and mixed logit specifications. We only give these normalized coefficients for those coefficients used in the forecasting experiments in the next section. Since the RP model is expected to give more accurate estimates of the body-type coefficients, we use the "RP-Specific" coefficients for these variables. However, we use the "SP-Specific" coefficient for emissions because of the problems with the RP model discussed in Section 4.2. Comparisons between these normalized coefficients show that the biggest differences between the joint scaled logit and mixed logit models are for the "SP-Unique" coefficients. This is not surprising since three of the error components only affect the SP data, and the differences are similar to those found just using the SP data in Section 4.1. However, there are substantial differences between the SP mixed logit model in Table 2 and the joint RP/SP mixed logit model we use for forecasting in the next section. These differences are primarily due to the use of RP body-type coefficients in the forecasting model, which highlights the importance of joint RP/SP modeling to capture the strengths and avoid the weaknesses of each type of data.

5. Scenario forecasts

Although there are some differences in the normalized coefficients between the MNL and mixed logit models described in Sections 4.1 and 4.3, the main differences between these models are due to the different substitution patterns caused by the different error specifications. The easiest way to see these differences is to compare forecasts for new alternatives for the various models. This sections presents the results of some forecasting experiments using a more realistic description of available vehicles than in Brownstone and Train (1998). The full set of vehicles we consider is given in Table 5, and is taken from a comprehensive set of vehicle technology forecasts prepared by the California Energy Commission as an input to the microsimulation model described in Bunch et al. (1996). We chose the year 1998 since that was originally the year that California would begin to mandate the sale of a substantial number of alternative-fuel vehicles. The operating fuel costs are derived assuming that gasoline costs $1.20/gallon and electricity costs 6 cents/KWH, and all prices are in 1995 dollars. The "MPG" column in Table 5 gives mileage in gasoline equivalents for CNG and methanol. The scenario presented in Table 5 is still unrealistic since it excludes used vehicles. The forecasts will therefore overstate the survey respondents actual demand for alternative-fuel vehicles.

The vehicle classes described in Table 5 present a very optimistic view of electric vehicle technology since they exclude battery replacement costs. Some estimates of these costs indicate that they might exceed the fuel costs (listed in the "cents/mile" column) if averaged over 10 000 annual miles per year. Measuring acceleration as time to reach 30 miles per hour also paints a rosy picture of electric vehicles, since their acceleration capabilities dramatically reduce as speed increases. Of course, this bias towards electric vehicles should not affect the comparison between the different models' forecasts since they are all based on the same data given in Table 5.

Table 6 gives the results of some forecasting experiments for the SP MNL and mixed logit models in Table 2. Table 7 gives the same forecasts for the joint RP/SP MNL and mixed logit models in Table 4. Note that these are unweighted forecasts over the 4656 respondents with

Table 5
"1998" Scenario definition

Alt. No.	Class	Cost	Dom/Imp	Body type	Fuel	Models	Price	Cents/mile	mpg	Range	0–30 Acc	Topspeed	Pollution	Stations
1	1	Low	I	Mini	Electric	1	17 518	1.33		75	3.9	92	0.0	0.1
2	1	High	I	Mini	Gasoline	4	19 986	4.90	25	281	3.0	120	0.8	1.0
3	2	Low	D	Sub compact	Electric	1	19 562	1.47		75	3.6	98	0.0	0.1
4	2	Low	I	Sub compact	Gasoline	20	13 524	4.53	26	344	3.5	110	0.8	1.0
5	2	High	I	Sub compact	Gasoline	17	31 415	5.48	22	277	2.9	122	0.8	1.0
6	3	Low	D	Compact	Gasoline	20	14 814	4.63	26	389	3.4	111	0.8	1.0
7	3	High	I	Compact	Gasoline	30	35 689	5.77	21	306	3.0	121	0.8	1.0
8	3	Low	D	Compact	cng	1	21 057	4.14	29	147	3.7	105	0.4	0.1
9	4	Low	D	Midsize	Gasoline	23	18 695	5.73	21	345	3.3	114	0.8	1.0
10	4	Low	D	Midsize	Methanol	3	18 985	5.38	22	215	3.2	115	0.6	0.7
11	4	High	I	Midsize	Gasoline	17	36 127	6.39	19	310	2.6	133	0.8	1.0
12	4	High	D	Midsize	Gasoline	6	39 382	5.92	20	334	2.8	126	0.8	1.0
13	4	High	D	Midsize	Methanol	1	36 504	6.00	20	193	2.5	134	0.6	0.7
14	5	Low	D	Large	cng	1	26 561	5.24	23	150	3.4	112	0.4	0.1
15	5	Low	D	Large	Gasoline	11	23 083	5.88	20	398	3.1	117	0.8	1.0
16	5	High	D	Large	Gasoline	9	50 000	6.70	18	390	2.9	122	0.8	1.0
17	6	Low	D	Sports	Electric	1	26 414	1.96		75	3.0	109	0.0	0.1
18	6	Low	I	Sports	Gasoline	12	19 641	5.41	22	344	3.0	121	0.8	1.0
19	6	High	D	Sports	Gasoline	2	38 999	6.78	18	247	2.0	154	0.8	1.0
20	6	High	I	Sports	Gasoline	30	53 309	6.18	19	270	2.6	132	0.8	1.0
21	7	Low	D	Compact p.u.	Electric	1	21 669	2.05		75	3.2	84	0.0	0.1
22	7	Low	I	Compact p.u.	Gasoline	11	15 000	5.61	21	362	3.2	94	0.8	1.0
23	7	High	D	Compact p.u.	Gasoline	4	20 401	7.23	17	299	3.1	95	0.8	1.0
24	8	Low	D	Standard p.u.	Gasoline	15	18 839	8.34	14	324	3.2	94	0.8	1.0
25	8	Low	D	Standard p.u.	Methanol	1	19 129	7.84	15	202	3.1	95	0.6	0.7
26	8	High	D	Standard p.u.	Gasoline	4	24 175	7.96	15	336	3.8	84	0.9	1.0
27	9	Low	I	Minivan	Electric	1	29 785	2.36		75	3.3	82	0.0	0.1
28	9	Low	D	Minivan	Gasoline	17	22 000	6.49	19	370	3.3	91	0.8	1.0
29	9	High	I	Minivan	Gasoline	3	27 533	6.96	17	354	3.4	90	0.8	1.0
30	10	Low	D	Minivan	Gasoline	18	19 741	8.40	14	357	3.2	94	0.8	1.0
31	10	Low	D	Minivan	Methanol	1	20 031	7.89	15	223	3.2	95	0.6	0.7
32	10	High	D	Minivan	Gasoline	2	24 820	9.43	13	326	3.4	90	0.8	1.0
33	11	Low	D	Compact SUV	Gasoline	11	23 100	7.16	17	338	3.2	91	0.8	1.0
34	11	High	I	Compact SUV	Gasoline	3	30 000	8.28	14	299	3.0	95	0.8	1.0
35	12	Low	D	Standard SUV	Gasoline	3	25 651	9.08	13	396	3.4	91	0.9	1.0
36	12	High	D	Standard SUV	Gasoline	4	27 629	9.41	13	385	3.4	90	0.9	1.0
37	13	Low	I	Mini SUV	Gasoline	7	15 223	5.28	23	261	3.7	86	0.8	1.0

Table 6
SP model scenario forecast market shares (%)

Alt. No.	Body type	Fuel	SP MNL			SP mixed logit		
			Gas	Non-EV	Full	Gas	Non-EV	Full
1	Mini	Electric			2.28			7.38
2	Mini	Gasoline	4.35	3.77	3.48	4.10	2.79	2.24
3	Sub compact	Electric			2.10			5.53
4	Sub compact	Gasoline	7.15	6.20	5.72	16.81	11.45	9.03
5	Sub compact	Gasoline	2.29	1.98	1.83	0.68	0.46	0.38
6	Compact	Gasoline	7.45	6.46	5.96	16.61	11.31	8.96
7	Compact	Gasoline	1.96	1.69	1.56	0.43	0.29	0.25
8	Compact	cng		2.35	2.17		9.66	7.75
9	Midsize	Gasoline	5.51	4.78	4.41	5.51	3.77	3.16
10	Midsize	Methanol		4.76	4.39		11.61	8.70
11	Midsize	Gasoline	2.12	1.84	1.70	0.45	0.31	0.27
12	Midsize	Gasoline	1.93	1.68	1.55	0.35	0.24	0.20
13	Midsize	Methanol		1.86	1.72		0.88	0.69
14	Large	cng		1.86	1.72		3.78	3.20
15	Large	Gasoline	5.01	4.34	4.01	3.81	2.61	2.19
16	Large	Gasoline	1.21	1.05	0.97	0.10	0.07	0.06
17	Sports	Electric			2.25			3.38
18	Sports	Gasoline	12.05	10.46	9.64	14.25	9.75	8.03
19	Sports	Gasoline	3.46	3.00	2.77	0.55	0.38	0.33
20	Sports	Gasoline	1.73	1.51	1.39	0.09	0.06	0.05
21	Compact p.u.	Electric			0.55			0.77
22	Compact p.u.	Gasoline	2.49	2.16	1.99	3.12	2.14	1.78
23	Compact p.u.	Gasoline	1.40	1.22	1.12	0.82	0.57	0.51
24	Standard p.u.	Gasoline	1.48	1.28	1.18	1.40	1.00	0.91
25	Standard p.u.	Methanol		1.31	1.21		2.48	2.14
26	Standard p.u.	Gasoline	1.07	0.93	0.86	0.50	0.36	0.32
27	Mini van	Electric			0.55			0.43
28	Mini van	Gasoline	1.97	1.70	1.57	1.38	0.95	0.82
29	Mini van	Gasoline	1.36	1.18	1.09	0.57	0.39	0.35
30	Mini van	Gasoline	1.32	1.14	1.05	1.38	0.98	0.90
31	Mini van	Methanol		1.15	1.06		2.49	2.15
32	Mini van	Gasoline	0.84	0.73	0.67	0.77	0.56	0.53
33	Compact SUV	Gasoline	8.69	7.53	6.95	5.39	3.75	3.31
34	Compact SUV	Gasoline	5.11	4.43	4.08	1.99	1.41	1.29
35	Standard SUV	Gasoline	7.20	6.24	5.76	6.71	4.86	4.51
36	Standard SUV	Gasoline	6.14	5.32	4.91	5.55	4.05	3.79
37	Mini SUV	Gasoline	4.71	4.09	3.77	6.68	4.56	3.74

complete SP data, so they do not represent overall population or vehicle market forecasts. [2] The "Gas" columns of Table 6 give the market share forecasts for a scenario only including the gasoline vehicles in Table 5. The "Non-EV" columns show the forecasts when the non-electric

[2] The mixed logit forecasts in Tables 6 and 7 are computed using 100 repetitions for each household. Experimentation with more repetitions showed that 100 repetitions is more than sufficient for 2 digit accuracy.

Table 7
Joint model scenario forecast market shares (%)

Alt. No.	Body type	Fuel	Joint MNL			Joint mixed logit		
			Gas	Non-EV	Full	Gas	Non-EV	Full
1	Mini	Electric			0.18			1.89
2	Mini	Gasoline	1.34	1.27	1.26	1.31	1.14	1.05
3	Sub compact	Electric			0.21			2.16
4	Sub compact	Gasoline	12.36	11.71	11.64	11.51	10.05	9.22
5	Sub compact	Gasoline	1.11	1.05	1.05	1.32	1.15	1.07
6	Compact	Gasoline	17.88	16.93	16.82	15.39	13.44	12.35
7	Compact	Gasoline	1.25	1.18	1.17	1.52	1.33	1.23
8	Compact	cng		0.24	0.24		1.99	1.80
9	Midsize	Gasoline	15.36	14.54	14.45	14.59	12.76	11.86
10	Midsize	Methanol		3.45	3.43		5.77	5.16
11	Midsize	Gasoline	1.35	1.28	1.27	1.59	1.39	1.30
12	Midsize	Gasoline	0.72	0.68	0.68	0.86	0.75	0.70
13	Midsize	Methanol		0.23	0.23		0.46	0.42
14	Large	cng		0.22	0.21		1.79	1.64
15	Large	Gasoline	7.22	6.84	6.80	6.91	6.05	5.62
16	Large	Gasoline	0.38	0.36	0.36	0.53	0.47	0.44
17	Sports	Electric			0.05			1.04
18	Sports	Gasoline	5.12	4.86	4.83	5.01	4.39	4.06
19	Sports	Gasoline	0.17	0.16	0.16	0.22	0.19	0.18
20	Sports	Gasoline	0.24	0.22	0.22	0.40	0.35	0.33
21	Compact p.u.	Electric			0.09			1.38
22	Compact p.u.	Gasoline	6.52	6.18	6.14	6.00	5.25	4.87
23	Compact p.u.	Gasoline	1.25	1.19	1.18	1.33	1.17	1.10
24	Standard p.u.	Gasoline	3.89	3.69	3.66	5.20	4.57	4.34
25	Standard p.u.	Methanol		0.56	0.56		1.22	1.12
26	Standard p.u.	Gasoline	0.69	0.66	0.65	0.92	0.81	0.77
27	Minin van	Electric			0.07			0.76
28	Mini van	Gasoline	7.00	6.64	6.60	7.09	6.23	5.82
29	Mini van	Gasoline	0.64	0.60	0.60	0.70	0.61	0.58
30	Mini van	Gasoline	4.82	4.56	4.53	5.72	5.03	4.78
31	Mini van	Methanol		0.61	0.61		1.19	1.10
32	Mini van	Gasoline	0.36	0.34	0.34	0.56	0.49	0.47
33	Compact SUV	Gasoline	6.07	5.74	5.71	5.93	5.20	4.89
34	Compact SUV	Gasoline	0.55	0.52	0.52	0.68	0.60	0.57
35	Standard SUV	Gasoline	1.48	1.40	1.39	2.04	1.80	1.72
36	Standard SUV	Gasoline	1.22	1.16	1.15	1.66	1.46	1.40
37	Mini SUV	Gasoline	1.00	0.95	0.94	1.02	0.89	0.83

(CNG and methanol) vehicles are added, and the "Full" column show the forecasts when all vehicles in Table 5 are available.

Generally the mixed logit models show much higher market shares for the alternative-fuel vehicles than the independent logit models. This is due to the large error components associated with the fuel constants in the mixed logit models. These error components imply some fraction of the sample will have a large enough fuel type constant to counteract the negative observed utilities associated with alternative-fuel vehicles. The IIA property of the independent logit models guarantees that a proportionate share of each new vehicle's market share must come from all other vehicles. Thus the market share for the mini electric vehicle (alternative number 1) draws a proportionate share from all vehicles in the Non-EV scenario. The mixed logit specifications generate the more reasonable prediction that the market share for the mini electric vehicle comes disproportionately from other mini and subcompact vehicles.

Tables 6 and 7 show that the joint models give quite different forecasts from the pure SP models. The SP models give high forecasts for non-gasoline vehicles- their total share for the full scenario is 20% for the MNL specification and 42% for the mixed logit specification. In contrast, the joint models give a non-gasoline share of 6% for the MNL specification and 18% for the mixed logit specification. The pure SP models also give very high forecasts for sports cars (12–16%) and sport utility vehicles (26–31%), while the joint models give more reasonable forecasts of approximately 6% for sports cars and 10% for sport utility vehicles. These differences are due to the use of RP body-type coefficients in the joint forecasting models.

6. Conclusions

Mixed logit models are a general and feasible class of models for joint RP/SP choice data. They can easily account for the scaling and unobserved error correlations typically found in these applications. The mixed logit specifications used here are particularly helpful when preference heterogeneity can be captured by a relatively small number of error components, since the order of integration or simulation is given by the number of these components rather than the number of discrete alternatives (as in, e.g., multinomial probit). However, modeling RP vehicle choices with any discrete choice model can be difficult due to the extremely large number of choices in the marketplace, and more work is needed to address this issue. In particular, procedures that rely on sampled choice sets for estimating non-IIA models require more investigation.

The alternative-fuel vehicle models presented here also highlight the advantages in merging SP and RP data. Even if we had a much larger RP sample size, RP data would still be plagued by multicollinearity and difficulties with measuring vehicle attributes. Nevertheless, RP data appear to be critical for obtaining realistic body-type choice information, and for scaling purposes. SP data are critical for obtaining information about attributes not available in the marketplace. Although SP data are subject to their own limitations, these can be overcome by careful joint modeling.

The models presented in this paper show large heterogeneity in preference for fuel types. Although some of this is could be due to respondents' different information sets and fundamental uncertainty, it is also likely that some preference heterogeneity could be explained by adding much more detailed demographic interactions to the models. (However, adding such variables makes producing forecasts much more complicated.) Our future research will examine a richer set of possible interactions as well as incorporating a very different SP experiment given to the Wave 2 respondents in our panel survey. We expect that the mixed logit model class will be useful in the search for these better models.

Acknowledgements

Tom Golob helped collect the data and conducted preliminary analyses upon which our analysis relies. Hongyan Sheng provided invaluable assistance with preparing the data for analysis. Collection of the data used in this paper was financed by Southern California Edison Co., Pacific Gas and Electric Co., the University of California Energy Institute and the California Energy Commission. Portions of this work were performed at the National Micro Economic Modelling Laboratory (MEMLab) of the University of Sydney, Australia. Of course, none of these people or agencies is responsible for any errors or representations that we make in this paper.

References

Ben-Akiva, M., Buldoc, D., 1996. Multinomial probit with a logit kernel and a general parametric specification of the covariance structure. Working paper, Department of Economics, MIT.

Ben-Akiva, M., Morikawa, T., 1997. Methods to estimate discrete choice models with stated and revealed preferences. Paper presented at 1997 NSF Symposium on Eliciting Preferences, Berkeley, California, July 1997.

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.

Brownstone, D., Bunch, D., Golob, T., Ren, W., 1996. Transactions choice model for forecasting demand for alternative-fuel vehicles. In: McMullen, S. (Ed.), Research in Transportation Economics, vol. 4, JAI Press, pp. 87–129.

Brownstone, D., Train, K., 1998. Forecasting new product penetration with flexible substitution patterns. Journal of Econometrics (in press).

Bunch, D.S., 1988. A comparison of algorithms for maximum likelihood estimation of choice models. Journal of Econometrics 38, 145–167.

Bunch, D.S., Brownstone, D., Golob, T.F., 1996. A dynamic forecasting system for vehicle markets with clean-fuel vehicles. In: Hensher, D.A., King, J. (Eds.), World Transport Research, vol. 4, pp. 189–203.

Bunch, D.S., Gay, D.M., Welsch, R.E., 1993. Algorithm 717: Subroutines for maximum likelihood and quasi-likelihood estimation of parameters in nonlinear regression models. ACM Transactions on Mathematical Software 19 (1), 109–130.

Hensher, D., Bradley, M., 1993. Using stated response data to enrich revealed preference discrete choice models. Marketing Letters 4 (2), 139–152.

Lee, L., 1992. On efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete response models. Econometrica 8, 518–552.

McFadden, D., Train, K., 1997. Mixed MNL models for discrete response. Department of Economics, UC Berkeley.

Mehndiratta, S., 1996. Time-of-day effects in inter-city business travel. Ph.D. thesis, Department of Civil Engineering, University of California, Berkeley.

Morikawa, T., 1994. Correcting state dependence and serial correlation in RP/SP combined estimation method. Transportation 21 (2), 153–166.

Revelt, D., Train, K., 1998. Incentives for appliance efficiency in a competitive energy environment: random-parameters logit models of households' choices. Review of Economics and Statistics (in press).

Swait, J., Louviere, J.J., 1993. The role of the scale parameter in the estimation and comparison of multinomial logit models. Journal of Marketing Research 30, 305–314.

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., 1998. Unobserved taste variation in recreation demand models. Land Economics (in press).