Title

A Dynamic household Alternative-fuel Vehicle Demand Model Using Stated and Revealed Transaction Information

Author

Sheng, Hongyan

Publication Date

1999

UNIVERSITY OF CALIFORNIA

IRVINE

A Dynamic Household Alternative-fuel Vehicle Demand Model Using Stated and Revealed Transaction Information

submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in Civil Engineering

by

Hongyan Sheng

1999

CHAPTER 1

INTRODUCTION

1.1 *Motivation*

Improving air quality has long been a big concern for society. The original Clean Air Act[1] was signed by president Nixon in 1970 in accordance to national clamor for environmental healing. In 1990, president Bush signed the Clean Air bill which made significant revisions to the original Clean Air Act. The Clean Air Act Amendments of 1990 establishes tighter pollution standards for emissions from automobiles and trucks. The new law also allows stricter emission limits for vehicles in California which can be met with any combination of vehicle technology and cleaner fuels. As a result, in the 1990s, California passed a law which mandates the introduction and sale of low-emission vehicles(e.g. natural gas vehicles) and zero-emission vehicles(e.g. electric vehicles). According to the levels set by California Air Resources Board, 10% of all vehicles sold in California must be electric vehicles by year 2003. Moreover, other states are actually considering following California's lead and adopting similar policies and incentive programs.

Since California accounts for 12 percent of U.S. new vehicle market already (according to Ren(1995)), vehicle manufacturers have an incentive to not only meet the state's emission mandate but also to take lead in the race of entering the niche market of alternative-fuel

[1] The Clean Air Act was enacted to protect and enhance the quality of the nation's air resources, to prevent and control air pollution, and to provide technical and financial assistance to state and local governments to develop pollution control program.

vehicles(AFV). Large auto manufacturers have made huge investment in developing all kinds of new technologies to reduce vehicle emission while at the same time trying to improve AFV performance relative to that of conventional gasoline vehicles. Auto maker DaimlerChrysler, has recently disclosed a new fuel-cell zero emission technology. This announcement brings more public awareness of the relative performance of AFVs and the prospect of mass production and future availability of AFVs. Auto makers who are designing and marketing AFVs, fuel suppliers who are preparing the necessary infrastructure, and state and federal government agencies who are making the policy decisions are eager to know about future market demand for AFVs. Therefore, forecasting alternative-fuel vehicle demand becomes very important for the decision making of both California and the other states, the Environmental Protection Agency (EPA) , the car manufacturers, and the fuel suppliers.

1.2 *Problems Faced and Issues Addressed*

However, since most alternative-fuel vehicles have not been commercially available in the market and their attributes are very different from those of conventional gasoline vehicles(such as reduced range and limited refueling options etc.), special stated preference(SP) experiments[2] have been designed to elicit households' preferences toward certain attributes of hypothetical future alternative-fuel vehicles (AFVs). (See Bunch et. al.(1992).)

[2]Stated preference is a method used to elicit respondents' preferences for changes in alternatives' attributes or for new alternative(s) by having them choose among hypothetical alternatives.

Previous work by Brownstone, Bunch, and Train(1999) shows that there are serious biases in these stated preference data. Therefore, this study seeks to use both the revealed preference(RP)[3] and the stated preference data in the demand model estimation. It is advantageous to combine the SP and RP information since they have complementary characteristics. Stated preference data is necessary for obtaining alternative-fuel vehicle attribute choices that are not available in the market; revealed preference information can be used to correct the possible biases associated with these SP data.

In the short run, households face the decision of whether to make vehicle transactions and what type of vehicles to purchase. Therefore, in order to come up with better short-run demand forecasts, I apply a dynamic vehicle transaction choice model using both the pure SP data and the joint RP and SP data. The model is dynamic in the sense that it is conditional on households' vehicle holdings and it incorporates households' vehicle type choices as well as households' transaction

choice[4]. In contrast, the models in former studies such as Brownstone, Bunch and Train (1999) can be viewed as "static", long run vehicle demand models because they simply look

[3]Revealed preference is respondents' actual observed choices among different existing alternatives.
[4]Vehicle type choices refer to households' preferences for vehicles in certain classes instead of detailed makes/models. Vehicle classes are determined by vehicle body type, high/low price tier, domest/import attribute, and vehicle vintage. Household transaction choices refer to households' choices for vehicle replacement, vehicle addition, or doing nothing etc..

at households' vehicle type choices without conditioning on households' existing vehicle information. The reason is that in the long run, households will go through vehicle transactions eventually and they only have to decide which vehicle to purchase. Moreover, I conduct the short-run demand forecasts for alternative-fuel vehicles based on the estimation results of the demand model. I also come up with more realistic forecasting scenario by combining new and used vehicle technologies and using the complete set of vehicle choices in the forecasting.

To provide more realistic forecasts, I also apply nested logit models besides the conditional multinomial logit models. It is well known that the standard logit exhibits restrictive "independence from irrelevant alternatives" property, which in many cases seems unrealistic. While allowing more flexibility, nested logit models can generate more realistic substitution patterns when a new alternative or product is introduced or when changes happen to existing alternatives. Therefore, this study serves as an example of how to forecast the demand for new products or technologies that mark considerable departures from existing products or technologies.

The main questions that I will address in this dissertation include:

(i) What are the important explanatory factors in affecting households' vehicle transaction decisions? How do households' transaction choices and attributes of their existing vehicles affect demand models?

(ii) How do the pure SP estimates differ from the pure RP model estimates? How do the forecasts differ for the pure SP model and the joint SP and RP model?

(iii) What are the differences between the standard logit and the nested logit model in terms of model estimates and forecasts?

(iv) What are the predicted market shares for future alternative-fuel vehicles?

1.3 *Overview of the study*

Chapter 2 reviews studies on households' vehicle transaction choice modeling, studies related to alternative-fuel vehicle demand forecasting and studies focusing on joint estimation of stated and revealed preference information. The major shortcomings of former vehicle transaction choice models come from several aspects.

These weaknesses can be summarized as:

(i) inclusion of single-vehicle households only;

(ii) modeling limited transaction choices, especially without considering household's choice of doing nothing;

(iii) developing separate models for one-vehicle and two-vehicle households;

(iv) use of stated preference vehicle transaction information only;

(v) no consideration of households' existing vehicle information;

(vi) lack of detailed households' demographic information and vehicle attributes;

(vii) use of unrealistic future vehicle choices as the forecast scenario.

In Chapter 3, I first introduce concepts of stated and revealed preference methods and compare their advantages and disadvantages. Then I describe the California Alternative-Fuel Vehicle Demand Forecasting Project and the panel surveys carried out by the project. Stated and revealed preference vehicle transaction choice data are collected through the panel surveys. Some of the statistics of the SP and RP data are also given. I also describe the theoretical background of conditional multinomial logit (MNL) and nested logit (NL) models. Moreover, I give the framework of the joint estimation of SP and RP information. Then I introduce and discuss the specifications, definition of explanatory variables and the estimation results for the RP MNL model, the SP MNL and SP NL models separately. Comparisons are also made between these model estimates.

The joint SP-RP models are presented in Chapter 4. I compare the joint SP-RP MNL and NL model estimates and discuss the advantages of the joint estimation by comparing the joint models against the pure SP or RP models. In Chapter 5, I conduct forecasts based on the joint models and the pure SP models separately in order to look at the differences of the model forecasts. Market shares of alternative-fuel vehicles are obtained. Moreover, I incorporate the new and used vehicle technologies and develop a more realistic forecasting scenario. I also conduct forecasting exercises to look at the different substitution patterns predicted by the conditional multinomial logit and nested logit models. In the last chapter, Chapter 6, I give the conclusion.

CHAPTER 2
LITERATURE REVIEW

2.1 *Studies on Households' Vehicle Transaction Choices*

There are few studies which model the household sector's demand for automobiles by explicitly considering households' vehicle transaction choices. However, in the short run, households are most likely to have to decide whether to add a vehicle, or replace one of their existing vehicles etc.. By incorporating households' vehicle transaction choices, we can make the model more dynamic in the sense that we are looking at changes of households' vehicle holdings.

In Hensher et.al. (1992), they describe the Dimensions of Automobile Demand Project which was carried out in the Sydney metropolitan area from 1981 to 1991. The main objective of the project was to collect a panel data set and develop policy-sensitive econometric models to estimate and predict households' demand for automobiles. In their book, they describe the purposes of the project, their approaches to the study, the theoretical background and empirical results of the static and dynamic vehicle choice/use models, and forecasts of automobile energy demand.

Hensher et. al.(1992) have also given an overview of the main empirical studies of household vehicle choice and usage models (pp. 8-12). They classify the studies by looking at the following conditions:

(i) model status in terms of considering vehicle holdings only, considering vehicle transaction only, or considering both;

(ii) fleet size i.e. what are the sizes of household's vehicle fleets?

(iii) type choice i.e. what are the choice sets defined in those studies? What kind of models(such as multinomial logit or nestes logit models etc.) are applied in those studies?

(iv) vehicle usage i.e. whether there are vehicle usage models in those studies? If there are, what are the dependent variables and what kind of methods are used?

(v) sample, i.e. how are the samples selected in those studies?

The static vehicle choice model is achieved by constructing a nested logit model of vehicle holdings. (The graph which they used to show the nested structure is on pp.36.) Basically, they consider a three-level nested logit model with first level being the fleet size choices, the second level being body-type choices and the third level being model type choices. Hensher et. al.(1992) find that explanatory variables such as annual household income, age of household head, number of children, number of workers etc. are very important in explaining households' choices of their vehicle fleet compositions. They also put some lifecycle stage dummies in the model, which are also significant, such as young adults without children, adults 35 to 64 years old without children and retired persons without children.

The dynamic vehicle choice model is an extension to the static model. It allows current choice behavior to depend on the past behavior by including some state variables. According to Hensher et. al.(1992), these state variables represent stocks of automobiles and stocks of habits, including the cumulative effect on the present choice of the most recent continuous experience in a state as well as habit persistence. The estimates of the dynamic vehicle choice model show that the experience effects are very significant, i.e. there is an important role of habit in shaping type-choice holding.

The static vehicle use model is a joint model of vehicle choice and vehicle use. It allows the use of each vehicle to endogenously depend on the use of other vehicle in the household. It also considers the role of vehicles held during the periods of vehicle use which were disposed of before the point of defining vehicle holdings. Three-stage least squares is used to obtain the parameter estimates. Variables such as number of workers, kilometers of other vehicles, fuel cost per vehicle kilometer and residential location etc. are found to be significant in the model estimation.

The last part of Hensher et. al. (1992) is about forecasting household auto energy demand. They look at different sources of influences on energy demand. Using the period of year 1985 to 1988, they conduct forecasts for fuel consumption for year 1995, 2000 and 2005.

The paper of Smith, Hensher, and Wrigley (1991) focuses on modeling household's revealed preference vehicle transactions. Using a four-wave panel data set of household vehicle transactions, they develop a dynamic discrete choice sequence model assuming that outcome sequences through time are of the beta-logistic form.[5] The beta-logistic model can separate out the time-varying macro effects such as the vehicle age from the time-invariant household-specific effects such as household income. It is assumed that only two choices are available for each household in each period: replacing a vehicle or keeping a vehicle. Smith et. al. (1991) find that the choice sequence of keep-keep-keep is most common and the sequence of replace-keep-keep is the second most common choice for their sample.

The sample used by Smith, Hensher and Wrigley(1991) is extracted from a household sample from the Sydney Metropolitan area. They study the sequence of household's annual vehicle transaction choice during the 1980-1985 period. They look at household vehicle fleet information, vehicle utilization rates and household socio-economic and demographic information during the four successive periods. Since they only investigate households with one vehicle in each of the four periods, their sample is relatively small consisting of 373 households only. The findings in

[5]Smith, Hensher and Wrigley(1991) try to identify the observed patterns of time-sequenced choice response. Such a choice sequence specification requires a functional form representing the distribution of selection probabilities. Since the beta distribution is suitable for modeling unknown variation in choice probabilities, combining logit form with the best beta distribution provides a way to separate out the intertemporal and individual-specific effects. Such a model is known as the beta logistic model.

Smith, Hensher and Wrigley(1991) suggest that some factors are very important in households' auto transaction choices such as the pattern of vehicle use, the recent previous history of transacting, and period-specific macro-effects.

There are no future alternative-fuel vehicles involved in their study. Moreover, since two or three vehicle households account for a large proportion of California population, I overcome the sample bias problem in Smith, Hensher and Wrigley(1991) by using not only one-vehicle households but also multiple-vehicle households. In addition, I assign a more realistic choice set to each household in each period. The choice set includes replacing a vehicle, adding a vehicle and doing nothing.

Ren(1995) develops stated preference vehicle transaction choice models conditioning on households' holding vehicles in order to forecast households' demand for alternative-fuel vehicles. But he models and forecasts for one-vehicle households and two-vehicle households separately. Due to the limitation of the stated preference experiments, households were asked to either add a vehicle or replace a vehicle. Thus, Ren(1995) only include these two kinds of vehicle transaction choices in his models. Using a vehicle technology scenario in 1998, Ren(1995) forecasts that the purchase share percentages for gasoline vehicles, electric vehicles, compressed natural gas vehicles and methanol vehicles are 63.6, 3.2, 19.3 and 14.3 correspondingly. The forecasts seem to be far away from reality today. Several factors may have caused the biased forecasts. One important factor may come from using the pure stated preference transaction data because there are serious biases in the SP data.

My study will overcome the shortcomings in Ren(1995) in several aspects. One is to estimate the alternative-fuel vehicle demand for one-vehicle and multiple-vehicle households together using one single model. Although this increases the complexity of the estimation model, it helps to explain most of the common effects over the whole population. The second is to obtain households' revealed preference vehicle transaction data and jointly estimate households' stated and revealed preference data together so as to correct the biases in the pure SP estimates. The third is to create a more complete choice set for households for the auto transaction choice model by including "doing nothing" besides vehicle replacement and addition. The fourth is to use a better vehicle classification scheme which reduces the measurement errors. Ren(1995) uses a rough vehicle classification scheme which contains 14 aggregated vehicle classes[6] and each of the vehicle class is determined only by vehicle make, model and series. However, in my study I use a more detailed vehicle classification scheme which will allow more variations of vehicle attributes. The new vehicle classes are determined not only by the make, model and series of the vehicles but also depends on the price tiers (high or low) and the origin (domestic or import) of the vehicles.

Another recent paper by Jong (1996) develops several submodels: a vehicle holding duration model, a vehicle type choice model and a vehicle usage model. The models are all based on respondents' actual vehicle purchasing information. The vehicle holding duration model estimates the time (in months) which elapses between purchase of a vehicle and its replacement. The vehicle type choice model explains households' preferences toward

[6] Ren(1995) uses the following 14 aggregated vehicle classes: mini cars, subcompact cars, compact cars, mid-size cars, large cars, luxury cars, sports cars, compact pickups, standard pickups, compact vans,

certain vehicle types. The vehicle usage model estimates the annual mileage for each panel car and annual energy use per kilometer for the car. Although Jong (1996) considers several aspects of households' vehicle demand together, there are some shortcomings in his paper. First, due to data availability all his models focus on vehicle replacement. Second, the panel survey is more a car survey than a car driver survey. (Household unit is not chosen as the unit of the study.) Therefore, it is very likely that not all the vehicles in a multiple-vehicle household will be taken into consideration in Jong (1996)'s models. Third, there is very limited household demographic information except for detailed driver characteristics. Fourth, some important vehicle attributes (such as acceleration time, range etc.) are neglected in the vehicle type choice model. My study will improve upon the vehicle demand model by interacting important vehicle attributes with detailed household demographic information (such as household member employment change, children-turning-16 etc.) and utilizing information on all of the households' existing vehicles.

Using the stated preference data from the same survey which Ren(1995)'s study is based on, Brownstone and Train(1999) apply "mixed logit" to forecast the demand for alternative-fuel vehicles in the long-run. Because it is assumed in multinomial logit models that the IIA property holds, MNL models predict that a change in the attributes of one alternative(or the introduction of a new alternative) changes the probabilities of the other alternatives proportionally such that the ratios of the probabilities remain the same. This substitution pattern may seem unrealistic in many cases. Even for nested logit models, it still

---

standard vans, mini sport utility vehicles, compact utility vehicles and standard sport utility vehicles.

assumes that the IIA property holds between alternatives within each nest. Therefore, neither multinomial logit models nor nested logit models provide us with a general substitution pattern between alternatives. However, mixed logit can generate general substitution patterns by specifying certain distribution of the stochastic part which is incorporated in the utility function to specify the correlation over alternatives and people.

According to Brownstone and Train(1999), mixed logit models can be represented by specifying the utility function as

$$U_i = b'X_i + [h_i + e_i]$$

Where U_i is the utility from alternative i; X_i is a vector of the observed attributes associated with alternative i and the person; β is a vector of estimated parameters which are assumed to be fixed over alternatives and people; η_i is the stochastic part which may be correlated over alternatives and people and whose distribution depends in general on underlying parameters and observed data relating to alternative i; ε_i is the stochastic part which has zero mean and is independently, identically distributed(iid) over alternatives and people. Therefore, by allowing any distribution for h while letting e be iid extreme value, we can write the conditional probability as a simple logit probability expression given the value of h:

$$L_i(h) = \frac{\exp(b'X + h_i)}{\sum_j \exp(b'X + h_j)}.$$

Thus the unconditional choice probability can be expressed as:

$$P_i = \int L_i(h)\, f(h \mid \Omega)\, dh\ .$$

The above probability is calculated through simulation because the integral does not have a closed form solution in general. The simulated choice probability is:

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

where η^r is the rth draw from its distribution, R is the number of draws of η, SP_i is the simulated probability and it is an unbiased estimate of P_i for any R. Therefore, the parameters are estimated by maximizing the simulated log-likelihood function $\Sigma_n \ln(SP_{ni})$ where n is an index to individuals. Brownstone and Train(1999) point out that different types of mixed logit models can be set up by specifying different distributions for η. Hence we can obtain different substitution patterns. According to McFadden (1996), any random utility model can be approximated by a mixed logit model including standard logit and nested logit models by specifying proper error-component structure of the utility function.

Using the stated preference choice data for alternative-fuel vehicles from a sample of 4654 California households, Brownstone and Train (1999) compare both the estimates of the standard logit and mixed logit model and the estimates of the probit and mixed logit model. They specify the mixed logit model by imposing certain error-component structures on non-electric vehicles, non-CNG vehicles, vehicle size, and vehicle luggage space. By doing so, they introduce certain correlation structures between pairs of vehicles with different fuel types, different sizes and luggage spaces. They find that the error components are very significant. At the same time, they allow the tastes of households regarding each vehicle attribute to vary in the population. To do so, they specify the other coefficients in the model to be independently normally distributed in the population so that the means and standard deviations are both estimated. They found that the ratios of estimated parameters are very similar in the standard logit and mixed logit models although the magnitudes of the estimated parameters in the non-stochastic portion of the utility function is generally larger in the mixed logit than the standard logit.

Furthermore, Brownstone and Train (1999) compare the standard logit model against the mixed logit models by identifying the differences in the predicted substitution patterns from these models when the forecasting scenario is changed (in other words, when a new alternative-fuel vehicle is introduced to the base scenarios or the price of the gasoline vehicle is changed). They find that the mixed logit models provide more reasonable and realistic substitution patterns although those models predict higher shares for alternative-fuel vehicles than the standard logit.

However, the predicted probabilities calculated in Brownstone and Train (1999) do not represent the actual forecasts. As it is mentioned in Brownstone and Train (1999), in order to use the models for forecasting, it is necessary to calibrate the model against real-world choices, develop a set of representative households for the forecast year and develop forecasts of vehicle characteristics that realistically represent future offerings or the range of possible future offerings.

Brownstone, Bunch and Train(1999) provide more realistic forecasts in several ways. First, they combine households' stated and revealed preference information about their purchased vehicle type choices. They find that the revealed preference data provide more realistic body-type choice information while the stated preference data is necessary for obtaining attributes of alternative-fuel vehicles which are not available in the marketplace. Second, they use a large set of households for the forecast with demographic information such as household size, annual income and education level. Third, they adopt a more realistic forecasting scenario file which contains a comprehensive set of new vehicle technology forecasts for year 1998 prepared by the California Energy Commission. As a result, Brownstone, Bunch and Train(1999) find that the joint SP/RP models give quite different forecasts from the pure SP model and the mixed logit models predict much higher shares for the alternative-fuel vehicles than the standard logit models.

The models in Brownstone, Bunch and Train(1999) can be viewed as "static", long run vehicle demand models because they simply look at households' vehicle purchasing type choices. Their models are not conditional on households' holding vehicle information and households' transaction choices are not incorporated in these models either. In contrast, the models in my study can be viewed as more "dynamic" and short run vehicle demand models in the sense that the models are conditioned on all of the households' holding vehicle information and the models incorporate households' vehicle type choice as well as their transaction type choices.

2.2 *Studies of Alternative-Fuel Vehicles*

As it is mentioned above, Ren(1995) apply the conditional multinomial logit model and use households' stated preference vehicle purchasing information to predict the demand for future alternative-fuel vehicles.

Besides Ren(1995), some earlier studies can also be found on predicting alternative-fuel vehicle demand. Examples are: SRI(1978), Beggs, Cardell and Hauseman(1981), Train(1980), Calfee(1985), Bunch et.al. (1993). Most of these studies focus on forecasting demand for electric vehicles using pure stated preference data. Ren(1995) has detailed discussions of all of these papers.

Golob et. al. (1997) develop models of household vehicle use behavior by type of vehicle. They make use of not only households' revealed vehicle utilization information but also households' stated vehicle utilization information for future alternative-fuel vehicles. They estimate the models separately for one-vehicle and two-vehicle households through specifying structural equation systems and they find that the joint RP-SP model is advantageous in simultaneously capturing the endogenous effects of vehicle reallocation along with perceived changes in utilization associated with electric vehicle characteristics. Households' behavioral vehicle use characteristics, physical vehicle characteristics and households' structural characteristics are used to model households' vehicle usage, the annual vehicle miles traveled (VMT). The model output can be applied to a dynamic microsimulation forecasting system which includes a sociodemographic transition model and vehicle transaction models.

2.3 *Studies of Estimating from Stated and Revealed Information Jointly*

Hensher(1993) points out the complementary characteristics of stated preference methods to revealed preference methods and the growing importance of stated preference methods in travel behavior research especially in travel demand modeling and prediction. Hensher(1993) introduces several different type of data which stated preference methods have been applied to and he also describes the procedure of stated preference experimental choice design in demand modeling applications. Basically, there are seven steps:

(i) identification of a set of attributes which are important in demand estimation and need to be included in the design;

(ii) selecting the measurement unit for each attribute;

(iii) specification of the number and magnitudes of attribute levels;

(iv) statistical design to generate alternatives by combining attribute levels and making them orthogonal;

(v) translation of the design into a set of questions and showcards for data collection purpose;

(vi) selection of appropriate estimation methods;

(vii) obtaining market share predictions by applying the estimated parameters in forecasting.

Hensher(1993) also discusses the method of joint estimation of SP-RP data by properly

$$U_{rp} = a + bX_{rp} + jY + e_{rp},$$
$$U_{sp} = d + bX_{sp} + gZ + e_{sp},$$
and
$$q^2 = \mathrm{var}(e_{rp}) / \mathrm{var}(e_{sp}),$$

scaling the SP data. The model can be expressed as:

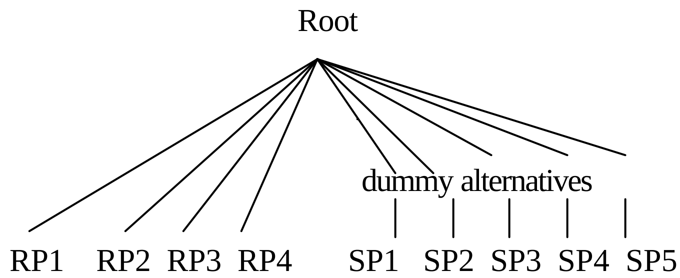
where X_{rp}, X_{sp} = a vector of observed variables common to RP and SP data;

Y, Z = vectors of observed attributes specific to RP and SP data respectively;

ε_{rp}, ε_{sp} = the unobserved effects associated with the RP and SP data;

θ = the scaling parameter, enabling the joint estimation of the two data sets.

In addition, Hensher (1993) points out that the scaling method is equivalent to an artificial nested logit structure with two subsets of alternatives. One subset is RP alternatives each of which is placed just below the "root" of the nest; the other subset is SP alternatives each of which is placed in a single-alternative "nest". The tree structure can be shown below.



Therefore, the joint RP and SP model can be estimated efficiently through the full-information maximum likelihood method on the nested structure.

Morikawa(1994) specifies the joint SP-RP model as the following to correct for the state dependence and correlation of the SP and RP data.

$$U_{in}^{RP} = V_{in}^{RP} + e_{in}^{RP},$$
$$U_{in}^{SP} = V_{in}^{SP} + e_{in}^{SP},$$
$$e_{in}^{RP} = l_{in} + v_{in}^{RP},$$
and
$$e_{in}^{SP} = q_i l_{in} + v_{in}^{SP},$$

Where RP indicates revealed preference data; SP indicates stated preference data; λ denotes individual and alternative-specific preferences and its probability density function is assumed to be the standardized multivariate normal distribution; θ_I is an unknown coefficient on λ; v is assumed to follow a white noise and Gumbel distribution. Therefore, the state dependence and serial correlation of SP and RP data are introduced through λ and v.

CHAPTER 3
STATED AND REVEALED PREFERENCE VEHICLE TRANSACTION MODELS

3.1 *Stated Preference Method versus Revealed Preference Method*

Stated preference is a method used to elicit respondents' preferences for new attributes, or for changes in existing attributes, by having them choose among hypothetical alternatives. Stated preference methods have been widely used in many studies especially in transportation research to identify behavioral responses to choice scenarios which are not revealed in the market.

There are both advantages and shortcomings of using stated preference methods.[7] Stated preference methods can generate alternatives and elicit preferences for new or non-existing alternatives; there are no measurement errors for attributes; multicollinearity can be avoided through orthogonal experiment design; ranges for explanatory variables can be extended to create a great deal of variation; and the choice set is specified beforehand. The major disadvantage of using SP data is that they may not be consistent with respondents' actual behavior. People may make implicit assumptions about the surveys, or they may attempt to influence policy, or they do not consider the constraints they have to face in the real world, or they do not need to make any real commitment about their choices. Therefore, their answers may be "biased".

---

[7] A similar comparison of stated and revealed preference methods can be found in Morikawa(1994) and Small and Winston(1998).

In contrast, revealed preference data reflect consumers' actual market behavior. Revealed preference shows respondents' actual observed choices among different existing alternatives. However, revealed preference methods can not forecast responses to non-existing alternatives, or to changes in the attributes of alternatives. Moreover, revealed preference data may include measurement errors, correlated attributes and limited attribute ranges. The choice set for revealed preference is ambiguous in some cases. Therefore, it is necessary to use stated preference methods in some studies and it is also advantageous for researchers to combine both stated and revealed preference information and take advantage of their complimentary characteristics in model estimations.

3.2 *Stated and Revealed Preference Vehicle Transaction Information*

I use two data sets in this study: one is households' observed actual vehicle transaction choices and the other is respondents' stated preferences for transactions associated with the given hypothetical alternative-fuel vehicles. Both of the data sets are from the panel survey conducted by University of California's Institute of Transportation Studies at Irvine and Davis for California Alternative-Fuel Vehicle Demand Forecasting Project.

3.2.1 *California Alternative-Fuel Vehicle Demand Forecasting Project*

The California Alternative-Fuel Vehicle Demand Forecasting Project was initiated and sponsored by Southern California Edison Company, Pacific Gas and Electric Company, and the California Energy Commission. The panel survey started in May 1993 and covered most of urbanized California excluding San Diego County. The survey sample was identified using geographically stratified pure random digit dialing. Respondents are geographically stratified by Zip Codes into 79 urbanized areas of California excluding San Diego County. The population studied was 22.7 million people(1990), about 75% of the total population of the state.

The panel survey consists of a three-wave survey collected at an interval of about 15 months. Each wave survey except the third-wave survey is complete through two stages: first stage is the computer-aided telephone interview (CATI-1) and the second stage is the mail-out questionaires (CATI-2) following the CATI-1 survey. The first two waves are used in this study[8].

---

[8]We are only able to use data from the first and second wave survey data because there is not enough funding to get the third-wave survey data ready for use. This panel survey is one of the largest regional survey of personal motor vehicle ownership and usage patterns and of stated future purchase references for alternative-fuel vehicles. Some basic data analysis of the survey is documented at ITS, University of California at Irvine.

3.2.2 *Stated Preference Vehicle Transaction Choice Data*

The initial survey, Wave-1 CATI-1 survey was completed for 7,387 households. It 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 (See Appendix I). If the next transaction was likely to involve a purchase, the survey asked for body type, size, and its approximate purchase price (including whether new or used). These data were used to customize the SP choice experiments that were later sent to the household so that the questions would be appropriate and relevant to the household. (Appendix II is an example experiment.)

The mail-out questionnaire contained two stated preference discrete-choice tasks for each household. Each task described six hypothetical vehicles, from which the households were asked to choose their preferred vehicle. These hypothetical vehicles included both alternative-fuel and gasoline vehicles, and the body types and prices were customized to include vehicles that were similar (but not identical) to the household's description of their next intended vehicle.[9] The survey also includes information on households' housing unit attributes related to ability to refuel EV or CNG vehicles at home.

[9] There are four fuel-type vehicles: gasoline, CNG, methanol and EV. In the Wave-1 CATI-2 SP questions, each households was given three out of the four fuel-type vehicles. In addition, these different fuel types are combined with vehicle body types such as car, truck or van to create six different hypothetical vehicles for households to choose from.

In addition, these households were asked about whether they intended to replace an existing vehicle with the preferred vehicle, or to simply add the preferred vehicle. If they indicated a replacement, the survey asked which vehicle was to be replaced (see Appendix II). In total, 4,747 households successfully completed the mail-out portion of the survey in Wave1 which represents a 66% response rate among the households that completed the initial CATI-1.

Compared to Census data, the sample is slightly biased toward home-owning, larger households with higher incomes. Among these 4,747 households only 1,626 households have valid stated preference vehicle transaction information and have 3 or fewer holding vehicles[10]. Therefore, only these 1,626 households are included in my study as the observations having stated preference data. But I have estimated the same pure stated preference vehicle type choice models using the 4,747 households and the 1,626 households and found no statistical differences between the model estimates.

[10] For the 4747 households, we only have desired and consistent vehicle class definitions for the 2243 households who have completed the Wave-2 CATI-1 information as well. This is because in the Wave-2 survey, we adjusted the vehicle class definitions to make them more meaningful and precise. The new vehicle classification scheme determines vehicle body types, import/domestic, high/low-price tier according to the makes, models and series of the vehicles. While the old vehicle classification scheme only identifies vehicle body types in a very rough manner. Therefore, the new vehicle class specification scheme is adopted and is only applicable to those households who participate in the wave-2 survey. Among the 2243 households, some do not have valid answers to the stated preference questions, some do not have valid vehicle information such as correct vehicle vintage, and some own more than 3 vehicles. Therefore, by excluding those households with conditions mentioned above, only 1626 households are left with the appropriate stated preference transaction data for the model estimations in this study.

3.2.3 *Revealed Preference Vehicle Transaction Choice Data*

Among the 7,387 households who received the initial telephone interview in the Wave1, only 2,857 households participated in the CATI-1 of the Wave-2 survey. The 2,857 households include 2,243 households out of the 4,747 households who completed the Wave-1 SP experiments and 614 households who only finished the Wave-1 CATI-1. In the Wave-2 CATI-1, households were asked about whether there is any change to their vehicle stock (for example, whether the household acquired any vehicle since the Wave-1 survey). Therefore, according to households' revealed data, we are able to figure out what kind(s) of vehicle transactions the households actually went through during the approximately one-year period between wave-1 and wave-2 surveys. [11]

A list of information included in Wave-1 CATI-1 and Wave-2 CATI-1 surveys are shown in Table 1 below.

[11] There are a set of rules that were set up to determine the transaction types that the households went through based on their vehicle acquiring/dropping information. One of the basic assumptions is that there is only one vehicle transaction occurring each six month interval. If a household acquires a vehicle within six months around the time it drops an existing vehicle, the vehicle transaction is considered to be a vehicle replacement. Otherwise, it is considered to be a vehicle addition. If the households have multiple-vehicle transactions, the order of the transactions is determined by the relative timings of the transactions.

Table 1. List of information included in Wave-1 CATI-1 and Wave-2 CATI-1 surveys

Information On	Wave-1 CATI-1 (05/93)	Wave-2 CATI-1 (08/94)
Households' Current Vehicles (make/model/year/color)	√*	√
Households' Future Vehicle Replacement Behavior (which to replace/when/what to purchase/type/size/expenditure)	√	N/A
Households' Residences (physical characteristics)	√	√
Households' Socioeconomic Characteristics (employment status/age/gender/ education level/driver status)	√	√
Household members' demographic changes (moving in/leaving/employment status changes/kid-turning-16 etc.)	N/A**	√
Household's location change(s)	N/A	√
Household's vehicle stock changes (vehicle entering into/leaving/reasons)	N/A	√

* Note that √ means that the corresponding information is available for the survey.
** Note that N/A means that the corresponding information is not available for the survey.

From the above table, we can see that both CATI-1 surveys contain information about households' demographic characteristics and vehicle holdings. While in Wave-2 CATI-1 survey, there is additional information on households' vehicle stock changes and demographic changes. Therefore, according to the records of households' vehicle stock changes, we are able to construct households' vehicle transaction information and their corresponding demographic change information.

Out of the 2857 households, 1097 households (about 38% of the sample) had at least one vehicle transaction (including vehicle replacement, addition, and deletion) while the other

1760 households (about 62% of the sample) did not change their vehicle stocks. Due to the large amount of work associated with identifying the "real" vehicle transactions[12] that happened within each sampled household, I only got to identify the first "real" vehicle transaction for each household even if some households may have had multiple vehicle transactions.

Among the 1097 households with vehicle transactions, only 749 households have valid information on their vehicle replacement or addition.[13] Out of the 749 households 665 households have 3 or fewer vehicles after their vehicle transactions and I take these households as the group of households who had real vehicle transactions in my study. I chose these households in our study because it is assumed that there is a single decision maker in each household who decides which vehicle to buy each period. Households ended up with more than three vehicles are more likely to have more than one "decision maker". Moreover, including households who have 4 or more vehicles will definitely add to the complexity of the estimation model. Besides, only 84 of the 749 households have 4 or more vehicles. Therefore, I exclude these 4-or-more-vehicle households from my study.

[12]Some of the vehicle changes are excluded as vehicle transactions according to the answers given by the respondents about how the vehicles left or came into the households. For example, if a vehicle is taken by some household member who left the household, we would consider this as a pure household demographic change instead of a vehicle transaction. A few households only had vehicle deletions, they are not studied here either.

[13]In order to look into the vehicle transactions, we have to have complete information on the purchased vehicles and households' existing vehicles. The information includes vehicle class, high/low price tier, import/domestic, and vehicle vintages. Such information can be used to identify the corresponding vehicle attributes which are necessary in the model estimations.

The distribution of the number of vehicles for the 665 transaction households is shown in Table 2 below.

Table 2: Distribution of Number of Vehicles Owned by the 665 Transaction Households

Number of Vehicles	Frequency	Percentage	Cumulative Percentage
1	107	16.09	16.09
2	350	52.63	68.72
3	208	31.28	100.00
In Total	665	100.00	

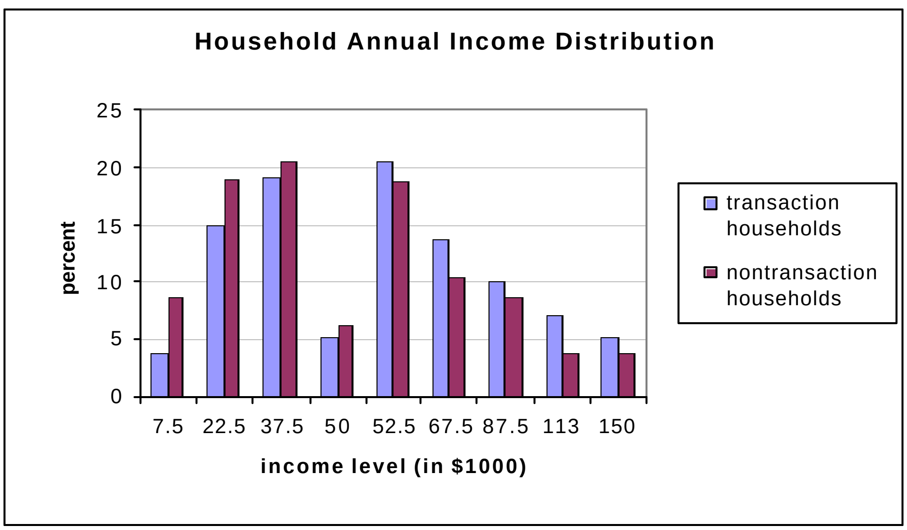
Similarly, for the 1760 households who did not have any vehicle changes over the period, 1616 households have valid information on households' existing vehicle information and 1561 of them are households with 3 or fewer vehicles. Therefore I only include these 1561 non-transaction households along with the 665 transaction ones in my study. The distribution of the number of vehicle for the 1561 non-transaction households is shown in Table 3.

Table 3: Distribution of Number of Vehicles Owned by the 1561 Non-Transaction Households

Number of Vehicles	Frequency	Percentage	Cumulative Percentage
1	662	42.41	42.41
2	748	47.92	90.33
3	151	9.67	100.00
In Total	1561	100.00	

The comparisons of the income distributions are shown in Figure 1. The detailed summary statistics are given in Appendix III. The statistics summarize the household annual income statistics for all the 2857 households in the wave-2 CATI-1 survey, for the 2226 households that are included in my study, for the 665 households who had vehicle transaction(s) among the 2226 households and for the other 1561 households without a vehicle transaction. As expected, those households who had vehicle transactions have a slightly higher average income than non-transaction households. The actual observed vehicle transaction information for these 2226 households becomes the households' revealed preference vehicle transaction data for my study.

Figure 1. Comparison of the Income distribution for the transaction households versus non-transaction households



Among the 665 households who had vehicle transactions during the study period, about 35.5 percent purchased brand new vehicles. The vintage distribution of the vehicles purchased by these households during this period is shown in Figure 2. Among the vehicles bought by the 665 households, the percentages for small cars, midsize or large cars, sport cars, pickup trucks, vans and sport utility vehicles are 33%, 24%, 7%, 13%, 11% and 12%

respectively. The purchased-vehicle body-type distribution is shown in Figure 3. According to the statistics, households who had vehicle transactions also have more adults and more members than the non-transaction households.

Figure 2. The vintage distribution of the vehicles purchased by the 665 households

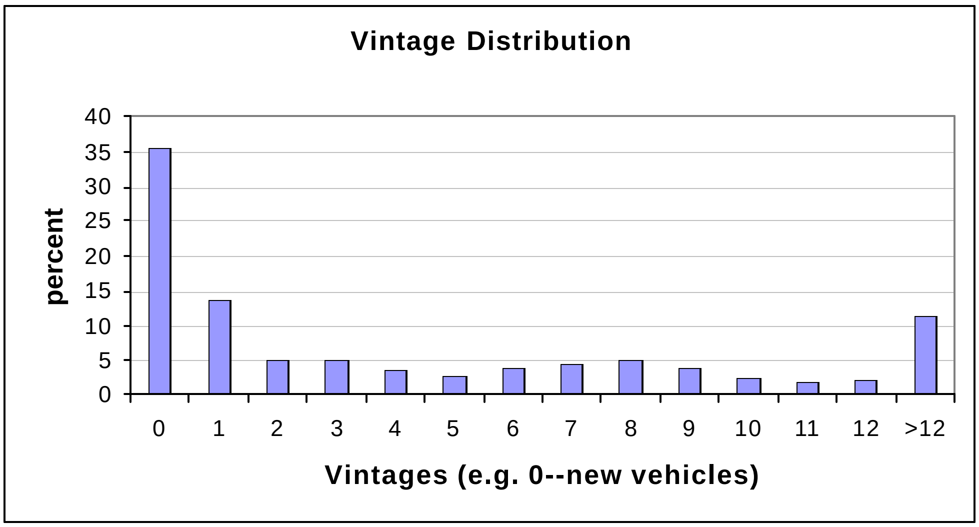
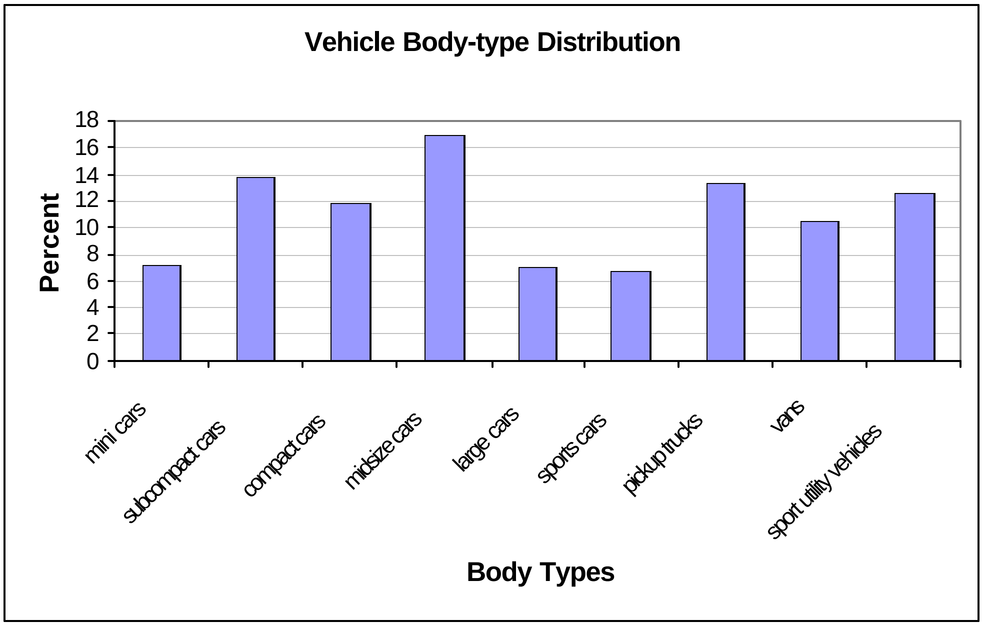


Figure 3. The body-type distribution of the vehicle purchased by the 665 households



3.3 *Theoretical Framework*

I assume that in each household there is a single decision maker, who decides which vehicle to buy, by maximizing the utility that the household gets from their vehicle fleet.

3.3.1 *Conditional Multinomial Logit Model*

Theoretically, the indirect utility function for household decision maker *n* choosing alternative *i* is given by the following equation(McFadden, 1974):

$$U_{in} = V_{in} + \varepsilon_{in}$$

where *n* is the observation index, *i* is the vehicle alternative index, V_{in} is the deterministic part of the utility function, and ε_{in} is the random component of the utility function. Here I assume ε_{in} is independently distributed with the extreme value distribution. I further assume V_{in} to be a linear combination of known characteristics of each household, its given transaction choices, and given alternative vehicles. Thus I have

$$V_{in} = \beta' X_{in}$$

where X_{in} is the vector of explanatory variables including all observable characteristics specific to the household, its holding vehicles, the SP/RP alternative vehicles and the SP/RP transaction choices; β is the coefficient vector of X_{in}. Therefore, the conditional probability of household *n* choosing alternative *i* given a set of choices can be expressed as:

$$P_{in} = P(y_{in} = 1) = \frac{\exp(V_{in})}{\sum_{j=1}^{J} \exp(V_{jn})} = \frac{\exp\left(b' X_{in}\right)}{\sum_{j=1}^{J} \exp\left(b' X_{jn}\right)}$$

where y_{in} is the vector of dependent variables: $y_{in} = 1$ when alternative *i* is chosen by the household *n*, and $y_{in} = 0$ otherwise. The log likelihood function is as follows:

$$L = \sum_{n=1}^{N} \sum_{j=1}^{J} y_{in} (b' X_{in} - \ln \sum_{j=1}^{J} \exp(b' X_{jn}))$$ where J is the number of alternatives.

The probability formula of the multinomial logit model implies that

$$\ln\left[\frac{P_{in}}{P_{jn}}\right] = \beta' X_{in} - \beta' X_{jn} .$$

Noticed that the ratio, P_{in}/P_{jn}, does not depend on the other alternatives following the original assumption of independent and homoscedastic disturbances. The property is called independence of irrelevant alternatives (i.e. the IIA property) and the ratio is called the odds ratio.

However, in reality this is a strong assumption because it is likely that a subset of the choices are correlated. There are several ways to test the validity of the IIA property. Hausman's specification test (see Hausman and McFadden(1984)) is one of them and its statistic is defined as the following:

$$\chi^2 = (\hat{b}_s - \hat{b}_f)'\left[\hat{V}_s - \hat{V}_f\right]^{-1}(\hat{b}_s - \hat{b}_f),$$

where *s* stands for the estimators based on the subset choices, *f* shows the estimators of the full set of choices, and $\hat{V}_s$ and $\hat{V}_f$ are the respective estimates of the asymptotic covariance matrices. The Hausman statistic is asymptotically distributed as chi-squared with K degrees of freedom, where K is the number of elements in the coefficient vector that is identifiable from the subset choice model.

3.3.2 *Nested Logit Model*

If the test mentioned above fail to support the IIA assumption, a nested logit model serves as an alternative to the multinomial logit model which frees up the IIA assumption of the multinomial logit model. One can construct the nested structure by grouping the alternatives into subgroups so that the odds ratio between any two alternatives from different subgroups is no longer independent of the attributes of the other alternatives while the odds ratio is still independent of the existence of the other alternatives for any two alternatives within the same subgroup. In other words, IIA still holds within each nest but not across nests.

Following Train(1986), the full set of altenatives J_n is divided into K subsets denoted by $B_n^1,...,B_n^K$. The utility function for household n choosing alternative *i* is expressed as

$$U_{in} = V_{in} + \varepsilon_{in} ,$$

where generalized extreme value(GEV) distribution is assumed for all ε_{in}s in J_n. The joint cumulative distribution of ε_{in}s is assumed to be

$$\exp\left\{-\sum_{k=1}^{K} a_k \left(\sum_{i \in B_n^k} \ell^{-e_{in} / l_k}\right)^{l_k}\right\},$$

where the parameter λ_k is a measure of the correlation of ε_{in}s within each subset B_n^k.

According to McFadden (1978), the probability for household *n* choosing alternative *i* in subset B_n^k can be expressed as the following given the above joint distribution of ε_{in}s:

$$P_{in} = \frac{e^{V_{in} / l_k}\left(\sum_{j \in B_n^k} e^{V_{jn} / l_k}\right)^{l_k - 1}}{\sum_l \left(\sum_{j \in B_n^l} e^{V_{jn} / l_l}\right)^{l^l}}.$$

The above probability function can be decomposed as follows:

$P_{in} = P_{in|B_n^k} . P_{B_n^k}$,

where $P_{in|B_n^k}$ is the conditional probability of household *n* choosing alternative *i* given that an alternative within the subset B_n^k is chosen; $P_{B_n^k}$ is the marginal probability of household *n* choosing any alternative in each subset. I can write the marginal and conditional probabilities as follows:

$$P_{in / B_n^k} = \frac{e^{Y_{in}^k}}{\sum_{j \in B_n^k} e^{Y_{jn}^k}}, \text{ and}$$

$$P_{B_n^k} = \frac{e^{W_n^k + l_k I_k}}{\sum_{l=1}^{k} e^{W_n^l + l_l I_l}}$$

where $I_k = \ln \sum_{j \in B_n^k} e^{Y_{jn}^k}$, the "inclusive value" which measures the average utility that a household gets from obtaining alternatives within subset B_n^k; vector Y contains attributes which are associated with the alternatives within each subset; vector W contains attributes which are associated with alternatives across each subset B_n^k (i.e. the attributes in $W_n{}^k$ do not vary across alternatives, i$\in B_n^k$).

The coefficient of the inclusive value λ_k is called the dissimilarity coefficient. It is related to the degree of correlation of unobservables for alternatives within B_n^k. When λ_k is equal to 1, the model is equivalent to a multinomial logit model in which the IIA property holds globally. When λ_k is close to 0, the unobservables for alternatives within nest B_n^k become perfectly correlated. McFadden(1978) proves that a sufficient condition for the nested logit model to be consistent with an underlying random utility maximizing model is for the coefficients of the inclusive values to lie within the unit interval.

There are two ways to estimate the parameters of the nested logit model. One is two-step maximum likelihood approach with limited information. The other is full information maximum likelihood estimation (FIML) which is more efficient. For FIML, coefficients are estimated by maximizing the log-likelihood function which can be expressed as

$$\ln L = \sum_{n=1}^{N} \sum_{J=1}^{J} P_{jn|B_n^k} * P_{B_n^k} * y_{jn},$$

where N is the number of households, J is the number of alternatives, K is the number of subgroups, y_{jn} is equal to 1 if alternative j (within node B_n^k) is chosen by household n and 0 otherwise.

3.3.3 *Methods of Joint Estimation of SP and RP Data*

Stated preference data is necessary because it provides attributes for alternative-fuel vehicles which are not available in the marketplace. As noted earlier, there are both advantages and shortcomings of using stated preference data. Revealed preference data reflect consumers' real market behavior. But the revealed preference method can not be used to obtain responses to non-existing alternatives. Thus, it is advantageous for researchers to combine both SP and RP data and take advantage of their complimentary characteristics in model estimations[14].

I use two data sets in this study: one is households' observed actual vehicle transaction choices and the other is respondents' stated preferences for transactions associated with the given hypothetical alternative-fuel vehicles. The idea of the RP/SP combined estimation method is to jointly estimate the RP and SP models instead of estimating the models separately using only SP or RP data.

According to the findings in my previous study (see Sheng(1997)), the common assumption that the trade-off relationship among different attributes are the same for SP and RP data is not always valid for each attribute. Therefore, I assume that the relative trade-offs between most attributes are different for the stated and revealed preferences except for the

[14] A similar comparison of stated and revealed preference methods can be found in Morikawa(1994) and Small and Winston(1998).

several major attributes such as net capital cost, operating fuel cost etc.. Thus, the deterministic parts of the utility functions can be expressed in more detail as below:

$$\begin{cases} V_{in}^{RP} = b'X_{in}^{RP} + a'Y_{in}^{RP} + w'Z_{in}^{RP} \\ V_{in}^{SP} = b'X_{in}^{SP} + l'Y_{in}^{SP} + g'W_{in}^{SP} \end{cases}$$

where X and Y are vectors of attributes that are common to both SP and RP data; Z and W are vectors of attributes that are specific to RP and SP data correspondingly. Variables in vector X share same β in both models while attributes in vector Y have different coefficients in the RP model, α and in the SP model, λ.

Since I assume that the random terms are IID Gumbel within either RP or SP data, MNL models can be derived for both RP and SP data. However, it is assumed in this paper that the variances for the SP and RP data are different. Following Morikawa(1994), their relationship is described below :

$$Var(e_{in}^{RP}) = m^2 Var(e_{in}^{SP}).$$

This means that I can estimate the RP and SP models together by simply pooling the RP and SP data while multiplying the utility function of the SP data by μ. Note that now the choice probability for the SP data is given as the following:

$$P_{in}^{SP} = \frac{\exp(m \cdot V_{in}^{SP})}{\sum_j \exp(m \cdot V_{jn}^{SP})}.$$

I use the "grid search" method to estimate the unknown parameter μ in STATA.[15] This is similar to other methods such as the sequential estimation method suggested in Morikawa et al. (1994).

[15] This method was suggested by David Brownstone. Basically I have to write a program to estimate the MNL model for each value of μ, compare the log-likelihood values and pick a value of μ that gives the highest log-likelihood value among all the models. The computing time and the accuracy of μ depend on the starting value of μ and the searching step length. This estimator is close to the true maximum likelihood estimator.

3.4 *Revealed Preference Vehicle Transaction Model Specification*

In the real world, households can choose to add another car, replace one of their existing vehicles, or do nothing i.e. keep their current vehicle composition.[16] Households also face a large universe of different vehicles to choose from. In order to reduce the model complexity, I have to sample the full set of choices and create a representative small choice set for the model estimation. Following from the IIA property, any reasonable sample from the full choice set yields consistent estimates of the full choice model parameters for the MNL models.)

Brownstone, Bunch and Train(1999) find serious problems with simple randomly sampled alternatives in the RP models[17]. Therefore, I strategically sampled 12 alternative vehicles with different attributes for each household from the universe of the existing new and used vehicles (there are 729 different vehicles in total). Basically, I stratified the sample according to the vehicle vintages so that each sampled choice set contains 3 new vehicles, 3 one-year or two-year old vehicles, 3 3-10 year old vehicles, and 3 more than 10-year old vehicles.

[16] Although there maybe some other choices as well, for example, disposal or scrappage, those choices are not considered in this study. Because only a few households made those choices and there does not seem to exist a very good way to model those choices.

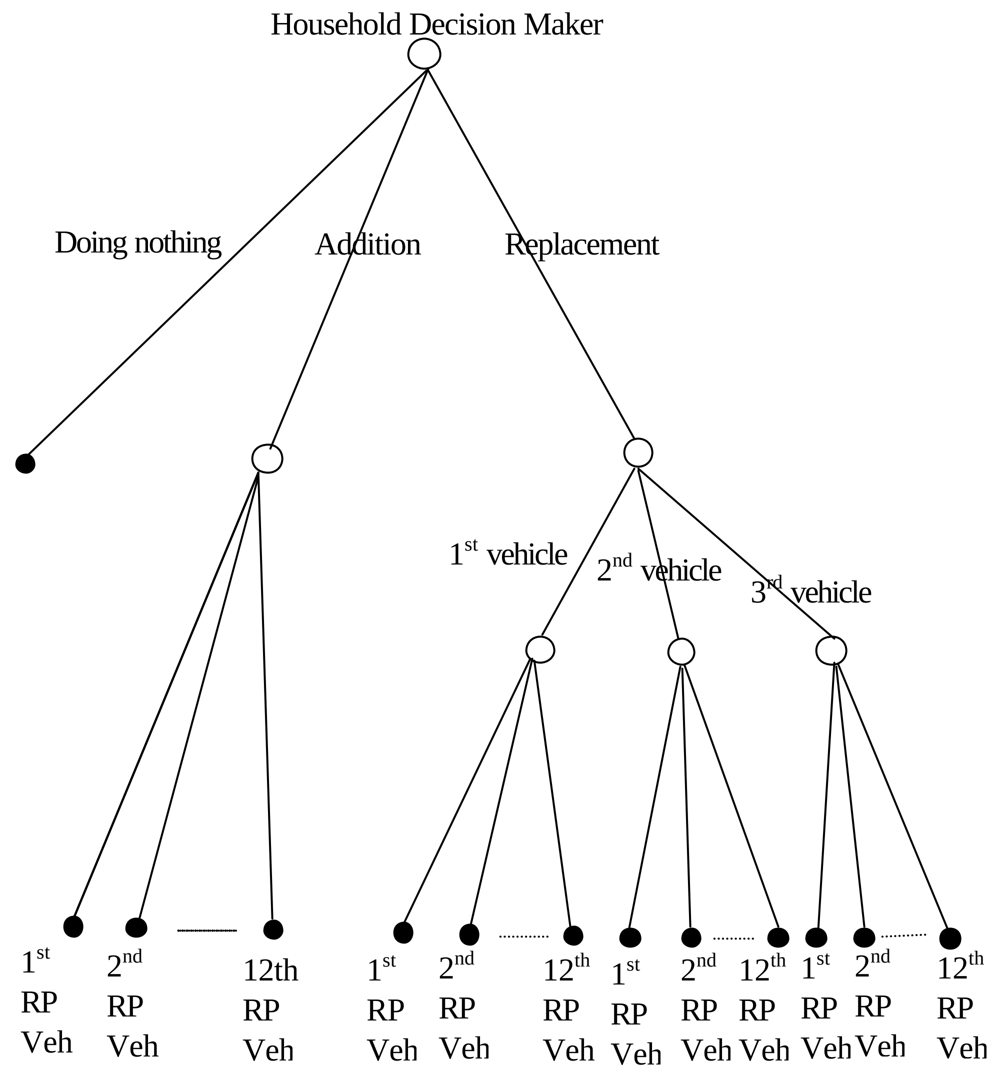
[17] Similar problem arises in this study. It is very likely that any sample of size 12 will contain only one or two new vehicles because there are all together 40 new vehicles out of the 729 vehicle choices. That will lead to implausibly high estimates for the new vehicle dummy variable since 236 (35.5%) of the 665 households chose new vehicles in the RP data.

3.4.1 *Description of the Model*

By assuming that there is a single decision maker in each household, I use a "hierarchy tree" structure to describe household's decision making process of vehicle transactions in the real world. The tree structure shown in Figure 4 describes the joint households' revealed preference vehicle transaction and type choice model. The upper level branches of the tree structure are the three major vehicle transaction choices: adding a vehicle, replacing a vehicle or doing nothing; the lower level branches involve the specific vehicle type choices under each transaction type choice.

Note that the "tree" structure is simply used to describe the process of households' transaction choice decisions and it does not necessarily imply any nested structure. Instead, it serves to better specify the MNL model by including the vehicle transaction choices as well as the vehicle type choices. The actual model is therefore a model which jointly estimates both the transaction type choices and the vehicle type choices. In other words, all the 48 alternatives in the lower level of the "tree" structure and the "doing nothing" choice are jointly estimated in the RP model. In addition, since I include all up-to-3-vehicle households in the model and the model is conditioned on households' holding vehicles, the "replacement" choice actually splits into three subgroups: replacing the first holding vehicle, replacing the second one, or replacing the third vehicle. (The order of the holding vehicles corresponds to the order in which respondents entered their vehicles.)

Figure 4. Multiple-Vehicle Household RP Transaction Tree



Compared to Ren (1995), the transaction model here has several advantages. First, this model is based on households' revealed preference data instead of stated preference data. The RP data reflect households' actual preferences under real-world scenario while the SP data is conditional on hypothetical vehicles which may have inherent biases. Second, this model endogenizes one of households' important transaction choices, "doing nothing". Because there is a limited number of households who actually go through vehicle transactions during each period, the information about households who end up doing nothing is valuable in justifying the whole sample and the probabilities for households' transaction choices. Third, I combine one-vehicle, two-vehicle and three-vehicle households together and model their choices using the same model. This helps us in looking at the overall effects for the whole sample and simplifies the process of having to estimate different models for these households.

Brownstone et. al.(1998) use a similar data set but their model focuses on households' vehicle type choices. Their model can be viewed as a more "static", long run vehicle demand model because they simply look at households that purchase vehicles. No holding vehicle information nor households' transaction choices are incorporated in their model. The model in my study can be viewed as a more "dynamic" short run vehicle demand model in the sense that it is conditioned on all of the households' holding vehicle information and it incorporates households' vehicle type choice as well as transaction type choices.

Unlike Smith, Hensher, and Wrigley (1991), "time" does not enter the models in this study except that vehicle vintage effects are taken into account. The models would be more “dynamic” if some vehicle duration measurement (i.e. household’s vehicle holding duration) is considered. However, I could not create such variables for the RP data because there are quite a few missing data for vehicle duration, although we did ask respondents when they obtained each of their holding vehicles.

Based on the “tree” structure and the basic assumptions that are made in the Methodology section, I apply the conditional multinomial logit model to the RP data. The estimation results are shown in Table 4. Note that the dependent variable in the RP model is a dummy variable which equals zero if the vehicle is not chosen and equals one otherwise. See Appendix IV for detailed summary statistics for the explanatory variables in the RP model.

3.4.2 *Revealed Preference MNL Transaction Model Estimation Results*

Table 4. Multiple-Vehicle Household RP Transaction Model Estimation Results

Explanatory Variables	Coeff.	T-STATs
Cost Variables		
net capital cost /ln(household income/$1k) (in $1k)	-0.356	-6.739
average fuel cost / ln(household income/$1k) (in cents/mile)	-0.504	-2.099
average fuel cost * non-transaction dummy	0.435	2.570
average fuel cost * addition dummy	0.393	1.858
Vehicle Performance Attributes		
average acceleration time (in seconds)	-0.057	-0.286
average top speed (in hundreds of miles/hour)	1.550	1.268
range of the purchasing vehicle (in hundreds of miles)	0.550	3.215
maximum range of holding vehicle(s) (in hundreds of miles)	2.590	2.462
squared maximum range	-0.355	-2.850
Vehicle Pollution Measurement		
tailpipe emission of the purchasing vehicle	0.334	3.710
Purchasing-Vehicle-Body-Type Dummies		
small cars	-0.255	-1.981
sport cars	-0.109	-0.339
pickup trucks	-0.211	-1.235
Vans	-0.940	-2.377
sport-utility-vehicles	0.655	2.242
mini-sport-utility-vehicles	-1.254	-1.587
Other Attributes of the Purchasing Vehicle		
natural logarithm of number of makes in the same-size class	0.797	10.460
luxury vehicles	-0.263	-1.546
foreign vehicles	-0.379	-2.928
new vehicles	1.191	6.077
1-year-old vehicles	0.419	2.038
natural logarithm of vehicle vintage	-0.316	-2.456
Household Portfolio-Choice Variables		
replacing-lower-value-vehicle dummy	0.228	1.418
replacing-same-body-type-vehicle dummy	0.675	5.482
oldest vehicle vintage	-0.050	-3.787
market value of the remaining holding vehicles (in $1k)	0.069	3.198
from no-car households to one-car households	0.209	1.592
from no-sport-car households to one-sport-car households	-0.398	-1.176
from no-van households to one-van households	0.908	2.458
from no-SUV households to one-SUV households	0.019	0.071
all domestic vehicle(s)	0.362	3.079
all imported vehicle(s)	0.474	4.158
Transaction-Type Dummy Variables		
"inertia" i.e. non-transaction choice	7.840	8.665
addition constant	0.123	0.303
Demographic Characteristics and Changes		
addition * number of vehicles less than or equal to number of drivers	1.818	13.394
addition * households with annual income<=$45k	-0.159	-0.967
addition * households with annual income>$90k	0.602	2.861
addition or replacement* household demographic change(s)	0.096	0.919
vans * household size >= 3	0.785	3.099
sport cars * household size >= 3	0.756	2.370

Note that: 1) the log-likelihood value for the above model is -3128.506;
2) the number of observations is 73458 and the number of households is 2226;
3) the pseudo R2 value is 0.595;
4) the "base" vehicle class is "midsize/large" cars.

3.4.3 *Explanatory Variables*

a. Net Capital Cost

The net capital cost represents the actual amount of money a household has to spend on the vehicle transaction. Therefore, the net capital cost is viewed as an important part of the total cost involved in the vehicle transaction. If a household simply adds a vehicle to its fleet, the net capital cost is the current price of the purchased vehicle; if a household purchases a vehicle to replace one of its existing vehicles, the net capital cost is the difference between the current market values of the purchased vehicle and the vehicle being replaced; if a household has no vehicle transactions, then the net capital cost is equal to zero. In order to capture the "income effect" on the net capital cost, I divide the net capital cost by the natural logarithm of household's annual income (both in $1,000). The "income effect" of the net capital cost is found in many other studies such as Brownstone and Train(1999).

The result in Table 4 shows that the net capital cost has a negative sign, as we expect, and it is significant. Given its nonlinear form, it shows that households with lower income are more sensitive to the net capital cost.

b. Average Operating Cost

The average operating cost is the average fuel cost of a household's new fleet after the vehicle transaction. It is equal to the average fuel cost of all of the existing vehicles if the household does not have any vehicle transactions. The average operating cost is in terms of cents per mile and is calculated based on the assumption that the gasoline price is $1.20/gallon. All the prices are in terms of 1995 dollars. It does not include maintenance cost because there is no good measurement for the maintenance cost at present. Moreover, the average operating cost is also divided by the natural logarithm of households' annual income (in $1,000) because I expect that there is an "income effect" on operating cost as well.

In order to look into the effect of operating cost on households with different types of vehicle transaction choices, I interact the average operating cost with the "inertia" dummy variable, i.e. the non-transaction constant and the "addition" dummy variable. The results in Table 4 shows that the average operating cost is negative and significant except for households without a vehicle transaction or with a vehicle addition. This indicates that households replacing vehicles prefer to purchase vehicles which will lead to lower average fuel cost for their fleets, but households adding a vehicle do not pay too much attention to fuel cost. The lower-income households are more sensitive to this than higher-income households.

c. Average Acceleration Time

Acceleration time is the average time in seconds required to reach 30 mile/hour from a complete stop. The average acceleration time of the household's new fleet is shown in Table 4 to be almost "zero". Unlike the stated preference experiment in which we can control for the variations and correlation for different variables through experimental design, in the revealed preference data there is actually not much variation for acceleration time and there is most likely high collinearity between acceleration time and other vehicle attributes such as range and top speed. Therefore, the estimation for the average acceleration time in the RP model is expected to be imprecise.

d. Average Top Speed

Top speed is the highest speed that a vehicle can attain and it is in terms of hundreds of miles per hour. Similar to the average acceleration time, the estimated coefficient for the average top speed in the RP model is also expected to be imprecise although it shows to be significantly positive in Table 4.

e. Range of the Purchasing Vehicle

Range is defined as the distance a vehicle can travel between refuelings/rechargings (in hundreds of miles). There is little variation among the ranges of the purchased vehicles. Therefore the estimate is also expected to be imprecise although it is positive and significant in Table 4.

f. Maximum Range for the New Holding Vehicles

Multiple-vehicle households, probably do not care about the range of every vehicle in their fleet as long as they have at least one vehicle whose range is long enough for them to make certain trips. This is because those households with multiple vehicles can put vehicles with different ranges into different uses. Therefore, I put maximum range in the RP model in addition to the range variable. Because there is little variation of the ranges in the RP data, it is expected that the maximum range has a high correlation with the range variable.

In order to reflect the marginal effect of the maximum range variable on the utility function, the square term of the maximum range is included as well. From the estimates in Table 4, we can see that the coefficient on the squared range term is negative and significant which indicates decreasing marginal utility due to a unit increase of the maximum vehicle range. Nevertheless this marginal utility is still positive for ranges as large as 365 miles[18] .

g. Vehicle Pollution Level

Vehicle pollution level is a measurement of vehicle tailpipe emission. It is defined as the tailpipe emissions as a fraction of comparable 1995 new gas vehicles. However, we do not have accurate data to serve as a good measurement for the pollution variable. We used the pollution limits in effect in the RP data when the vehicles were new. In the RP data, the pollution variable is highly correlated with vehicle vintage and vehicle body types such as

[18] The number can be calculated by solving the equation for marginal utility of range. It is equal to [2.590/(2*0.355)]*100 i.e. 365 miles.

trucks. Therefore, it is difficult to interpret the estimate for the pollution variable in our RP data although it shows to be positive and significant in Table 4.

h. Vehicle Body Type Dummy Variables

Vehicle body type is considered to be a very important vehicle attribute. We classify vehicle body types into several categories: small cars (including mini cars, subcompact cars, and compact cars), large cars (including midsize and large size cars), sport cars, trucks (including compact pickups and standard pickups), vans (including compact and standard vans), mini sport utility vehicles and sport utility vehicles (including compact and standard sport utility vehicles). In the RP model, I leave the large-size car category as the base category. Therefore, the negative coefficient of small cars shows that households prefer large-size cars to small cars. From Table 4, we can see that except for sport utility vehicles, vehicles with other body types are less preferred by consumers.

i. Import/Domestic Vehicles

A vehicle is considered to be an import vehicle if the vehicle has an import nameplate. Otherwise, it is considered to be a domestic vehicle. In order to look at households' choices of their fleets, two dummies are created: one is a dummy indicating that every vehicle of household's fleet is an import vehicle, the other is a dummy indicating that every vehicle of household's fleet is a domestic vehicle. The left out category is the case where households own both import and domestic vehicles. According to the estimates in Table 4, households prefer to have all import vehicles or all domestic vehicles.

j. Luxury Vehicles

A vehicle is viewed as a luxury vehicle, i.e. higher-price-tier vehicle if the vehicle is a "luxury" model within a vehicle body-class. The coefficient of luxury vehicle dummy is negative and insignificant.

k. Vehicle Vintages

The vehicle vintage shows how old the vehicle is. It equals zero if the vehicle is a new vehicle; it equals one if the vehicle is one-year old and so on. In the RP model, I create a dummy for new vehicles, and one dummy for one-year-old vehicles because we expect that there is a big difference between the willingness to pay for a new vehicle and a one-year-old vehicle for consumers. Then I take the natural logarithm of vehicle vintages for vehicles more than one-year old since we expect that the depreciation rate for vehicles is of non-linear form. From Table 4, we can see that a new vehicle is worth much more than a one-year old vehicle by comparing their coefficients and vehicle value drops gradually according to a log-linear form as the vehicle gets older.

l. Number of Vehicles in Class

I calculate the natural logarithm of the number of vehicles in a class. The result in Table 4 confirms the hypothesis that if there are more vehicles in some class, consumers prefer the vehicles in that class more.

m. Transaction-Type-Choice Dummies

I put the non-transaction or "inertia" and the "addition" dummy in the RP data to capture the unobserved transaction cost associated with non-transaction or adding a vehicle. The left out choice is replacing a vehicle. Both the "inertia" dummy and the "addition" dummy are alternative-specific constants. Putting in the two dummies helps explain the unexplained factors in the RP model, therefore it is crucial to include them in the estimation.

n. Household Demographic Characteristics and Changes

According to many former studies such as Train(1986), a household's demographic characteristics play an important role in its vehicle transactions. So I interact several demographic factors with the transaction-type-choice dummies such as "inertia" or "addition" to reflect different effects brought by the demographic factors.

For example, I look at the choices for those households whose number of vehicles per driver is less than or equal to one. It turns out that the coefficient on the dummy is positive and significant which shows that households who have numbers of vehicles less than or equal to the number of drivers are more likely to add vehicles to their fleet. I also interact "addition" with the high-income-class dummy and the low-income-class dummy. The results in Table 4 show that high-income households are more likely to add vehicles to their fleets and the effect is significant. However, when I put in the RP model the dummy which stands for households who have demographic change(s) and choose to add or replace a vehicle, the coefficient is positive but not at a very significant level. This may be because the other demographic factors have already helped explain major effects.

o. Household Portfolio Choice

It is observed that multiple-vehicle households purchase vehicle(s) and put them into different uses. So, maximizing the utilities that they can get from holding different combinations of vehicles is their objective. Therefore, I look into households' "portfolio choices" for their fleets.

There are several variables I put in the RP model which are associated with households' vehicle portfolio choices. The dummy variable, "replacing a cheaper vehicle" stands for households that replaced an existing vehicle with a more expensive one. Its coefficient is positive, though not significant. Another dummy, "replacing an existing vehicle with the same body type as that of the purchased vehicle" is positive and very significant. This implies that most households are trying to maintain the vehicle body type composition of their fleets when they decide to make a vehicle replacement.

From the estimation results, I also find that the market value of the remaining household vehicle(s) of household's fleet is crucial -- remaining vehicles are the original vehicle(s) which were kept after the vehicle transaction. The coefficient is positive and very significant which may be interpreted as the remaining vehicle value is part of the household’s wealth. This finding is consistent with what is found in Ren(1995).

3.5 *Stated Preference Vehicle Transaction Model Specification*

Households' stated preference transaction data comes from the first-wave SP experiment in which each household was given vehicle addition and replacement choices and was asked to choose among the six given hypothetical alternative-fuel vehicles.[19]

3.5.1 *Description of the SP Multinomial Logit Model*

The "tree" structure for the SP multinomial logit transaction model is similar to the one for the RP model in Figure 1. The exceptions are that there are only 6 alternatives instead of 12 to choose from (at each of the lower-level branches) and households do not get to choose "do nothing". Note that the "tree" structure only serves to better specify the MNL model and no nested structure is imposed by this structure. The dependent variable is equal to zero if the hypothetical vehicle is not chosen and equal to one otherwise.

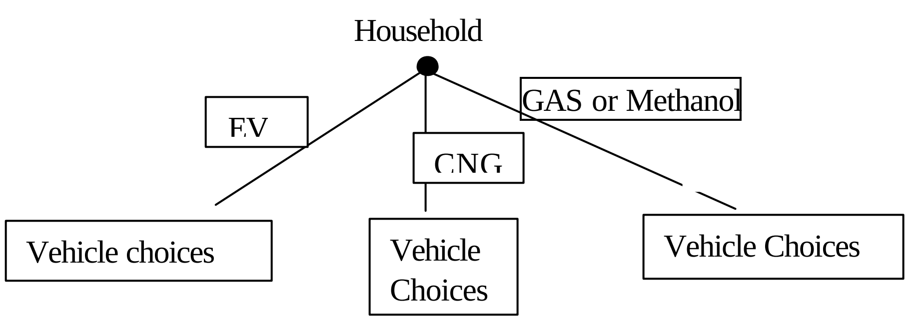
[19] The total number of unique alternatives is 120. The alternatives comes from the combinations of four fuel types (electricity, CNG, methanol and gas), five size classes(mini, subcompact, compact, midsize and large) and six body types (regular car, sports car, truck, van, station wagon, sports utility vehicle). Each household was given three vehicles with different fuel types and size classes and they were presented in either of two specified body types. Therefore, each household faced six hypothetical vehicles to choose from.

Based on the “tree” structure representing the SP vehicle transaction model, I apply the conditional multinomial logit model to the SP data. The SP MNL model estimation results is shown in Table 4 with estimated coefficients listed in column 2 of Table 4.

3.5.2 *Description of the SP Nested Logit Model*

Motivated by former studies such as Bunch and Bradley (1995), Brownstone and Train (1999), I impose three non-overlapping nests on the SP data: one consisting of electric vehicles, one consisting of compressed natural gasoline vehicles, and one consisting of gasoline and methanol vehicles. The nested structure can be described in Figure 5 below.

Figure 5. Description of the Nested Structure for the SP Transaction Data



The SP nested logit(NL) estimates are achieved by fully maximizing the log-likelihood function described in the theoretical framework section. The estimated coefficients are listed in column 3 of Table 5.

3.5.3 *Comparisons of the Estimation Results of MNL and NL Models*

Both the MNL and NL estimates are shown in Table 5. Column 2 and Column 3 of Table 5 give the estimated parameters and standard errors respectively for the multinomial logit and nested logit models with the same specification.

Table 5. Multiple-Vehicle Household SP Transaction Model Estimation Results

Explanatory Variables	MNL Coeff. (T-STATs)	NL Coeff. (T-STATs)
Cost Variables		
net capital cost /ln(household income/$1k) (in $1k)	-0.234 (-5.378)	-0.363 (-5.172)
average fuel cost/ln(household income/$1k) `(in cents/mile)	-0.240 (-3.860)	-0.316 (-3.581)
average fuel cost * addition dummy	0.131 (0.980)	0.124 (0.905)
Vehicle Performance Attributes		
average acceleration time (in seconds)	-0.047 (-1.711)	-0.081 (-1.932)
average top speed (in hundreds of miles/hour)	0.343 (1.709)	0.484 (1.609)
range of the purchasing vehicle (in hundreds of miles)	0.318 (6.553)	0.636 (5.051)
maximum range of holding vehicle(s) (in hundreds of miles)	0.263 (1.415)	0.488 (2.062)
squared maximum range	-0.047 (-1.508)	-0.080 (-2.136)
Vehicle Pollution Measurement		
tailpipe emission of the purchasing vehicle	-0.334 (-2.128)	-0.040 (-1.836)
Hypothetical-Vehicle-Body-Type Dummies		
small cars	-0.309 (-4.137)	-0.267 (-2.668)
sport cars	0.734 (1.596)	0.806 (1.639)
pickup trucks	-0.614 (-5.296)	-0.722 (-5.760)
Vans	-0.867 (-4.021)	-0.797 (-3.363)
sport-utility-vehicles	0.557 (1.188)	0.527 (1.057)
mini-sport-utility-vehicles	-1.011 (-1.134)	-0.777 (-0.820)
Specific Attributes of Alternative Vehicles		
electric vehicles	-0.128 (-0.634)	-0.345 (-0.613)
electric sport cars	-0.387 (-1.015)	-0.316 (-0.624)
electric pickup trucks	-0.261 (-1.119)	-0.298 (-1.184)
compressed natural gasoline vehicles	0.400 (2.990)	0.625 (1.236)
methanol vehicles	0.537 (4.512)	0.953 (4.805)
electric vehicle * having college education	0.379 (2.723)	0.807 (2.470)
addition constant * electric vehicles	0.675 (4.485)	0.669 (4.225)
refueling/recharging station availability	0.507 (3.613)	0.756 (3.035)
Household Portfolio-Choice Variables		
replacing-lower-value-vehicle dummy	0.333 (2.150)	0.348 (2.115)
replacing-same-body-type-vehicle dummy	1.148 (13.223)	1.188 (13.414)
oldest vehicle vintage	-0.017 (-1.590)	-0.017 (-1.556)
market value of the remaining holding vehicles (in $1k)	0.099 (6.238)	0.132 (6.347)
from no-car households to one-car households	-0.262 (-1.575)	-0.315 (-1.786)
from no-sport-car households to one-sport-car households	-0.035 (-0.069)	-0.031 (-0.057)
from no-van households to one-van households	0.207 (0.959)	0.135 (0.572)
from no-SUV households to one-SUV households	0.926 (1.838)	1.045 (1.943)
vans * household size >=3	0.902 (5.477)	1.045 (5.200)
sport cars * household size >=3	-0.892 (-1.430)	-1.188 (-1.876)
addition constant	-0.434 (-1.475)	-0.398 (-1.322)
addition constant * households with annual income <=$45k	0.264 (1.874)	0.308 (2.124)
addition constant * households with annual income >=$90k	-0.045 (-0.163)	-0.115 (-0.412)
Inclusive Values (in nested logit model)		
inclusive values for electric or CNG vehicles		0.459 (6.069)
inclusive values for Gasoline or Methanol vehicles		0.482 (6.094)

Note that: 1) the log-likelihood values for the multinomial logit and nested logit model are -3953.660 and –3938.135 respectively;

2) the number of observation is 26472 and the number of households is 1626;
3) the pseudo R2 is 0.1197 for the multinomial logit model and it is 0.3242 for the nested logit model;
4) the "base" vehicle size is "midsize/large" cars; the "base" fuel type is "gasoline".

Most variables in the SP models have similar definitions to the variables in the RP model. However, some variables are specific to the SP data. First of all, "electric vehicles", "compressed natural gasoline vehicles", and "methanol vehicles" are alternative-fuel-vehicle dummies. The MNL estimates in Table 4 show that coefficients of CNG and methanol are positive and significant while the electric vehicle coefficient is negative and not significant. The coefficient for methanol vehicles is positive but not significant in the nested logit model. It is difficult to explain these results because we do not have detailed attributes for the alternative-fuel vehicles such as dual fuel availability which may affect consumers' preferences for alternative-fuel vehicles. Adding in such attributes may be considered in future work. I interact the electric-vehicle constant with sport-car and pickup-truck dummies: both coefficients are negative but not significant. However, when I interact the electric-vehicle dummy with the household-member-having-college-education dummy, the interaction term is positive and significant. This is consistent with the finding in former studies such as Brownstone and Train(1999). Another important attribute for alternative-fuel vehicles is "refueling/recharging station availability" which defines the percentages of available refueling/recharging stations. This attribute is also significantly positive as expected.

a) MNL estimates versus NL estimates

The nested logit model has a better fit than the multinomial logit model. The absolute value of its log-likelihood is much smaller. The result of a log-likelihood ratio test also shows that the nested logit model is better. The likelihood ratio test statistic is equal to -2(-3953.660-(-3938.135)) which is 31.05. However, the critical value for chi-squared distribution with 2 degrees of freedom (which is equal to the number of restrictions) is only 5.99 which is much smaller than 31.05.

For the nested logit model, the estimated coefficient of the inclusive values for electrical and CNG vehicles is 0.459[20] while it is 0.482 for the gasoline and methanol vehicles. Therefore, we can see that the two coefficients are very close in magnitude and they are both very significant which shows that there is strong correlation among the unobservables for alternatives within the nests.

All the coefficient estimates have the same signs in both the NL model and the MNL model. However, most of the coefficients in the NL model are larger in magnitude, except that several body-type related attributes have smaller coefficients. This finding is consistent with the findings in Brownstone and Train (1999) where they find that the parameter estimates are usually larger in mixed logit models than standard logit models. The finding is also expected because nested logit models allow certain correlation structures, impose less

[20]I allow the coefficients of the inclusive values for the electrical and CNG vehicles to differ first and obtain separate estimates for them. However, the estimation results show that there is no significant

stringent restrictions, and the variance of the disturbance term is smaller. Therefore, the parameters in the nested logit are expected to be larger since the utility function is normalized by the variance of the disturbance term.

Some of the ratios of the estimated parameters (which are economically meaningful) are similar in the standard logit and nested logit. For example, the ratio of average top speed to net capital cost is 1.47 in the standard logit and 1.33 in the nested logit. However, some of the ratios are quite different in the two models. For example, the ratio of household-member-having-college-education-choosing-an-electric-vehicle to the net capital cost is 1.62 in the standard logit while it is 2.22 in the nested logit.

difference between the two estimates. Therefore, I constrain the two coefficients to be the same in the nested logit model since it does not affect the model estimation.

b) SP model estimates versus RP model estimates

Since the MNL and NL estimates are similar for the SP data, I will simply compare the SP MNL estimates against the previous RP MNL model estimates. From Table 4, we can tell that there are some big differences between the estimates of these two models. First, some of the attributes have different signs in the SP and RP models and they are both significant. For example, vehicle's tailpipe emission level is significantly negative in the SP model but significantly positive in the RP model.[21] Some other attributes also have different signs although they may not be significant in both or either of the models. For example, sports car enters the SP model insignificantly with a positive sign while it enters the RP model insignificantly with a negative sign. Second, the magnitude of the attributes can be very different in the SP and RP models even though the estimates have the same signs. For example, the coefficient for the average top speed is much larger in the RP model (1.550) than in the SP model (0.343). Third, the ratio of the parameters in the two models can be very different as well. For example, the ratio of pickup trucks to small cars is 1.99 in the SP model and is 0.83 in the RP model.

[21] Since the pollution variable is not well defined in the RP data, I will estimate this variable separately for the RP and SP data in the joint model and use the SP estimate in the forecasts.

However, most attributes, especially the vehicle body type variables, are considered to be poorly estimated in the SP model compared to the RP model because of the special design of the stated preference experiment in our survey. Therefore, I estimate many of the vehicle attributes separately in the joint estimation models and simply use their estimates for the RP data in the forecasts to correct possible SP biases. For the attributes that are specific to the SP data, I combine their estimates in the joint model and use them in the forecasting which is necessary in terms of predicting alternative-fuel vehicle market shares. Alternative-fuel dummies such as electric, CNG or methanol are examples of this kind of variable.

Chapter 4
Joint Revealed and Stated Preference Vehicle Transaction Model

4.1 *Model Specification*

From the discussion of the differences between the pure SP and pure RP estimates, we can see that the relative trade-offs can be very different even for the common attributes of the SP and RP data, for example, vehicle body-type variables. Therefore, how should we decide which variables should be "pooled" together (i.e. assigned same coefficients) and which variables should not? There are several factors which could affect this choice. One factor is that the estimates are quite different for the same variables in the RP and SP data. The other factor is that according to the characteristics of the RP and SP data, it is expected that either the RP or SP data provides better estimates for some variables. For example, the body type dummy variables are not well presented in the SP data because of the bias of the SP experimental design and it is expected that the body type estimates in the SP data are biased. Therefore, vehicle body type dummies are estimated separately for the RP and SP data and we can see from Table 6 that there are big differences between the estimates. Another factor is that statistically it is not possible to pool some attributes with quite different estimates (in terms of the magnitudes of the coefficients and the significance levels) in the RP and SP models. For example, I do not merge the variable, "replacing-same-body-type-vehicle dummy" for the RP and SP data.

Based on my experience and the findings in former studies such as Brownstone, Bunch and Train(1999)[22], most of the generic attributes , for example, net capital cost, range, average acceleration time etc. are "pooled" together after simply scaling the SP data. However, most body-type related variables are estimated separately for the SP and RP data. Thus, there are three types of attributes in the joint SP-RP models: one is the common generic attributes whose coefficients are the same for the SP and RP data; one is the other common attributes whose coefficients are estimated separately for the SP and RP data; and one is attributes which are specific to the SP or RP data.

The conditional multinomial logit estimation results for the joint SP and RP data are given in column 2 of Table 6. In order to relax the basic IIA assumption of the MNL model, I also apply the nested logit model to the joint SP and RP data set. The nested structure is specified the same as the one in Figure 2 where electric vehicles, CNG vehicles and gasoline or methanol vehicles are gathered into three different groups. The FIML nested logit estimation results for the joint SP-RP data are given in column 3 of Table 6.

[22]Brownstone, Bunch and Train(1999) mention that coefficients acting as "alternative specific constants" ("ASC's") might be much less likely to "pool" than coefficients associated with "generic attributes". The body-type coefficients are expected to behave similarly to "ASC's" besides the differences which could be caused by the different data generation processes of the SP and RP choices. Therefore, vehicle body type variables are estimated separately.

As specified in the theoretical framework section of chapter 3, a scale factor is introduced to account for the different variances for the random error components in the RP and SP data. The scale factor is estimated (in both the joint MNL and joint NL models) using a program written to search for the value which maximizes the log-likelihood function. It is expected that the SP data is more "noisy" than the RP data. Therefore, the magnitude of the scale factor is expected to be within the unit interval. It turns out that the scale factor is 0.70 for the joint MNL model, while it is 0.92 for the joint NL model. Therefore, the MNL scale factor is not only smaller than one but also smaller than the NL scale factor. This finding is similar to the results in Brownstone, Bunch and Train(1999) although the scale factor for their mixed logit model is greater than one.

Table 6. Multiple-Vehicle Household Joint RP/SP Transaction Model Estimation Results

Explanatory Variables	MNL Coeff. (T-STAT)	NL Coeff. (T-STAT)
Cost Variables (Joint RP/SP)		
net capital cost /ln(household income/$1k) (in $1k)	-0.351 (-9.069)	-0.370 (-8.927)
average fuel cost/ln(household income/$1k) (in cents/mile)	-0.342 (-4.196)	-0.318 (-3.814)
average fuel cost * non-transaction dummy	0.351 (2.745)	0.292 (2.357)
average fuel cost * addition dummy	0.285 (2.085)	0.226 (1.938)
Vehicle Performance Attributes		
average acceleration time (in seconds) (joint RP/SP)	-0.088 (-2.316)	-0.099 (-2.401)
average top speed (in hundreds of miles/hour) (joint RP/SP)	0.556 (2.036)	0.578 (1.957)
range of the purchasing vehicle (in hundreds of miles) (joint RP/SP)	0.471 (7.417)	0.611 (6.263)
maximum range of holding vehicle(s) (in hundreds of miles) for RP	2.954 (3.273)	3.180 (3.249)
maximum range of holding vehicle(s) (in hundreds of miles) for SP	0.260 (1.011)	0.433 (1.788)
squared maximum range for the RP data	-0.393 (-3.751)	-0.427 (-3.757)
squared maximum range for the SP data	-0.043 (-1.011)	-0.070 (-1.774)
Vehicle Pollution Measurement		
tailpipe emission of the purchasing vehicle for the RP data	0.302 (3.626)	0.323 (3.831)
tailpipe emission of the purchasing vehicle for the SP data	-0.356 (-1.594)	-0.367 (-1.601)
Purchasing-Vehicle-Body-Type Dummies		
small cars for the RP data	-0.259 (-2.084)	-0.255 (-2.049)
small cars for the SP data	-0.153 (-1.461)	-0.113 (-1.090)
sport cars for the RP data	-0.056 (-0.179)	-0.025 (-0.072)
sport cars for the SP data	1.177 (1.826)	0.947 (1.814)
pickup trucks for the RP data	-0.281 (-1.853)	-0.272 (-1.794)
pickup trucks for the SP data	-0.966 (-6.130)	-0.843 (-6.492)
vans for the RP data	-1.076 (-2.893)	-1.019 (-2.766)
vans for the SP data	0.237 (1.332)	0.186 (1.244)
sport-utility-vehicles for the RP data	0.591 (2.184)	0.628 (2.313)
sport-utility-vehicles for the SP data	0.874 (1.328)	0.640 (1.205)
mini-sport-utility-vehicles for the RP data	-1.430 (-1.871)	-1.262 (-1.641)
mini-sport-utility-vehicles for the SP data	-1.297 (-1.034)	-0.793 (-0.786)
Other Attributes of the Purchasing Vehicle (RP only)		
natural logarithm of number of makes in the same-size class	0.793 (10.611)	0.803 (10.655)
luxury vehicles	-0.249 (-1.662)	-0.191 (-1.247)
foreign vehicles	-0.371 (-3.010)	-0.398 (-3.204)
new vehicles	1.122 (6.053)	1.131 (6.089)
1-year-old vehicles	0.354 (1.762)	0.335 (1.673)
natural logarithm of vehicle vintage	-0.380 (-3.229)	-0.417 (-3.525)
Household Portfolio-Choice Variables		
replacing-lower-value-vehicle dummy for the RP data	0.139 (0.906)	0.163 (1.072)
replacing-lower-value-vehicle dummy for the SP data	0.658 (3.148)	0.483 (2.834)
replacing-same-body-type-vehicle dummy for the RP data	0.659 (5.383)	0.666 (5.451)
replacing-same-body-type-vehicle dummy for the SP data	1.383 (12.028)	1.102 (12.230)
oldest vehicle vintage (RP only)	-0.043 (-4.538)	-0.036 (-4.218)

market value of the remaining holding vehicles (in $1k) (joint RP/SP)	0.111 (7.136)	0.115 (7.853)
from no-car households to one-car households for the RP data	0.199 (1.524)	0.199 (1.486)
from no-car households to one-car households for the SP data	-0.351 (-1.548)	-0.319 (-1.713)
from no-sport-car households to one-sport-car households - RP	-0.439 (-1.306)	-0.440 (-1.309)
from no-sport-car households to one-sport-car households - SP	-0.078 (-0.112)	-0.040 (-0.069)
from no-van households to one-van households for the RP data	0.931 (2.528)	0.927 (2.520)
from no-van households to one-van households for the SP data	-1.401 (-0.226)	-1.106 (-9.818)
from no-SUV households to one-SUV households for RP data	0.043 (0.159)	0.044 (0.161)
from no-SUV households to one-SUV households for SP data	1.174 (1.661)	0.998 (1.744)
all domestic vehicle(s)	0.359 (3.153)	0.352 (3.087)
all imported vehicle(s)	0.479 (4.287)	0.483 (4.318)
Specific Attributes of Alternative Vehicles (SP only)		
electric vehicles	-0.100 (-0.354)	-0.280 (-0.532)
electric sport cars	-0.528 (-0.982)	-0.365 (-0.693)
electric pickup trucks	-0.346 (-1.056)	-0.305 (-1.130)
compressed natural gasoline vehicles	0.587 (3.129)	0.596 (1.257)
methanol vehicles	0.752 (4.483)	0.905 (4.644)
electric vehicle * having college education	0.535 (2.723)	0.753 (2.503)
addition constant * electric vehicles	0.953 (4.520)	0.716 (4.263)
refueling/recharging station availability	0.637 (3.204)	0.734 (2.961)
Transaction-Type Dummy Variables		
“inertia” i.e. non-transaction choice (RP only)	7.120 (14.101)	7.788 (12.952)
addition constant for the RP data	-0.071 (-0.247)	0.020 (0.078)
addition constant for the SP data	0.123 (0.303)	-0.307 (-1.081)
Demographic Characteristics and Changes		
addition*number of vehicles less than or equal to number of drivers (joint RP/SP)	1.808 (13.338)	1.811 (13.361)
addition * households with annual income<=$45k for RP data	-0.131 (-0.836)	-0.113 (-0.725)
addition * households with annual income<=$45k for SP data	0.304 (1.599)	0.270 (1.808)
addition * households with annual income>$90k for RP data	0.569 (2.724)	0.558 (2.674)
addition * households with annual income>$90k for SP data	0.015 (0.040)	-0.059 (-0.200)
addition or replacement* household demographic change(s) (joint RP/SP)	0.094 (0.905)	0.094 (0.917)
vans * household size >= 3 (joint RP/SP)	0.906 (5.089)	0.795 (4.970)
sport cars * household size >= 3 for the RP data	0.758 (2.377)	0.749 (2.350)
sport cars * household size >= 3 for the SP data	-1.208 (-1.387)	-1.209 (-1.787)
Inclusive Values (in the nested logit model)		
Inclusive values for electric and CNG vehicles (constrained to be equal)		0.525 (7.015)
Inclusive value for the gasoline or methanol vehicles (combined in single group)		0.544 (7.144)

Note that: 1) the log-likelihood values for the joint multinomial logit and nested logit models are -7043.851 and –7031.303 respectively;

2) the number of observations is 99930 and the number of households is 3852;
3) the pseudo R2 value is 0.423 for the joint multinomial logit model while it is 0.610 for the joint nested logit model;
4) the "base" vehicle class is "midsize/large" cars and the "base" fuel type is gasoline.

4.2 *Comparisons of the Estimation Results of Joint MNL and NL Models*

a) joint MNL estimates versus joint NL estimates

Similarly to the SP models, the nested logit model gives a better fit for the joint data -- the absolute value of the log-likelihood function is much smaller than that of the multinomial logit. The coefficients of the inclusive values for the three "nests" (about 0.5) are close to each other and are very significant, which reflects a strong correlation of the unobservables of alternatives within each nest.

Compared with the joint MNL model, all the estimated parameters in the joint nested logit have the same sign except for the addition dummies. However, the addition dummies are not significant in either of the models. The magnitude of the coefficients for some variables are larger in the nested logit than the standard logit. But the magnitude for most body-type related variables and fuel cost variables is somehow smaller in the nested logit than the standard logit. The significance levels are generally similar in both models for most variables although there are a few exceptions. For example, the CNG dummy variable is significant in the standard logit while insignificant in the nested logit. The ratios of the estimated coefficients are quite similar in both of the models. For example, the ratio of average acceleration time to net capital cost is 0.25 in the joint MNL model and it is 0.26 in the NL model. The ratios can be very different for some variables in the models since some variables even have different signs in the models as it is mentioned above.

b) Joint SP-RP Model versus Pure RP or SP Model

To be consistent, I simply compare the conditional multinomial logit estimates of the joint data, the pure RP data and the pure SP data against each other.

The common variables which are "pooled" together in the joint model all have common signs, and most of them have similar significance levels in both the joint and pure RP models (See Table 4 and Table 6.). However, there are a few exceptions. For example, average acceleration time and average top speed are both significant in the joint SP-RP model while they are insignificant in the pure RP model. The magnitude of the coefficients for most "pooled" common variables is smaller in the joint model than the pure RP model except for variables such as average acceleration time, natural logarithm of vehicle vintage, and the dummy of household with 3 or more members buying vans. The ratios of the coefficients for these variables in the two models can be very different. For example, the ratio of average top speed to net capital cost is much smaller (equal to 1.58) in the joint model than the RP model (equal to 4.35).

For the common variables which are estimated separately for the RP and SP data in the joint model, we can see from Table 6 that there are large differences between the corresponding SP data and RP data estimates. For example, the RP coefficient for van dummy is significantly negative while its coefficient is positive and insignificant for the SP data. As we expect, most body type related variables have very different estimates for the pure RP and pure SP data. Therefore, it is useful to distinguish their differences and use the corresponding RP estimates in the forecasting to correct for the specific body-type biases associated with the SP experiment design.

Compared to the pure SP estimates, most SP specific variables have larger coefficients in magnitude in the joint model than the pure SP model except for the electric vehicle dummy variable, which is insignificant in both of the models. However, the ratios of most of these variables are similar in both of the models. For example, the ratio of methanol to net capital cost is 2.14 in the joint model and 2.29 in the pure SP model. Moreover, all the “pooled” common variables and the SP specific variables have same signs in both of the models.

Chapter 5
SCENARIO FORECASTS

5.1 *New Vehicle Technology File for Year 1998*

As pointed out in Brownstone, Bunch and Train (1999), the easiest way to see the differences among the various models is to compare their forecasting results.

To compare the pure SP and the joint SP-RP MNL models, I first use the same vehicle technology file as the one used by Brownstone, Bunch and Train (1999). The detailed description of the vehicle choices is given in Appendix VI. This description is considered to reflect the new vehicle market and the available vehicle technology in year 1998[23] since it is based on expert predictions. However, there are limitations of the vehicle technology file. For example, information about the battery replacement costs for electric vehicle is not included. The fuel costs are calculated by dividing the fuel price by vehicle fuel efficiency assuming that gasoline price is $1.20/gallon and electricity costs 6 cents/KWH. All the prices are in 1995 dollars. Appendix VI assumes that electric vehicle technology would be available for mini cars, subcompact cars, sports cars, compact pickup trucks and minivans in 1998, and that CNG and methanol vehicle technologies would also be available for limited types of vehicles.

[23]We choose year 1998 because it was originally the year that California would begin to mandate the sale of certain number of alternative-fuel vehicles, although in our survey households were actually asked to choose from hypothetical AFVs designed for year 1998. The new vehicle technology file is provided by the California Energy Commission.

As to the predicted performance of AFVs in year 1998, electric vehicles are predicted to have very limited range, 75 miles for every type of vehicle, and to have top speed that are generally lower than those of similar type gasoline vehicles. In reality, the ranges of the electric vehicles offered by auto manufacturers are around 75 miles or even less. According to the statistics on the web site of Electric Vehicle Association of the Americas, the range of EV1 offered by General Motors is about 70 miles to 90 miles while the range for S-10 Electric offered by Chevrolet is only about 40 to 60 miles. The performance of methanol vehicles is predicted to be comparable with that of similar type gasoline vehicles.

5.2. *Forecasting Results Based on the New Vehicle Technology File*

a) The Pure SP and the Joint SP-RP MNL Model Forecasts

Taking the new vehicle technology file as given, I first forecast the demand for each type of new vehicles in year 1998 using the estimated parameters in the pure SP MNL model. Then I conduct another run using the estimates from the joint SP-RP MNL model. The results are given in Table 7. Note that the forecasts are not weighted. The forecasts are for the 2226 households who have complete RP data. Therefore they do not represent overall vehicle market forecasts.

In the joint SP and RP model, some common variables of the SP and RP data are pooled and estimated together, which is one of the strengths of the joint estimation. The common coefficients are used in the forecast. While for the other common variables, they are estimated separately in the joint model because most of them are associated with vehicle body types which are estimated with biases in the SP data. Therefore, I choose to use the RP estimates for these coefficients in the forecast[24].

[24] The Pollution variable is an exception to this case. Because we do not have a good measurement of vehicle pollution in the RP data, I choose to use the SP estimate of the pollution coefficient in the forecast.

Table 7. Predicted Market Shares of Pure SP versus Joint SP-RP MNL Model

Index	Body Type	Hilo	D/I	Fuel Type	SP MNL Model		Joint RP/SP MNL Model	
					Full Scenario	Gasoline Vehicles Only	Full Scenario	Gasoline Vehicles Only
1	Mini Car	low	I	electric	2.38		0.29	
2	Mini Car	high	I	gasoline	2.95	3.70	0.86	0.90
3	Subcompact	low	D	electric	2.16		0.33	
4	Subcompact	low	I	gasoline	5.09	6.38	9.82	10.24
5	Subcompact	high	I	gasoline	1.58	1.97	0.90	0.95
6	Compact Car	low	D	gasoline	5.51	6.90	14.92	15.57
7	Compact Car	high	I	gasoline	1.47	1.83	1.09	1.14
8	Compact Car	low	D	cng	1.98		0.18	
9	Midsize Car	low	D	gasoline	5.04	6.31	12.43	12.98
10	Midsize Car	low	D	methanol	3.45		1.77	
11	Midsize Car	high	I	gasoline	1.92	2.40	0.92	0.96
12	Midsize Car	high	D	gasoline	1.80	2.24	0.47	0.49
13	Midsize Car	high	D	methanol	1.34		0.10	
14	Large Car	low	D	cng	1.89		0.14	
15	Large Car	low	D	gasoline	4.69	5.86	5.91	6.17
16	Large Car	high	D	gasoline	1.18	1.47	0.33	0.34
17	Sport Car	low	D	electric	3.58		0.08	
18	Sport Car	low	I	gasoline	9.00	11.27	3.98	4.16
19	Sport Car	high	D	gasoline	2.49	3.11	0.10	0.11
20	Sport Car	high	I	gasoline	1.34	1.66	0.20	0.21
21	Compact Truck	low	D	electric	0.80		0.11	
22	Compact Truck	low	I	gasoline	2.09	2.62	5.55	5.78
23	Compact Truck	high	D	gasoline	1.14	1.43	1.05	1.10
24	Standard Truck	low	D	gasoline	1.20	1.50	5.20	5.43
25	Standard Truck	low	D	methanol	0.85		0.46	
26	Standard Truck	high	D	gasoline	0.93	1.16	0.84	0.87
27	Mini Van	low	I	electric	0.58		0.08	
28	Mini Van	low	D	gasoline	1.54	1.93	7.67	8.02
29	Mini Van	high	I	gasoline	1.07	1.34	0.60	0.63
30	Standard Van	low	D	gasoline	1.42	1.77	9.26	9.68
31	Standard Van	low	D	methanol	0.98		0.70	
32	Standard Van	high	D	gasoline	0.91	1.13	0.63	0.66
33	Compact SUV	low	D	gasoline	7.41	9.26	7.29	7.62
34	Compact SUV	high	I	gasoline	4.31	5.38	0.61	0.63
35	Standard SUV	low	D	gasoline	6.41	7.99	2.55	2.66
36	Standard SUV	high	D	gasoline	5.47	6.82	1.97	2.05
37	Mini SUV	low	I	gasoline	2.05	2.57	0.61	0.64

The “Gasoline Vehicle Only” column gives the forecasts using a scenario with all new gasoline vehicles in the full vehicle choice set. The “Full Scenario” column gives the forecasts using the full scenario described in Appendix VI. We can see that the pure SP model and the joint SP-RP model give quite different forecasts. The pure SP vehicle transaction model gives much higher market shares to alternative-fuel vehicles including electric vehicles, CNG and methanol vehicles while giving a lower market share to gasoline vehicles.

According to the forecast numbers in Table 7, the comparison of alternative-fuel vehicle market shares using the full set of alternative vehicles is given in Table 8.

Table 8. Alternative-Fuel Vehicle Market Shares in Percentages

Fuel Type	Pure SP MNL Model	Joint SP-RP MNL Model
Electric	9.50	1.29
CNG	3.87	0.47
Methanol	6.62	3.18
Gasoline	80.00	95.01

The pure SP MNL model gives a total market share of 20% for the non-gasoline vehicles while the joint SP-RP MNL model gives a non-gasoline share of 5%. The joint SP-RP vehicle transaction model provides more realistic forecast. However, it is unlikely that even 5% of the vehicles purchased in 1998 are alternative-fuel vehicles.

In addition to comparing the market share for alternative-fuel vehicles, I also make a comparison of the market shares for gasoline vehicles with different body types just using the new vehicle technology file. The comparison is given in Table 9.

Table 9. Forecasting Market Shares for Gasoline Vehicles with Different Body Types

Vehicle Body Type	Pure SP Model	Joint RP/SP Model
Small Cars	20.8	27.9
Midsize/Large Cars	18.3	21.2
Sport Cars	16.0	4.6
Trucks	6.7	13.3
Vans	6.2	19.4
SUVs	32.0	13.6

The sport utility vehicles account for more than 30% of the whole vehicle market according to the pure SP model forecasts which sounds unreasonably high. In contrast, the joint RP/SP model gives much lower market share forecasts for sport utility vehicles and sport cars. It seems that the joint RP/SP model provides better forecasts[25].

[25]In order to explore the relevance of choosing the coefficients for forecasting, I use the same set of coefficients in this joint model forecast except that I use the SP estimates for the coefficients of vehicle body types instead of their RP estimates. The forecasted market shares for different body type vehicles are far out of reality, with vans predicted to account for more than 40% of the vehicle market. Nevertheless, the forecasted market share for alternative-fuel vehicles dropped from about 5% to 4% under this scenario.

Because according to the sale statistics of new vehicles in 1997 reported by Transportation Energy Data Book (See Table 6-1, Edition 18, 1998), the market shares for small cars, midsize/large cars, sport cars, trucks, vans and sport-utility vehicles are 30.3, 25.0, 0.6, 17.4, 10.8 and 15.9 respectively. It is obvious that the pure SP model over predicts the market share for sport cars while underestimates the market share for sport utility vehicle dramatically. However, the joint SP-RP model forecasts are much more close to the reality.

b) the Pure SP MNL and NL Forecasts

In order to compare the conditional multinomial logit pure SP model and the nested logit pure SP model, I conduct another run forecasts using the estimated parameters from the pure SP NL model. Comparisons of the forecasts are based on the full set of new vehicle choices and are given in Table 10.

From Table 10 we can see that the pure SP nested logit model predicts much larger market shares for alternative-fuel vehicles compared to the predictions of the corresponding standard logit. For example, the predicted market shares for electrical vehicles and CNG vehicles are 14.46 and 10.11 percent respectively based on the nested logit while the standard logit only predicts 9.50 and 3.87 percent respectively. This finding is consistent with the results in former studies. Brownstone and Train (1999) find that the mixed logit gives larger shares for alternative-fuel vehicles and their argument is that the mixed logit specified with the alternative-fuel error components incorporates the fact that some people actually prefer alternative-fuel vehicles regardless of the measured attributes. Similar argument can be made with the nested logit model here to support our forecasts.

Table 10. Predicted Market Shares of Pure SP MNL versus Pure SP-RP NL Model

Index	Body Type	Hilo	D/I	Fuel Type	**Pure SP MNL Model**	**Pure SP NL Model**
1	Mini Car	low	I	electric	2.38	4.73
2	Mini Car	high	I	gasoline	2.95	1.50
3	Subcompact	low	D	electric	2.16	3.89
4	Subcompact	low	I	gasoline	5.09	4.34
5	Subcompact	high	I	gasoline	1.58	0.46
6	Compact Car	low	D	gasoline	5.51	5.02
7	Compact Car	high	I	gasoline	1.47	0.37
8	Compact Car	low	D	cng	1.98	5.74
9	Midsize Car	low	D	gasoline	5.04	3.39
10	Midsize Car	low	D	methanol	3.45	3.17
11	Midsize Car	high	I	gasoline	1.92	0.49
12	Midsize Car	high	D	gasoline	1.80	0.41
13	Midsize Car	high	D	methanol	1.34	0.49
14	Large Car	low	D	cng	1.89	4.37
15	Large Car	low	D	gasoline	4.69	2.97
16	Large Car	high	D	gasoline	1.18	0.20
17	Sport Car	low	D	electric	3.58	3.33
18	Sport Car	low	I	gasoline	9.00	6.24
19	Sport Car	high	D	gasoline	2.49	0.51
20	Sport Car	high	I	gasoline	1.34	0.14
21	Compact Truck	low	D	electric	0.80	1.37
22	Compact Truck	low	I	gasoline	2.09	2.78
23	Compact Truck	high	D	gasoline	1.14	1.03
24	Standard Truck	low	D	gasoline	1.20	1.36
25	Standard Truck	low	D	methanol	0.85	1.35
26	Standard Truck	high	D	gasoline	0.93	0.78
27	Mini Van	low	I	electric	0.58	1.14
28	Mini Van	low	D	gasoline	1.54	1.87
29	Mini Van	high	I	gasoline	1.07	0.96
30	Standard Van	low	D	gasoline	1.42	2.09
31	Standard Van	low	D	methanol	0.98	1.95
32	Standard Van	high	D	gasoline	0.91	0.99
33	Compact SUV	low	D	gasoline	7.41	8.70
34	Compact SUV	high	I	gasoline	4.31	3.35
35	Standard SUV	low	D	gasoline	6.41	8.58
36	Standard SUV	high	D	gasoline	5.47	6.52
37	Mini SUV	low	I	gasoline	2.05	3.41

In order to further compare the differences in the nested logit and standard logit forecasts, I use two models (similar to the ones in Brownstone and Train(1999)) to study the different substitution patterns in the three models implied by the nested logit and standard logit.

The base scenario consists of five gasoline vehicles (one mini car, one subcompact car, one compact car, one midsize car and one large car) and one mini electric car. The two models are considered under two different scenarios:

(i) a subcompact electric car is introduced to the base scenario (scenario I);

(ii) one midsize methanol car is introduced to scenario I (scenario II);

I use the same representative household in these models. The representative households have 3 or more members, its annual income is $50k, one household member has college education, and it owns a 6-year-old subcompact car and an 11-year-old minivan already. Table 11 gives the changes in the probabilities (not realistic forecasts) that result from the four scenarios i.e. from the introduction of new vehicles. The table shows how different substitution patterns are generated by nested logit and standard logit models, which are essentially caused by the differences in the basic assumptions of these two models.

Table 11. Changes in Predicted Percentages for Households under Different Scenarios

Percentage Changes in Predicted Probabilities	Scenario I (Base scenario. Add a subcompact EV.)	
	Logit	Nested Logit
Mini Gas Car	-12	-7
Subcompact Gas Car	-12	-7
Compact Gas Car	-12	-8
Midsize Gas Car	-12	-7
Large Gas Car	-12	-7
Mini EV	-12	-34
Subcompact EV	12	14
Total EV Probabilities	26	30

Percentage Changes in Predicted Probabilities	Scenario II (Base: scenario I. Add one midsize methanol car)	
	Logit	Nested Logit
Mini Gas Car	-17	-18
Subcompact Gas Car	-17	-18
Compact Gas Car	-17	-17
Midsize Gas Car	-17	-18
Large Gas Car	-17	-17
Mini EV	-17	-7
Subcompact EV	-17	-8
Methanol Car	17	15

In part 1 of Table 11, a new subcompact electric car is introduced to the base scenario with five gas cars and a mini electric car. Nested logit predicts that about 34% households who chose to buy the mini electric car will switch to buy the subcompact electric car, while only 7% or 8% of the households who chose to buy the five gas cars will choose the subcompact electric car. In contrast, because of the IIA property, standard logit predicts that the subcompact electric car will draw proportionately (about 12%) from the households who chose to purchase the five gas cars and the mini electric car. Therefore, we can see that the nested logit provides more realistic forecasts of the substitution patterns. This difference comes from the specification of the "nest" structure in the nested logit model which reflects the fact that some households actually prefer electric vehicles regardless of the measured attributes.

In part 2 of Table 11, a new methanol car is introduced to Scenario I with five gas cars and two electric cars. The nested logit predicts that more households (about 17% or 18%) who chose the five gas cars will choose to buy the new methanol car (which is considered to have similar performance as the gas cars). But only 7% or 8% of the households who bought electric cars are predicted to switch to the purchase of the new methanol car.

Therefore, we can see that the nested logit model provides more realistic substitution patterns in general when a new alternative-fuel vehicle is introduced. The standard logit model provides unrealistic forecasts because it always predicts that the new vehicle introduced will draw proportionately from each of the existing vehicles. However, the nested logit incorporates the fact that some households prefer vehicles with certain fuel types. It does this by specifying the correlation structures among the error components of vehicle choices with different fuel types.

5.3 *Forecast Results Based on Both the New and Used Vehicle Technologies*

The major disadvantage of the current forecasting scenario file is that it does not have any description of used vehicles and it is not a very good description of the total vehicle market. Therefore, I try to incorporate the used vehicle technologies into the original vehicle technology file. I include in the scenario file another 689 used vehicle alternatives with vintages ranging from 1-year-old to 22-year-old. Thus the new scenario file consists of 726 vehicle alternatives in total (37 new vehicles and 689 used vehicles) and it is a much better representation of the whole vehicle market. The new forecasting results using the present scenario file shows that the predicted market share for alternative-fuel vehicles (including electric vehicles, CNG and Methanol vehicles) drops from the original 5% to 2.09%. Therefore, we can see that it is very useful to include used vehicle technology in the vehicle scenario file. The forecasts are all based on the estimates from the joint SP-RP multinomial logit model.

Table 12. Predicted Market Share for Alternative-Fuel Vehicles Using the New and Used Vehicle Technology File

Fuel Type	Using New Vehicle Technology File	Using New and Used Vehicle Technology File
Electric	1.29	0.49
CNG	0.47	0.13
Methanol	3.18	1.47
Gasoline (New)	95.0	58.79
Gasoline (Used)		39.12
Gasoline (Total)	95.0	97.91

Using the scenario file consisting of both new and used vehicles, the predicted market shares for new and used gasoline vehicles are 58.79 and 39.12 percent respectively according to the numbers shown in Table 12. The forecasts are more realistic because the used vehicle market is taken into consideration. By using the same scenario file, I am also able to come up with the predicted market shares for new and used gasoline vehicles with different body types. The statistics are shown in Table 13 below.

Table 13. Predicted Market Shares for New and Used Vehicles with Different Body Types Using the New and Used Vehicle Technology File

Different Body Type Vehicles	New	Used
Small Cars	17.6	11.4
Midsize/Large	12.8	9.3
Sport Cars	2.6	3.3
Trucks	8.0	5.5
Vans	11.1	4.6
Sport Utility Vehicles	8.0	5.8

To make realistic forecasts, one should use such a vehicle scenario file contain both new and used vehicle choices; one should also use a set of households that represents the demographics at that time; one should use a more reasonable estimation model such as nested logit models or mixed logit models as well. To show the usefulness of the nested logit models, one can conduct the similar forecast task as I explained in section 5.2. One can include the used vehicle choices in the scenario file besides the new ones, use the representative set of households and conduct forecast under the different scenarios. It will be interesting to see how the substitution patterns will matter to new and used vehicles when new alternative-fuel vehicles are introduced. To perform such a task, a nest structure on new and used vehicles are expected. However, due to the limited number of new vehicle choices in the sampled choice set, I did not find sufficient evidence to support this kind nested structure. Therefore, future work may consider developing more complicated nested logit models for such forecast purpose.

CHAPTER 6

CONCLUSIONS

Assuming that households make their vehicle type choices together with their vehicle transaction choices, I develop and estimate separate revealed and stated preference vehicle transaction choice models in which households' vehicle transaction choices are jointly estimated with households' vehicle type choices. In addition, a joint SP-RP model is estimated utilizing both the stated and revealed preference data. Since it is assumed that the variances of the random error components for the SP and RP are different, a scale factor is incorporated in the joint estimation to take account of the difference. The scale factor is estimated to be less than 1 which means that the SP data is more "noisy" than the RP data.

I test the validity of IIA property of multinomial logit model and I develop the nested logit model in which same-fuel-type vehicle-choices are grouped together. The nested structure specification reflects the fact that some households prefer vehicles with certain fuel types regardless of their other attributes. I apply the same nested structure to both the pure SP data and the joint SP-RP data. According to the log-likelihood values of the models, nested logit models fit the data better than the standard logit models. The estimated coefficients are similar in the nested logit to those in the standard logit. The coefficients of the inclusive values for different "nests" are within the unit interval and they are all very significant. This reflects that there is a strong correlation among the error components of vehicle choices within each "nest".

Most variables in the models have the expected signs and are significant. The variables include cost, vehicle performance, purchased vehicle body type dummies and other attributes, households' "portfolio choice" variables, transaction type dummies, households' demographic characteristics and some SP/RP specific variables. The results show that some of these variables are very important. For example, some household portfolio choice variables such as the "market value of the remaining holding vehicles" and some household demographic factors such as "number of vehicles less than or equal to the number of drivers in the household" play important roles, and the results are quite robust.

In order to further compare the differences between the nested logit and the standard logit, I conducted short-run demand forecasts using the pure SP data and the same new vehicle technology file. I find that the nested logit gives higher market shares to alternative-fuel vehicles, compared to gasoline vehicles, because nested logit incorporates the fact that some households prefer AFVs no matter their other attributes.

Different forecasts are also carried out using the estimates from the joint SP-RP multinomial logit model and the new vehicle technology file. By comparing the forecasts with the ones based on the pure SP multinomial logit model, I show that the joint SP-RP model provides better forecasts. Because the joint MNL model predicts about 5% market share for EV, CNG and Methanol vehicles for year 1998 while the prediction based on the pure SP MNL model is about 20% -- which is far from reality.

In order to make the forecasts more realistic, I then incorporate the used vehicle technologies into the forecasting scenario file. The predicted market share for alternative-fuel vehicles drop from the original 5% to about 2%. Therefore, I find that taking account of the used vehicle technologies is very important in forecast.

REFERENCES

Beggs, Steven D., Scott Cardell, and J. Hausman (1981), "Assessing the Potential Demand for Electric Cars", *Journal of Economrtrics* 16, pp. 1-19.

Brownstone David, D. Bunch, and K. Train, 1999, " Joint Mixed Logit Models of Stated and Revealed Preferences for Alternative-Fuel Vehicles",

Brownstone David and K. Train, 1999, "Forecasting New Product Penetration with Flexible Substitution Patterns," *Journal of Econometrics,* V89, pp.109-129.

Calfee, John E. (1985), "Estimating the Demand for Electric Automobiles Using Fully Disaggregated Probabilistic Choice Analysis", *Transportation Research B* 19B.4, pp.187-301.

David S. Bunch et al. (1992), "Demand for Clean-Fuel Vehicles in California: A Discrete-Choice Stated Preference Pilot Project," IN: Transportation research. Part A, Policy and practice, Vol. 27A, No. 3.

Golob, TF, Bunch, DS, and Brownstone, D. (1997), "A vehicle use forecasting model based on revealed and stated vehicle type choice and utilisation data", *Journal of Transportation Economics and Policy*, V31, N1.

Greene H. William(1997), "Econometric Analysis", Third Edition, Princeton-Hall International, Inc.

Hausman, J., and D. McFadden (1984), "Specification Tests for the Multinomial Logit Model." Econometrica, 46.

Hensher D.A., N.C. Smith,F.W, Milthorpe and P.O. Barnard (1992), "Dimensions of Automobile Demand", Elsevier Science Publishers B.V.

Jong, Gerard de (1996), "A Disaggregate Model System of Vehicle Holding Duration, Type Choice and Use", Transportation Research, Volume 30B, No. 4, pp. 263-276.

MaFadden, D.(1974), "Conditional Logit Analysis of Qualitative Choice Behavior", Frontiers in Econometrics, P.Zarembka, Ed, Academic Press, New York, pp.105-142.

Morikawa, Taka (1994), "Estimation of Discrete Choice Models from Serially Correlated RP and SP data", *Transportation*, Volume 21, pp. 153-165.

Ren, Weiping, (1995), “A Vehicle Transaction Choice Model”, Dissertation, Department of Economics, University of California, Irvine.

Sheng, Hongyan (1997), “A Practical Investigation of Incorporating Stated and Revealed Preference Information”, research paper, Department of Economics, University of California, Irvine.

Small, Kenneth A. and Clifford Winston (1998), “The Demand for Transportation: Models and Applications”, Working Paper, Institute of Transportation Studies, University of California, Irvine.

Smith, N.C., D.A. Hensher, and N. Wrigley (1991), “A Discrete Choice Sequence Model: Method and An Illustrative Application to Automobile Transactions”, *International Journal of Transport Economics*, Vol. XVIII, No. 2.

Davis, Stacy C., (1998), “Transportation Energy Data Book: Edition 18”, Oak Ridge National Laboratory, U.S. Department of Energy, ORNL-6941.

SRI (1978), *Electric Vehicle News.*

Train, Kenneth (1980b), “The Potential Market for Non-gasoline-Powered Automobiles”, *Journal of Transportation Research* 14A, pp. 405-414.

Train, Kenneth (1986), “Qualitative Choice Analysis, Theory, Econometrics and an Application to Automobile Demand”, The MIT Press.

Appendix I: Preliminary Survey Questions about Next Intended Vehicle Purchase

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)

Apeendix II. Wave-1 CATI-1 Stated Preference Vehicle Transaction Choice Questions

Which vehicle did you choose? Was it the:

1. Vehicle ‘A’ body type 1
2. Vehicle ‘A’ body type 2
3. Vehicle ‘B’ body type 1
4. Vehicle ‘B’ body type 2
5. Vehicle ‘C’ body type 1
6. Vehicle ‘C’ body type 2
7. Don’t Know
8. Refused

Did your household choose this vehicle as a replacement vehicle or as an additional vehicle?

1. replacement
2. Additional
3. Don’t know
4. Refused

You have crossed off the vehicle you would replace. In your household did the Chosen Vehicle replace your

0. Not Applicable
1. Household Vehicle 1
2. Household Vehicle 2
3. Household Vehicle 3
4. Household Vehicle 4
5. Household Vehicle 5
6. Household Vehicle 6

Appendix III: Statistics on Household Annual Income Distributions (in $1,000)

Sample Category	Number of Obs.	Mean	Std. Dev.	Min.	Max.
the whole wave2 CATI1 Sample	2857	53.63	34.09	7.5	150
the total households with RP data included in my study	2226	52.71	32.76	7.5	150
the 665 households with vehicle transactions	665	58.34	33.75	7.5	150
the 1561 households without any vehicle change	1561	50.31	32.04	7.5	150

Appendix IV: Summary Statistics for the Explanatory Variables in the RP Models

Variables	Mean	Std. Dev.	Min	Max
Cost Variables				
net capital cost /ln(household income/$1k) (in $1k)	2.687	3.697	-8.771	45.482
average fuel cost / ln(household income/$1k) (in cents/mile)	6.079	1.127	0	12.342
average fuel cost * non-transaction dummy	0.047	0.284	0	5.271
average fuel cost * addition dummy	0.606	0.860	0	5.085
Vehicle Performance Attributes				
average acceleration time (in seconds)	3.569	0.464	0	6.219
average top speed (in hundreds of miles/hour)	1.044	0.084	0	1.55
range of the purchasing vehicle (in hundreds of miles)	4.150	0.897	0	5.75
maximum range of holding vehicle(s) (in hundreds of miles)	4.506	0.409	0	5.75
squared maximum range	20.473	3.627	0	33.063
Vehicle Pollution Measurement				
tailpipe emission of the purchasing vehicle	1.510	0.871	0	6.113
Purchasing-Vehicle-Body-Type Dummies				
small cars	0.244	0.429	0	1
sport cars	0.106	0.308	0	1
pickup trucks	0.144	0.351	0	1
vans	0.130	0.337	0	1
sport-utility-vehicles	0.154	0.361	0	1
mini-sport-utility-vehicles	0.022	0.146	0	1
Other Attributes of the Purchasing Vehicle				
natural logarithm of number of makes in the same-size class	1.852	0.979	0	3.638
luxury vehicles	0.389	0.488	0	1
foreign vehicles	0.457	0.498	0	1
new vehicles	0.242	0.429	0	1
1-year-old vehicles	0.123	0.328	0	1
natural logarithm of vehicle vintage	1.182	1.144	0	3.091
Household Portfolio-Choice Variables				

replacing-lower-value-vehicle dummy	0.417	0.493	0	1
replacing-same-body-type-vehicle dummy	0.188	0.391	0	1
oldest vehicle vintage	10.792	6.210	0	22
market value of the remaining holding vehicles (in $1k)	7.527	6.650	0	67.458
from no-car households to one-car households	0.086	0.280	0	1
from no-sport-car households to one-sport-car households	0.078	0.268	0	1
from no-van households to one-van households	0.098	0.297	0	1
from no-SUV households to one-SUV households	0.132	0.339	0	1
all domestic vehicle(s)	0.294	0.456	0	1
all imported vehicle(s)	0.231	0.421	0	1
Transaction-Type Dummy Variables				
“inertia” i.e. non-transaction choice	0.030	0.171	0	1
addition constant	0.364	0.481	0	1
Demographic Characteristics and Changes				
addition * number of vehicles less than or equal to number of drivers	0.088	0.283	0	1
addition * households with annual income<=$45k	0.164	0.370	0	1
addition * households with annual income>$90k	0.033	0.179	0	1
addition or replacement* household demographic change(s)	0.332	0.471	0	1
vans * household size >= 3	0.040	0.197	0	1
sport cars * household size >= 3	0.034	0.181	0	1

Appendix V: Summary Statistics for the Explanatory Variables in the SP Models

Variables	Mean	Std. Dev.	Min	Max
Cost Variables				
net capital cost /ln(household income/$1k) (in $1k)	3.100	2.239	-8.526	17.371
average fuel cost / ln(household income/$1k) (in cents/mile)	1.415	0.542	0.200	4.621
average fuel cost * addition dummy	0.539	0.766	0	4.621
Vehicle Performance Attributes				
average acceleration time (in seconds)	3.919	0.838	2.265	6
average top speed (in hundreds of miles/hour)	1.016	0.140	0.55	1.465
range of the purchasing vehicle (in hundreds of miles)	2.373	0.934	0.5	4
maximum range of holding vehicle(s) (in hundreds of miles)	4.083	0.861	0.5	5.7
squared maximum range	17.416	5.611	0	1
Vehicle Pollution Measurement				
tailpipe emission of the purchasing vehicle	0.359	0.301	0	1
Hypothetical-Vehicle-Body-Type Dummies				
small cars	0.269	0.444	0	1
sport cars	0.024	0.153	0	1
pickup trucks	0.196	0.397	0	1
vans	0.257	0.437	0	1
sport-utility-vehicles	0.035	0.183	0	1
mini-sport-utility-vehicles	0.003	0.054	0	1
Specific Attributes of Alternative Vehicles				
electric vehicles	0.254	0.435	0	1
electric sport cars	0.008	0.090	0	1
electric pickup trucks	0.049	0.216	0	1
compressed natural gasoline vehicles	0.248	0.432	0	1
methanol vehicles	0.251	0.434	0	1
electric vehicle * having college education	0.118	0.322	0	1
addition constant * electric vehicles	0.094	0.291	0	1
refueling/recharging station availability	0.494	0.351	0.1	1

Household Portfolio-Choice Variables				
replacing-lower-value-vehicle dummy	0.566	0.496	0	1
replacing-same-body-type-vehicle dummy	0.277	0.448	0	1
oldest vehicle vintage	7.048	6.018	0	22
market value of the remaining holding vehicles (in $1k)	9.189	8.082	0	68.921
from no-car households to one-car households	0.072	0.258	0	1
from no-sport-car households to one-sport-car households	0.017	0.128	0	1
from no-van households to one-van households	0.218	0.413	0	1
from no-SUV households to one-SUV households	0.030	0.170	0	1
vans * household size >=3	0.070	0.255	0	1
sport cars * household size >=3	0.004	0.067	0	1
addition constant	0.369	0.482	0	1
addition constant * households with annual income <=$45k	0.174	0.379	0	1
addition constant * households with annual income >=$90k	0.023	0.151	0	1

Appendix VI: Year 1998 Vehicle Description of the Current Forecast Scenario File

No	Class	Hilo	D/I	Body Type	Fuel	No. of Models	Price ($)	Fuel Cost	Station Availability
1	1	low	I	Mini Car	electric	1	17518	1.33	0.1
2	1	high	I	Mini Car	gasoline	4	19986	4.90	1.0
3	2	low	D	Subcompact	electric	1	19562	1.47	0.1
4	2	low	I	Subcompact	gasoline	20	13524	4.53	1.0
5	2	high	I	Subcompact	gasoline	17	31415	5.48	1.0
6	3	low	D	Compact Car	gasoline	20	14814	4.63	1.0
7	3	high	I	Compact Car	gasoline	30	35689	5.77	1.0
8	3	low	D	Compact Car	cng	1	21057	4.14	0.1
9	4	low	D	Midsize Car	gasoline	23	18695	5.73	1.0
10	4	low	D	Midsize Car	methanol	3	18985	5.38	0.7
11	4	high	I	Midsize Car	gasoline	17	36127	6.39	1.0
12	4	high	D	Midsize Car	gasoline	6	39382	5.92	1.0
13	4	high	D	Midsize Car	methanol	1	36504	6.00	0.7
14	5	low	D	Large Car	cng	1	26561	5.24	0.1
15	5	low	D	Large Car	gasoline	11	23083	5.88	1.0
16	5	high	D	Large Car	gasoline	9	50000	6.70	1.0
17	6	low	D	Sport Car	electric	1	26414	1.96	0.1
18	6	low	I	Sport Car	gasoline	12	19641	5.41	1.0
19	6	high	D	Sport Car	gasoline	2	38999	6.78	1.0
20	6	high	I	Sport Car	gasoline	30	53309	6.18	1.0
21	7	low	D	Compact Truck	electric	1	21669	2.05	0.1
22	7	low	I	Compact Truck	gasoline	11	15000	5.61	1.0
23	7	high	D	Compact Truck	gasoline	4	20401	7.23	1.0
24	8	low	D	Standard Truck	gasoline	15	18839	8.34	1.0
25	8	low	D	Standard Truck	methanol	1	19129	7.84	0.7
26	8	high	D	Standard Truck	gasoline	4	24175	7.96	1.0
27	9	low	I	Mini Van	electric	1	29785	2.36	0.1
28	9	low	D	Mini Van	gasoline	17	22000	6.49	1.0
29	9	high	I	Mini Van	gasoline	3	27533	6.96	1.0
30	10	low	D	Standard Van	gasoline	18	19741	8.40	1.0
31	10	low	D	Standard Van	methanol	1	20031	7.89	0.7
32	10	high	D	Standard Van	gasoline	2	24820	9.43	1.0
33	11	low	D	Compact SUV	gasoline	11	23100	7.16	1.0
34	11	high	I	Compact SUV	gasoline	3	30000	8.28	1.0
35	12	low	D	Standard SUV	gasoline	3	25651	9.08	1.0
36	12	high	D	Standard SUV	gasoline	4	27629	9.41	1.0
37	13	low	I	Mini SUV	gasoline	7	15223	5.28	1.0

Appendix VI (continue)

No	Class	Hilo	D/I	Body Type	Fuel	Range (miles)	Accel. Time	Top Speed	Pollut. level

							(seconds)	(miles)	
1	1	low	I	Mini Car	electric	75	3.9	92	0.0
2	1	high	I	Mini Car	gasoline	281	3.0	120	0.8
3	2	low	D	Subcompact	electric	75	3.6	98	0.0
4	2	low	I	Subcompact	gasoline	344	3.5	110	0.8
5	2	high	I	Subcompact	gasoline	277	2.9	122	0.8
6	3	low	D	Compact Car	gasoline	389	3.4	111	0.8
7	3	high	I	Compact Car	gasoline	306	3.0	121	0.8
8	3	low	D	Compact Car	cng	147	3.7	105	0.4
9	4	low	D	Midsize Car	gasoline	345	3.3	114	0.8
10	4	low	D	Midsize Car	methanol	215	3.2	115	0.6
11	4	high	I	Midsize Car	gasoline	310	2.6	133	0.8
12	4	high	D	Midsize Car	gasoline	334	2.8	126	0.8
13	4	high	D	Midsize Car	methanol	193	2.5	134	0.6
14	5	low	D	Large Car	cng	150	3.4	112	0.4
15	5	low	D	Large Car	gasoline	398	3.1	117	0.8
16	5	high	D	Large Car	gasoline	390	2.9	122	0.8
17	6	low	D	Sport Car	electric	75	3.0	109	0.0
18	6	low	I	Sport Car	gasoline	344	3.0	121	0.8
19	6	high	D	Sport Car	gasoline	247	2.0	154	0.8
20	6	high	I	Sport Car	gasoline	270	2.6	132	0.8
21	7	low	D	Compact Truck	electric	75	3.2	84	0.0
22	7	low	I	Compact Truck	gasoline	362	3.2	94	0.8
23	7	high	D	Compact Truck	gasoline	299	3.1	95	0.8
24	8	low	D	Standard Truck	gasoline	324	3.2	94	0.8
25	8	low	D	Standard Truck	methanol	202	3.1	95	0.6
26	8	high	D	Standard Truck	gasoline	336	3.8	84	0.9
27	9	low	I	Mini Van	electric	75	3.3	82	0.0
28	9	low	D	Mini Van	gasoline	370	3.3	91	0.8
29	9	high	I	Mini Van	gasoline	354	3.4	90	0.8
30	10	low	D	Standard Van	gasoline	357	3.2	94	0.8
31	10	low	D	Standard Van	methanol	223	3.2	95	0.6
32	10	high	D	Standard Van	gasoline	326	3.4	90	0.8
33	11	low	D	Compact SUV	gasoline	338	3.2	91	0.8
34	11	high	I	Compact SUV	gasoline	299	3.0	95	0.8
35	12	low	D	Standard SUV	gasoline	396	3.4	91	0.9
36	12	high	D	Standard SUV	gasoline	385	3.4	90	0.9
37	13	low	I	Mini SUV	gasoline	261	3.7	86	0.8

Note that the operating fuel costs (in cents/mile) are calculated assuming that gasoline price is $1.20/gallon and electricity costs 6 cents/KWH, and all prices are in 1995 dollars.