Title
Forecasting Electric Vehicle Ownership and Use in the California South Coast Air Basin

Authors
Golob, Thomas F.
Brownstone, David
Bunch, David S.
et al.

Publication Date
1996-08-01

Forecasting Electric Vehicle Ownership and Use in the California South Coast Air Basin

UCI-ITS-WP-96-9

Thomas F. Golob [1]
David Brownstone [2]
David S. Bunch [3]
Ryuichi Kitamura [4]

[1] Institute of Transportation Studies, University of California, Irvine

[2] Department of Economics and Institute of Transportation Studies, University of California, Irvine

[3] Graduate School of Management, University of California, Davis

[4] Department of Transportation Engineering, Kyoto University, Japan

August 1996

Institute of Transportation Studies
University of California, Irvine
Irvine, CA 92697-3600, U.S.A.
http://www.its.uci.edu

Prepared for the Southern California Edison Company

TABLE OF CONTENTS

1. OVERVIEW

1.1 Objectives and Scope

This research deals with demand for automobiles and light-duty and medium-duty trucks. Planners concerned with energy consumption, air quality and the provision of transportation facilities must have dependable forecasts of vehicle ownership and use from both the residential (personal-use vehicle) sectors and the fleet (commercial and governmental) sectors. As long as vehicles evolved slowly, it was possible to base such forecasts on extrapolations of observed demand. However, in an era of increasing environmental awareness, mandated in part by the Clean Air Act Amendments (US EPA, 1990), government agencies are now concerned with promoting clean-fuel vehicles; vehicle manufacturers are faced with designing and marketing clean-fuel vehicles; and suppliers of fuels other than gasoline must plan infrastructure and pricing policies.

In California, and potentially also in a number of Northeast States, stringent vehicle emission standards have been adopted or proposed and specific zero-emissions and ultra-low-emissions vehicle mandates are in place. The California Air Resources Board (CARB) requires that new cars sold in the state emit 80 percent less hydrocarbons by the year 2000, and 50 to 75 percent less carbon monoxide and nitrogen oxide. CARB has also mandated the production and sale of zero-emission (presumably electric) vehicles, beginning with 2 percent of annual car sales in 1998 and increasing to 10 percent in 2003. Elsewhere in the United States, clean-air and fuel-management legislation (U.S. DOE, 1994) specifically targets fleets as markets for clean-fuel vehicles. Research is needed to establish the extent to which there is demand for clean-fuel vehicles. In reaction to this need, the Southern California Edison Company and the California Energy Commission is sponsoring a project to develop a dynamic demand forecasting model for clean-fuel vehicles in California. In this paper we briefly describe the forecasting system being developed and summarize some preliminary results.

Clean-fuel vehicles are potentially different from conventional gasoline or diesel vehicles in terms of many attributes that can be of prime importance to consumers. Such attributes include, but are not limited to: range between refueling, overnight recharging requirements (electric vehicles), the potential availability of at-home refueling (compressed natural gas vehicles), the limited availability of refueling or recharging stations, vehicle performance levels, and cargo carrying capacity, as well as substantial differences in capital and operating costs.

This research builds upon previous efforts to provide quantitative estimates of demand for electric and alternative fuel vehicles. These estimates are useful for evaluating incentive polices, vehicle design and marketing strategies, and fuel demand management. It is not possible to discuss all of these precursor studies here, but, on the residential vehicles side, important studies are those of Beggs and Cardell (1980), Beggs, Cardell and Hausman (1981), Hensher (1982), Calfee (1985), Greene (1989; 1990), and Train (1980). On the fleet side, Berg, *et al.* (1984), Easton (1991), Hill (1987), and Macro (1992) provide pioneering studies.

The model system is designed to forecast demand for vehicles and also fuel usage for each type of vehicle in each of a number of geographical regions. To determine the impact of electric vehicle recharging on the electric transmission and distribution system, the system forecasts recharge demand for electric vehicles by time of day in each of approximately forty districts that correspond to distribution planning areas. Currently, peak electricity demand in California occurs during summer afternoons, and minimum demands occur between midnight and 6:00 A.M. Therefore, electric vehicle recharging will be much cheaper and less polluting if it takes place during late night hours when electricity is generated by hydroelectric and other clean baseline plants. The current version of the system produces charging profiles under the assumption that consumers plug in their vehicles in an unconstrained fashion; the data source was a distribution of plug-in times from a vehicle trials program. Future versions of the system will use behavioral models to examine the extent to which consumers are willing to recharge electric vehicles off-peak at lower rates.

Forecasts are produced for various vehicle 'classes': All conventional-fuel and clean-fuel vehicle types that are anticipated to be available have been included. Makes and models of vehicles are grouped into relatively homogeneous classes with similar attributes, such as emission levels. The model system uses 14 residential vehicle body-type-and-size classes (7 car classes and 7 light truck classes) and 5 fuel technology types. Vehicles are further classified according to 10 model-year vintage groupings. The fleet demand sub-model also contains a medium-duty truck class and a small bus (shuttle bus) class in addition to all of these light-duty truck and car classes.

Since we are primarily interested in forecasting the demand for new types of vehicles, the model must be able to forecast the technology adoption process. This requirement rules out the classic static vehicle demand models, such as Train (1986). Our system produces a separate forecast for each period, with each period's forecast depending on all the previous forecasts. The current system does not account for vehicle demand from state and federal government rental car fleets: this will be added in future versions.

1.2 Basic Structure Of the forecasting system

The forecasting system is comprised of two main subsystems, the residential (personal-vehicle) demand subsystem and the fleet demand subsystem.

The forecasting system starts from a baseline database of households and commercial fleets, and then simulates a sequence of vehicle transactions at six-month intervals so that vehicle stocks are dynamically determined. Results are reported annually. The data for the baseline year, 1993, are derived from large-scale surveys of household vehicle holdings, and from a large-scale survey of fleets, augmented by vehicle registrations data. The forecasting method is similar to Hensher (1992), in which the household population is represented by a relatively small number of "synthetic" households. The present use of a large sample of actual households and fleets instead of a synthetic sample requires more computation, but the results should be more accurate.

Both the residential and fleet demand subsystems are based on transactions models. These models predict whether a vehicle transaction will occur during the current period and what type of transaction it will be. The inputs to the models are the current characteristics of the household (or fleet) and the current vehicle inventory and utilization. Since vehicle type decisions are discrete, the models can only provide probabilities that a particular household or firm will choose a particular type of vehicle. Forecasting a particular choice from these models requires simulating an actual choice, which introduces some random noise into the forecasting process. Fortunately, the effect of this randomness disappears when forecasts for individual households or fleets are aggregated to predict market demand. The predicted changes in vehicle holdings and utilization are then combined with initial holdings to forecast vehicle stocks for the next period (Brownstone, *et al.*, 1994).

1.3 Data

1.3.1 Household Survey

Since we are concerned with the demand for a new product that does not yet exist, we asked respondents to make choices among hypothetical vehicles. These "stated preference" questions (Louviere, 1988) have been successfully used in a pilot study of consumer preferences for alternative fuel vehicles (Bunch, at al., 1993; Golob, at al. 1993). This pilot study, sponsored by the California Energy Commission, confirmed that information about attribute trade-offs gained through our "stated preference" method are consistent with results of previous studies of actual vehicle purchase behavior (e.g., Train, 1980, 1986; Hensher, 1992).

Stated preference questionnaires require that respondents receive different hypothetical vehicles according to a pre-specified experimental design. The questionnaires also contain enough background information so that respondents can fairly evaluate the hypothetical vehicles. In addition to stated preference questions, we also ask extensive questions about respondents' existing vehicle stock and utilization. The remainder of this section gives more detail about the three main data sets used to calibrate our models.

The first wave of our personal vehicle panel survey was carried out in June and July, 1993. The sample was identified using pure random digit dialing and was geographically stratified into 79 areas covering most of the urbanized area of California. A total of 7,387 households completed the initial computer-aided telephone interview (CATI). This initial CATI interview collected information on: household structure, vehicle inventory, housing characteristics, basic employment and commuting for all adults, and stated intentions for the next vehicle transaction.

The data from the initial CATI interview were used to produce a customized mail-out questionnaire for each sampled household. This questionnaire asked more detailed questions about each household member's commuting and vehicle usage, including information about sharing vehicles in multiple-vehicle and multiple-driver households. The mail-out questionnaire also contained two "stated preference" (SP) choice experiment tasks for each household. Each of these tasks described three hypothetical vehicles, from which the households were asked to choose their preferred vehicle. These hypothetical vehicles included both clean-fuel and gasoline vehicles, and the body types and prices were customized to be similar (but not identical) to the household's description of their next intended vehicle purchase. Households then indicated which of their current vehicles they would replace with the preferred vehicle, or if the preferred vehicle would be purchased as an additional vehicle.

After the households received the mail-out questionnaires, they were again contacted for a final CATI interview. This interview collected all the responses to the mail-out questions. Additional questions about the household's attitudes towards clean-fuel vehicles were also included in this interview. A total of 4747 households completed all phases of the survey.

1.3.2 Fleet Site Survey

The first task in surveying commercial and local government (city, county and regional) fleet operators was to establish a comprehensive list of fleets from which a survey sample could be drawn. This also established a 'fleet universe.' Many small to medium size fleet operators are not currently registered in fleet databases available from fleet managers' associations, governmental agencies, or commercial market research firms. Moreover, these databases are not generally up to date on the number and type of vehicles operated in a given fleet. Consequently, a comprehensive list of potential fleets was obtained from the 26.5 million records of the California Department of Motor Vehicles registration file.

A rule-based algorithm was developed to match and combine all vehicle registrations with a high probability of being from the same company or individual at the same site, taking into account differences in registrations due to abbreviations and spelling. Most clean-air mandates target fleet sites with ten or more vehicles, so all potential sites with five or more registrations were investigated because of the likelihood that registration sites would still be fragmented into two or more components based on unresolved differences in names or addresses. Since substantial numbers of households own or lease five or more vehicles, and many households even own ten or more vehicles, a knowledge-based system using rules and predicate logic for conflict resolution was developed to separate households from businesses. A sample was then drawn from the identified registration sites, and survey results were used to factor the total list of registration sites in order to estimate the universe of commercial and local government fleet sites.

The survey of 2,100 fleet sites was conducted as a combined CATI and mail-back questionnaire. The CATI portion of the survey established the fleet inventory and business functions, and gathered data on multi-site fleet operations. In the customized mail-back questionnaire, fleet operators provided detailed operation and acquisition data on up to two selected types of vehicles currently in their fleets. In the mail-out SP tasks, the operators chose future fleets of the selected types from among hypothetical conventional-fuel and alternative-fuel vehicles, and they allocated the chosen vehicles to the tasks typically performed by the fleet. There were also questions concerning organizational decision making and opinions about alternative-fuel vehicles.

1.4 Forecasting System Components

Figure 1 is a schematic representation of the model system. The system has three types of components:

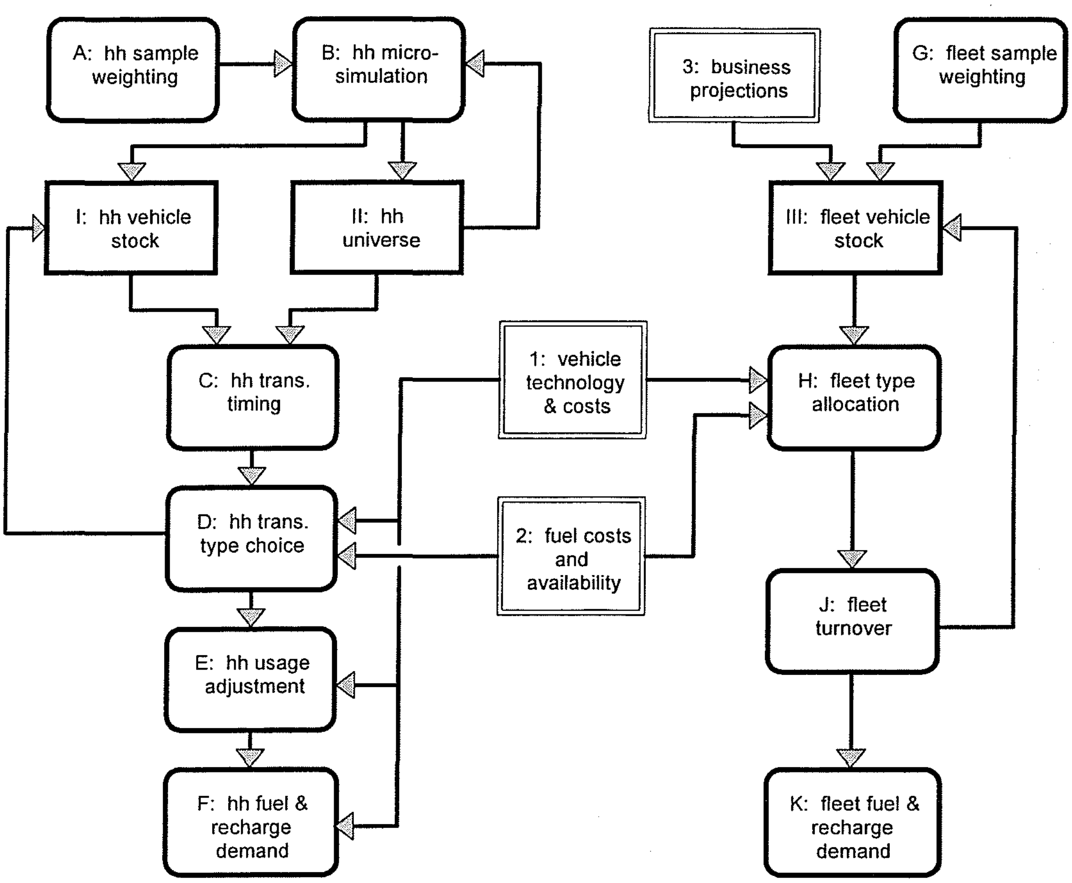
> *Exogenous datasets* that drive the forecasts are depicted as doubly-outlined boxes which are labeled 1 through 3.
>
> The principal *endogenous datasets* are depicted as heavy-outlined boxes which are labeled with the Roman numerals I through III.
>
> The component *models* are depicted in boxes with rounded corners and are labeled with the letters A through K (skipping I).

1.4.1 Exogenous Inputs

The key inputs to the residential vehicle forecasting subsystem are vehicle technology, and fuel costs and availability. Vehicle technology (Box 1 in Figure 1) includes numerical values for both historical and future vehicle attributes, including fuel type, refueling or recharging range, price, operating costs, vehicle tailpipe emissions, payload, and performance. Although it is relatively easy to forecast these attributes two to three years ahead, it is very difficult to predict the state of new technology ten or more years ahead. Forecasts from the model system crucially depend on future vehicle technology, and users of the model system will need to continually update this information as time progresses. Since the model produces forecasts for each year, it is also important to forecast when new technology vehicles will be introduced. Finally, the model system assumes that manufacturers are willing to provide as many vehicles as demanded at the forecast vehicle price.

Fuel costs and availability (Box 2 in Figure 1) is another exogenous input to the model system. Although fuel costs are typically very difficult to forecast, we only need accurate forecasts of relative fuel prices. The prices of three of the fuels considered in our model -- gasoline, compressed natural gas, and electricity -- have tended to move together with the price of crude oil during the past decade. However, if crude oil prices start to rise substantially, then the off-peak electricity price may diverge from recent patterns since in California off-peak electricity is primarily generated by hydroelectric power. Fuel infrastructure describes the availability of alternative clean fuels. For compressed natural gas and methanol this is expressed as the ratio of the number of service stations relative to gasoline.

Figure 1.1: Schematic Diagram of Forecasting System



Many proposed incentives (such as, sales tax and vehicle registration fee subsidies) simply lower the capital and/or operating costs of these vehicles, so the effects of these incentives can be modeled by changing the appropriate cost variables in the vehicle technology and fuel cost files. Other proposed incentives, such as free parking, solo driver access to high-occupancy vehicle (carpool) lanes, or extended vehicle warranties, cannot currently be captured in the vehicle technology or fuel technology inputs. The forecasting system is being expanded in 1995 so that both the residential and fleet demand subsystems will be sensitive to such incentives.

1.4.2 Sample Weighting

The 7387 survey households must be weighted to accurately represent the target population. We first created sampling weights that just accounted for the geographic stratification and the differential number of household telephone lines. These weights were then adjusted using statistical matching to the 1993 U.S. Census Current Population Survey (CPS) so that the weighted sample matched the CPS joint distributions of household composition, age, and income. Finally, these weights were further adjusted using a binomial logit model to account for non-random selection from the original 7387 survey households down to the forecasting sample. Sample weighting is the subject of *Chapter 5*.

1.4.3 Household Microsimulation

Model B in Figure 1 is a suite of dynamic competing-risks hazard models which age each household, and simulate births, deaths, divorces, children leaving home, etc. Once the new household structure is determined, other models in Box B determine the household's income and employment status. The models produce an updated Household Universe File (Dataset II) which is used as the starting point for aging the household in the next period; this cycling is depicted by the feedback from Dataset II to Box B in Figure 1. The household microsimulation models are mostly calibrated from the Panel Study of Income Dynamics (Hill, 1992) because the personal vehicle survey does not track households over a sufficiently long time period to be used as a calibration source. The household microsimulation model is documented *Chapter 6* and in Kazimi (1995) and Kazimi and Brownstone (1995).

1.4.4 Transactions Timing

Model C in Figure 2 takes the updated household and current (aged) vehicle holdings as inputs. It then decides whether or not a vehicle transaction takes place during this period. The period length is set at six months, in order to limit the number of transactions per period to one, but model system outputs are given annually. A vehicle transaction is defined to include: disposing of an existing vehicle, replacing an existing vehicle with another one, or adding a new vehicle to the household's fleet. The transactions timing model is documented in *Chapter 7*.

1.4.5 *Transactions Type Choice*

If the simulation from Transactions Timing Model B predicts that a vehicle transaction has taken place, the Transaction Type Choice Model in Box D determines exactly what type of transaction takes place. The household's vehicle holdings are updated accordingly, and these are used as inputs to the vehicle utilization model in Box E as well as starting values for the next period's forecast (the feedback loop from Box D to Dataset I). The model outputs for each year accumulate the probabilities of all actions to the total numbers of vehicles owned or leased by type and vintage. For new vehicles, this represents market penetration. The transactions type choice model is documented in <u>Chapter 8</u> and in Ren (1995) and Ren, *et al.* (1995).

1.4.6 *Vehicle Usage Adjustment*

A utilization model, Box E, then takes the updated vehicle holdings and household structure and predicts changes in the annual vehicle miles traveled (VMT) for each household vehicle. The model, described in Golob, Bunch and Brownstone (1995), is estimated on combined revealed preference (RP) and stated preference (SP) data from the household survey. The RP data involves reported usage levels for existing household vehicles, and the SP data involves responses to questions concerning how chosen hypothetical future vehicles would be used by various household members. Structural equation models are used to capture VMT and driver allocation for each household vehicle as a function of vehicle age, type, operating cost, range, and household characteristics. These models predict changes in VMT due to vehicle aging and driving aging effects, even if households make no vehicle transactions and all household characteristics are unchanged. Forecasts of VMT are generated by calculating expected usage at the beginning and end of each period and applying the percent changes in expected levels to the observed VMT base level, thus preserving sample heterogeneity. The usage model is documented in <u>Chapter 9</u> and in Golob, Bunch and Brownstone (1995).

1.4.7 *Fleet Sample Weighting*

The fleet survey weighting is described in <u>Chapter 10</u>. The DMV file was processed using rule-based algorithms to assign each vehicle to the household, business and local government, rental, and state and federal government sector. We then matched all of the business fleet vehicles by address to get the distribution of fleet sizes. Finally, we computed the distribution of vehicle type and vintage for all vehicles in fleets with size greater than or equal ten. The fleet weights were then set so that our weighted fleet sample matched these distributions.

1.4.8 Fleet Type Allocation Model

The fleet allocation model is based on a multinomial logit choice frequencies model estimated using the SP vehicle allocation tasks in the fleet survey. This model is documented in <u>*Chapter 11*</u> and in Golob, *et al.*, (1995).

1.4.9 Fleet Turnover

Finally, the fleet turnover and fuel demand system components are described in <u>*Chapter 12.*</u>

1.4.10 Acknowledgments and References.

This report closes with acknowledgments (Chapter 13), and a comprehensive listing of all references (Chapter 14).

2. BASE CASE FORECASTS

2.1 Introduction

A major deliverable of the ITS project is a base case forecasting scenario, and also the associated forecasts. It is not meaningful to discuss market penetration of electric vehicles (or any other types of vehicles) without rigorously stating the all underlying assumptions regarding future market conditions and competitors. The base case scenario ("BCS") provides a set of assumptions for all relevant market characteristics that are required by the current version of the forecasting system. Assumptions fall into two major categories: availability and characteristics of future vehicle technologies, and future characteristics of the refueling infrastructure. It should be clear that forecasting results critically depend on assumptions made for these items.

Data sets containing this information reside in two user-controlled input files: (1) vehicle technology, and (2) fuels forecast. The vehicle technology file establishes: (1) the market availability of various gasoline and alternative-fuel vehicle types in the years 1976 to 2010, and (2) forecasted vehicle characteristics that affect consumer demand for vehicles (e.g., purchase price, body type, range, acceleration, etc.). The fuels forecast file contains forecasts of fuel prices and service station availability for four fuels (gasoline, methanol, natural gas, and electricity) during the period 1994 to 2010. The file formats are given in Chapter 3, sections 3 and 4, respectively. The numerical values required for the base case scenario are quite detailed, and appear in Appendix C. The assumptions in the BCS were developed in early 1995 through collaboration and consultation with staff of the California Energy Commission (CEC), and also SCE. They will be revised sometime in the future as part the ITS team's ongoing research program. We reiterate that results depend on the assumptions contained in these files, and should not be discussed out of context. The ITS forecasting system allows users to change the assumptions and re-run the model.

In the following sections we first review the BCS assumptions regarding market availability of alternative fuel vehicles during the period 1994 to 2010, and give general background information regarding technology assumptions. We then summarize a variety of forecasting system results related to future vehicle penetration and usage in the Southern California Edison service area. The results will be given in an aggregated form that represents the major populated areas in Southern California (except San Diego). More detailed results may be obtained for individual SCE districts using GIS technology developed by SCE, but this type of analysis is beyond the scope of the ITS final report.

2.2 Market Availability of Alternative Fuel Vehicles

Two critical determining factors of future market penetration of alternative fuel vehicles (AFV's) are: (1) the types of vehicles that will be available, and (2) the timing of vehicle introduction. Two major defining vehicle characteristics are (1) vehicle class (which relate to body type and size), and (2) fuel type (fuel or fuels that the vehicle runs on). The categories used by the current system appear in Tables 2.1 and 2.2.

Table 2.1: Vehicle Classes.

Type	Size
1. Car	Mini
2. Car	Subcompact
3. Car	Compact
4. Car	Intermediate
5. Car	Large
6. Car	Luxury
7. Car	Sports (or, "Sports car")
8. Pickup	Compact
9. Pickup	Standard
10. Van	Compact (or, "Minivan")
11. Van	Standard
12. Sport utility vehicle	Small
13. Sport utility vehicle	Large
14. Sport utility vehicle	Mini
15. Truck	(10-14 KGVW)*
16. Bus	Shuttle (or, "Shuttlebus")*

*Used in the commercial fleet model, but not the personal vehicle model.

Table 2.2: Fuel Types.

1. Gasoline
2. Methanol ("flex fuel")
3. Dedicated Compressed Natural Gas (CNG)
4. Dual Fuel CNG ("bi-fuel")
5. Electric (dedicated)

The BCS vehicle technology file contains historical data on gasoline vehicles during the period 1976 to 1994. Gasoline vehicles have been available since 1976 for all vehicle types but two: gasoline minivans and mini-sport utility vehicles have only been available since 1979.

The forecasting system begins its microsimulation calculations in the year 1994, which is the first year that AFV's might potentially be available. The BCS assumes that nineteen types of AFV's will eventually be available in the market, giving a total of 33 vehicle types (AFV plus gasoline). It assumes that no alternative fuel vehicles will be widely available in the market until 1997. See Table 2.3. Nine of the nineteen AFV's appear in 1997, with the remainder appearing in 1998.

Table 2.3: BCS Market Availability of Alternative Fuel Vehicles.

Fuel type	Class	Size	1997	1998+
Methanol	Car	Intermediate		
	Car	Large		
	Car	Luxury		
	Pickup	Standard		
	Van	Standard		
Ded. CNG	Car	Subcompact		
	Car	Compact		
	Car	Intermediate		
	Car	Large		
	Pickup	Standard		
	Van	Compact		
	Van	Standard		
Dual CNG	Pickup	Standard		
Electric	Car	Mini		
	Car	Subcompact		
	Car	Compact		
	Car	Sports		
	Pickup	Compact		
	Van	Compact		

2.3 Vehicle Attributes

To predict vehicle choices, the models require information regarding certain attributes that have value to customers. For a listing of the attributes required for each vehicle, see Table 2.4. The format for the file allows data for both the personal vehicle model and commercial fleet model to be stored in the same location.

Table 2.4: Vehicle Attributes.

1. Vehicle Class Code	Integer, 1 to 16 (see Table 2.1)
2. Vehicle Fuel Type Code	Integer, 1 to 5 (see Table 2.2)
3. Vintage	Year (1976 to 2010)
4. Purchase Price (New)	Dollars
5. Fuel economy	MPG (for non-EV's) or, Miles per Kilowatt Hour
6. Acceleration Time (0 to 30 MPH)	seconds
7. Top Speed	MPH
8. Number of Models	
9. Service Station Refuel Time	Minutes
10. Refuel Time Two (= Home refuel time for personal electric vehicles, or slow-fill on-site refuel time for CNG fleet vehicles.)	Hours
11. Refuel Time Three (= On-site refuel time for electric fleet vehicles, or fast-fill on-site refuel time for CNG fleet vehicles.)	Minutes
12. Emissions Index	Fraction from 0 to 1. 1 = 1994 gasoline vehicle.
13. Range on Full Tank (Full Charge)	Miles
14. Luggage Space Index	Fraction from 0 to 1. 1 = comparable gasoline vehicle
15. Market Availability	1 = Vehicle is available, 0 = Not available.

Appendix C (which is quite lengthy), contains tables of vehicle attributes for the various vehicle types enumerated in Table 2.1. These tables are highly detailed, and require much time to digest. To simplify issues somewhat, we present some excerpts in section 2.6 below, along with more background and discussion.

2.4 Fuel Forecasts

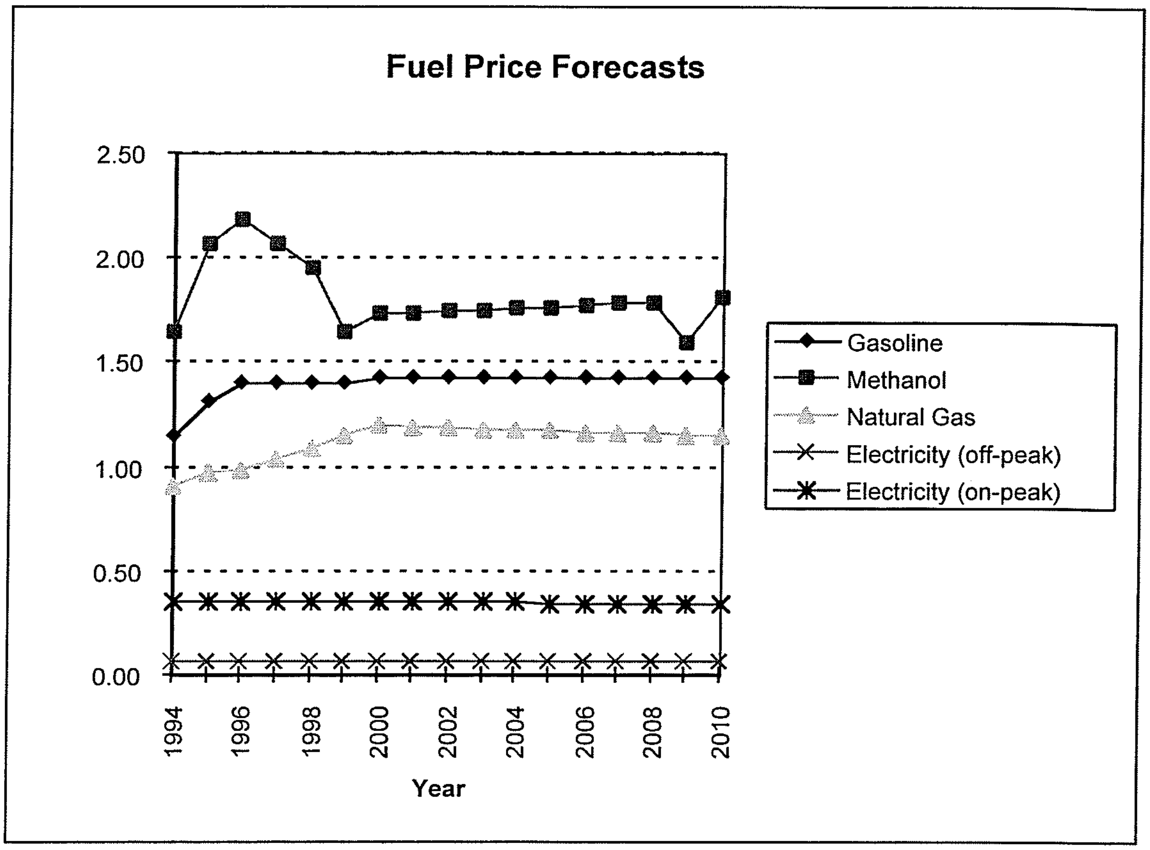
In addition to details on vehicles, the model requires information about future fuel prices and infrastructure. Decisions to purchase vehicles depend on fuel operating costs, which in turn depend on both vehicle fuel efficiency and the cost of fuel at the pump (or at the electrical outlet). In addition, purchase decisions depend on the availability of refueling stations. After vehicles have been purchased, the amount they are driven (annual vehicle miles traveled) depends on fuel operating cost. For a list of fuels forecast attributes required by the forecasting model, see Table 2.5.

The fuels prices used in the Base Case Scenario are the most recent ones available from the California Energy Commission (CEC). The electricity prices in California vary widely within the state: the values used here are those for the Los Angeles area. With regard to refueling station availability, the BCS assumes that the ratio of alternative fuel stations to gasoline stations is 0.1 for all years. A table of the detailed numerical values appears in Appendix B, but future fuel price trends are summarized in Figure 2.1.

Table 2.5: Fuels Forecast Attributes.

1. Year	1994 to 2010
2. Fuel Code	Integer, 1 to 4 (see below).
3. Fuel Cost One	Dollars per gallon at service station (for non-electric) or, Dollars per KWH (off-peak for electric)
4. Fuel Cost Two	On-peak for electric, Dollars per KWH
5. Fuel Availability Index (No. of stations relative to gasoline.)	Ratio from 0 to 1. 1 = gasoline service stations
Attribute 2. Fuel Codes	
1 - gasoline 2 - methanol 3 - compressed natural gas (CNG) 4 - electricity	

Figure 2.1



2.5 Selected Vehicle Technology Assumptions

When examining the forecasting results, it is important to remember that the results are predicated on the assumptions contained in the vehicle technology and fuels forecast input files. Changing the values in these files will produce different results. However, time and space limitations require us to select one specific Base Case Scenario and present results for that single case. Gaining a full understanding of the forecasting model and its behavior will require additional testing and re-running of simulations to see how the results are affected by changes in the attribute values such as purchase price, vehicle range, fuel availability, and so forth.

To make the discussion more manageable, we present in Tables 2.6-2.9 excerpts of attribute assumptions for two years: 1998 and 2005. These two years are of special interest because: (1) 1998 is the year of the first major CARB mandate, and (2) 2005 represents a ten-year time horizon relative to today's date. In addition to the details provided in the tables, we note the added assumptions: (1) refueling times for gasoline, methanol, and natural gas vehicles at service stations are 7, 7, and 5 minutes, respectively, (2) home recharging for electric vehicles takes 3 hours starting in 1998 (but 8 hours prior to 1998), (3) the service station availability index is 1 for gasoline and 0.1 for other alternative fuels. With the one noted exception, these assumptions apply to entire period 1994 to 2010.

There are a some noteworthy features in these tables. Various technological improvements are expected to occur for different vehicle types. Due to relatively low fuel prices, natural gas vehicles are less expensive to operate than gasoline and methanol vehicles. High methanol fuel prices cause methanol vehicles to be more expensive to operate than other vehicles. Gasoline vehicles have substantially more range than other vehicle types. . Ranges for electric and methanol vehicles improve over time. Electric vehicles are assumed to have substantially higher capital costs than other vehicles, but with the gap narrowing as time progresses. Other AFV's have capital costs comparable to gasoline. Natural gas vehicles are quite clean. Gasoline and methanol vehicles are much less clean, but they get cleaner over time. (Electric vehicles, of course, have zero tailpipe emissions.) In this BCS, we have followed previous work in which battery replacement costs are included as part of vehicle operating cost for electric vehicles.

Table 2.6: Gasoline Vehicle Assumptions in Base Case Scenario

Class	Size	Year	Price	MPG	Acceleration	Top Speed	Relative Emission	Range	Operatin Cost
Car	Mini	1998	13354	32.64	3.5	112	0.80	400	4.30
Car	Mini	2005	14207	34.09	3.3	112	0.52	400	4.18
Car	Subcompact	1998	12582	29.69	3.7	116	0.80	400	4.73
Car	Subcompact	2005	13424	30.99	3.5	116	0.52	400	4.59
Car	Compact	1998	17260	24.68	3.5	120	0.80	400	5.69
Car	Compact	2005	18084	25.70	3.3	120	0.52	400	5.54
Car	Intermediate	1998	19390	23.22	3.3	115	0.80	400	6.05
Car	Intermediate	2005	20167	23.39	3.1	115	0.52	400	6.09
Car	Large	1998	21025	20.91	3.2	109	0.80	400	6.72
Car	Large	2005	21884	21.48	3.0	109	0.52	400	6.63
Car	Luxury	1998	37799	19.62	3.0	141	0.80	400	7.16
Car	Luxury	2005	38592	19.67	2.8	141	0.52	400	7.24
Car	Sports	1998	17696	22.67	2.8	131	0.80	400	6.2
Car	Sports	2005	18610	23.23	2.6	131	0.52	400	6.13
Pickup	Compact	1998	13894	21.29	3.7	103	0.80	400	6.6
Pickup	Compact	2005	14679	21.66	3.5	103	0.52	400	6.57
Pickup	Standard	1998	17658	15.08	4.3	101	1.00	400	9.32
Pickup	Standard	2005	18389	15.32	4.0	101	0.86	400	9.29
Van	Compact	1998	20380	19.81	3.9	106	1.00	400	7.09
Van	Compact	2005	21199	20.14	3.6	106	0.86	400	7.07
Van	Standard	1998	18036	14.92	4.6	101	1.00	400	9.42
Van	Standard	2005	18861	15.68	4.3	101	0.86	400	9.08
SUV*	Compact	1998	22157	18.56	3.9	100	1.00	400	7.57
SUV	Compact	2005	23007	19.15	3.6	100	0.86	400	7.44
SUV	Standard	1998	24070	13.82	3.7	104	1.00	400	10.2
SUV	Standard	2005	24909	14.68	3.5	104	0.86	400	9.7
SUV	Mini	1998	14874	26.17	3.9	100	1.00	400	5.37
SUV	Mini	2005	15939	27.88	3.6	100	0.86	400	5.11

*"SUV" = "Sport Utility Vehicle.

Table 2.7: Methanol Vehicle Assumptions in Base Case Scenario

Class	Size	Year	Price	MPG	Acceleration	Top Speed	Relative Emission	Range	Operatin Cost
Car	Intermediate	1998	19621	25	3.1	115	0.80	267	7.87
Car	Intermediate	2005	20498	25	2.9	115	0.52	269	7.05
Car	Large	1998	21296	22	3.0	109	0.80	261	8.74
Car	Large	2005	22270	23	2.8	109	0.52	268	7.68
Car	Luxury	1998	37853	21	2.9	141	0.80	264	9.31
Car	Luxury	2005	38669	21	2.7	141	0.52	265	8.39
Pickup	Standard	1998	17927	16	4.0	101	1.00	300	12.12
Pickup	Standard	2005	18775	16	3.7	101	0.86	300	10.76
Van	Standard	1998	18305	16	4.3	101	1.00	300	12.25
Van	Standard	2005	19247	17	4.0	101	0.86	300	10.52

Table 2.8: Natural Gas Vehicle in Base Case Scenario

Class	Size	Year	Price	MPG	Acceleration	Top Speed	Relative Emission	Range	Operatin Cost
Car	Subcompact	1998	14902	30	4.0	106	0.09	180	3.60
Car	Subcompact	2005	15858	32	3.7	106	0.09	180	3.70
Car	Compact	1998	19580	25	4.0	110	0.09	180	4.34
Car	Compact	2005	20518	26	3.7	110	0.09	180	4.46
Car	Intermediate	1998	21710	24	3.7	105	0.09	180	4.61
Car	Intermediate	2005	22601	24	3.5	105	0.09	180	4.90
Car	Large	1998	23140	21	3.7	99	0.31	180	5.12
Car	Large	2005	24335	22	3.5	99	0.31	180	5.34
Pickup	Standard	1998	20516	15	4.9	91	0.31	180	7.10
Pickup	Standard	2005	21918	16	4.6	91	0.31	180	7.49
Van	Compact	1998	23266	21	4.2	96	0.31	180	5.29
Van	Compact	2005	24200	21	3.9	96	0.31	180	5.58
Van	Standard	1998	20898	15	5.2	91	0.31	180	7.17
Van	Standard	2005	22396	16	4.9	91	0.31	180	7.31
Pickup	Standard (Dual Fuel)	1998	21456	14	5.2	91	0.31	160	8.19
Pickup	Standard (Dual Fuel)	2005	22953	15	4.9	91	0.31	160	7.27

Table 2.9: Electric Vehicle Assumptions in Base Case Scenario

Class	Size	Year	Price	Miles per KWH	Acceleration	Top Speed	Relative Emission	Range	Operatin Cost*
Car	Mini	1998	27038	5	5.7	65	0.00	80	7.89
Car	Mini	2005	18924	5	5.3	71	0.00	113	7.79
Car	Subcompact	1998	32448	4	5.7	65	0.00	100	8.39
Car	Subcompact	2005	22706	4	5.3	71	0.00	141	8.27
Car	Compact	1998	37853	2	5.7	65	0.00	100	9.87
Car	Compact	2005	26492	3	5.3	71	0.00	132	8.76
Car	Sports	1998	40559	3	4.0	75	0.00	100	8.88
Car	Sports	2005	28384	3	3.7	87	0.00	132	8.76
Pickup	Compact	1998	32948	2	7.5	60	0.00	120	9.87
Pickup	Compact	2005	23063	2	6.3	69	0.00	138	9.73
Van	Compact	1998	48461	2	7.6	64	0.00	120	9.87
Van	Compact	2005	33916	2	6.4	71	0.00	138	9.73

2.6 Results of Microsimulation for Base Case Scenario

This section reviews some aggregated results obtained from a BCS run of the ITS forecasting model. The results cover the entire SCE region, including "districts" assigned to other utilities, during the years 1994 to 2010. The model itself provides results at the district level for use by SCE staff on their GIS systems (see Chapter 3 of the Final Report). However, ITS has not yet explored the model behavior at the district level or produced error bands for forecasts. This work is part of the ongoing research project, and is planned for 1995 and 1996 under funding from SCE and CEC.

Before exploring these results, it is useful to review a few features of the results produced by this model. The model is a scenario analysis tool that makes the following assumptions:

1. When vehicles are identified as "available in the market," it is assumed that all consumers are fully aware of the existence of these vehicles, and are also familiar with the vehicle attributes through advertising, education programs, personal experience or word-of-mouth, etc.

2. Vehicles are widely available throughout the market, i.e., full channels of distribution through dealerships are in place.
3. Supplies of vehicles are fully available at the assumed purchase price. Thus, this model produces results under the assumption of what automobile companies call "free expression."

2.6.1 Vehicle Penetration

During the period 1997 to 2010, alternative fuel vehicles are introduced in California and begin penetrating the market. Our model simulates the vehicle purchase and behavior of households in the SCE service area. To observe how the model behaves in regard to overall vehicle purchases over time, see Figure 2.2. The top line represents total vehicles in California as predicted by our model. The current version of our model predicts a substantial increase in vehicles during this period. One factor is that our microsimulation of demographic changes in California yields additional households at an average annual rate of 2.76%. Another factor (which cannot be seen from these aggregate figures), is that, in addition to the number of households increasing, the number of vehicles per household is forecasted to increase due to the occurance of simulated transactions in which households add vehicles rather than simply replacing them. In addition, the current version of our model uses transaction and disposal models based exclusively on stated intentions from the Wave 1 Survey. Generally, transaction rates are probably optimistic. Disposals represent planned disposals only, and do not include disposals due to unplanned events. In future versions of the model based on panel data, these will be adjusted. The scale of Figure 2.2 makes it difficult to show alternative fuel vehicles broken down by fuel type. For a plot of vehicle totals that excludes gasoline, see Figure 2.3. The number of alternative fuel vehicles in California grows from zero in 1996 to approximately 5.3 million in the year 2010.

An item of special interest, due to the CARB clean fuel vehicle mandates, is the market penetration of alternative-fuel vehicles in the years 1998, 2003, and 2010, respectively, as measured by new vehicle sales. Results for these years appear in Figure 2.4. Under the BCS modeling assumptions cited above, alternative fuel vehicles generally meet the CARB mandates. The strongest competitor under this scenario is Natural Gas. This is apparently due to a highly competitive combination of low operating cost, cleanliness, medium-level range, and capital cost relative to gasoline vehicles. Electric vehicles meet the 2% 1998 mandate, but fall a bit short of the 10% 2003 target due to the heavy competition from Natural Gas and Methanol. However, it should be noted that the BCS purchase price and operating cost assumptions are quite conservative for EVs. The Gasoline share drops to 65% in 1998 and stays there as the technology for these vehicles continues to improve. A more detailed plot of purchase shares over time is given in Figure 2.5.

Figure 2.2

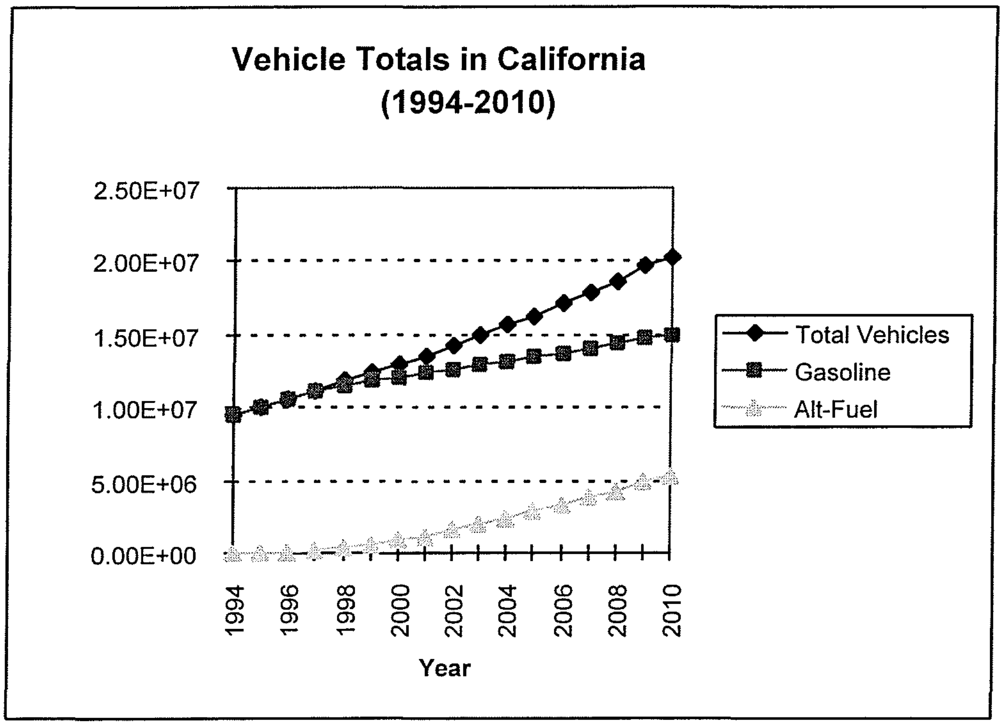


Figure 2.3

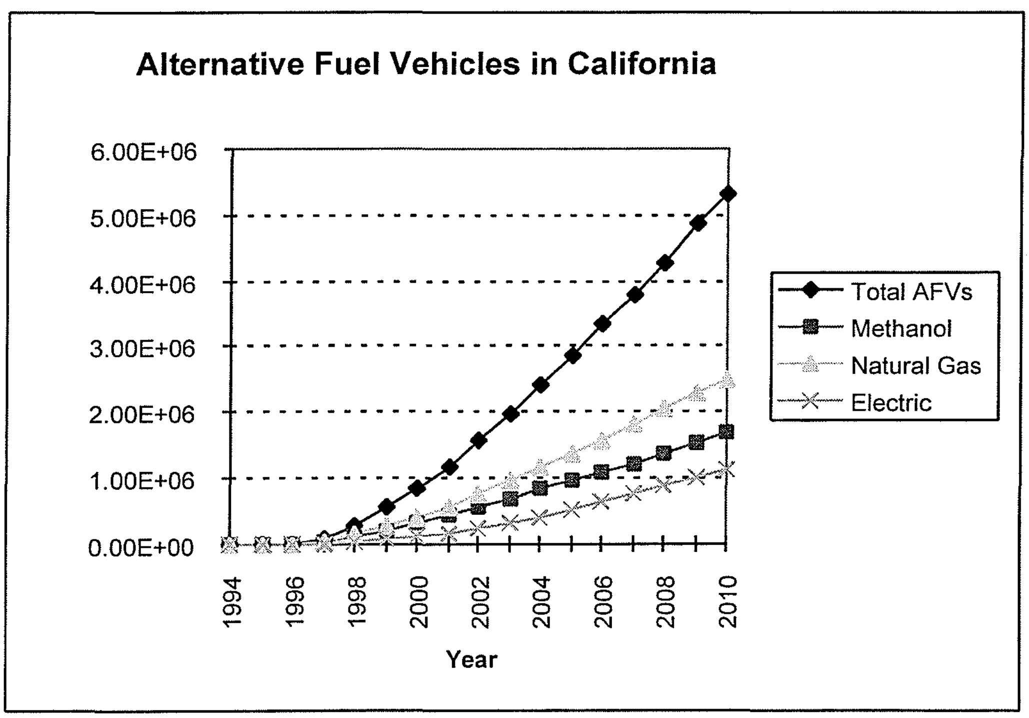


Figure 2.4

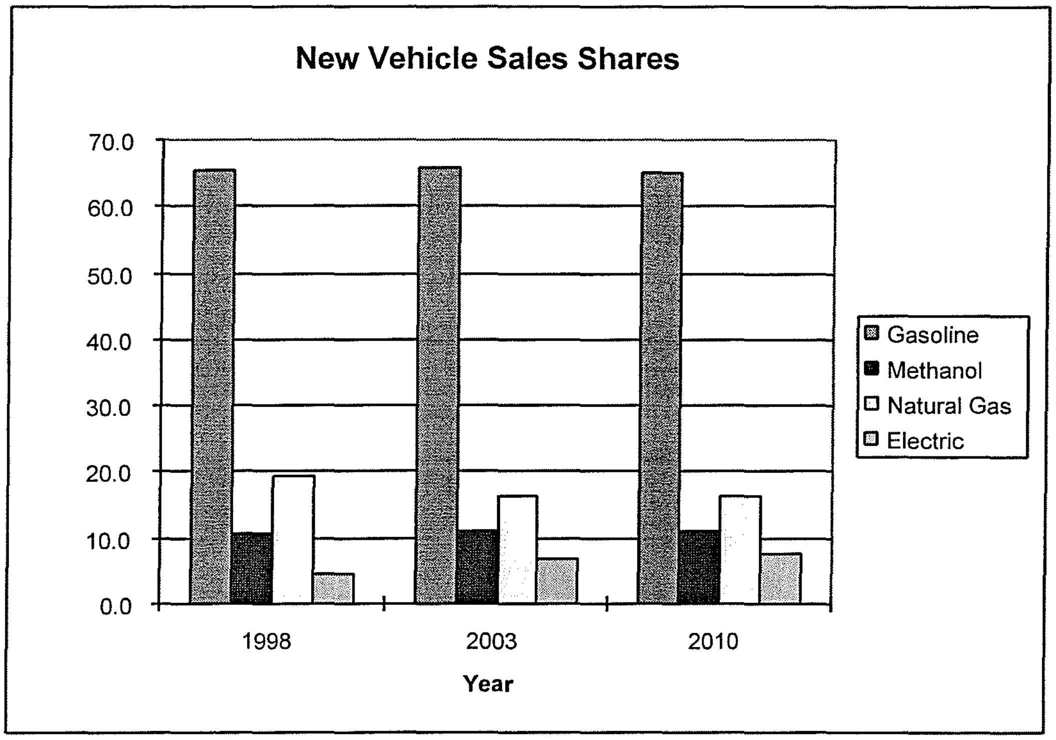
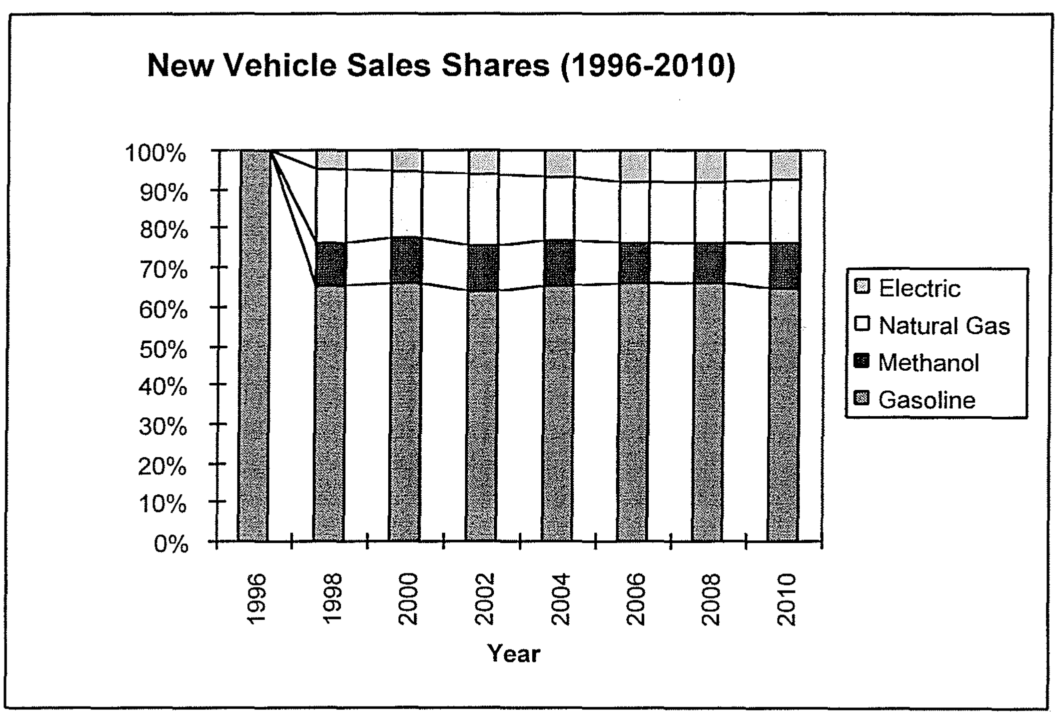
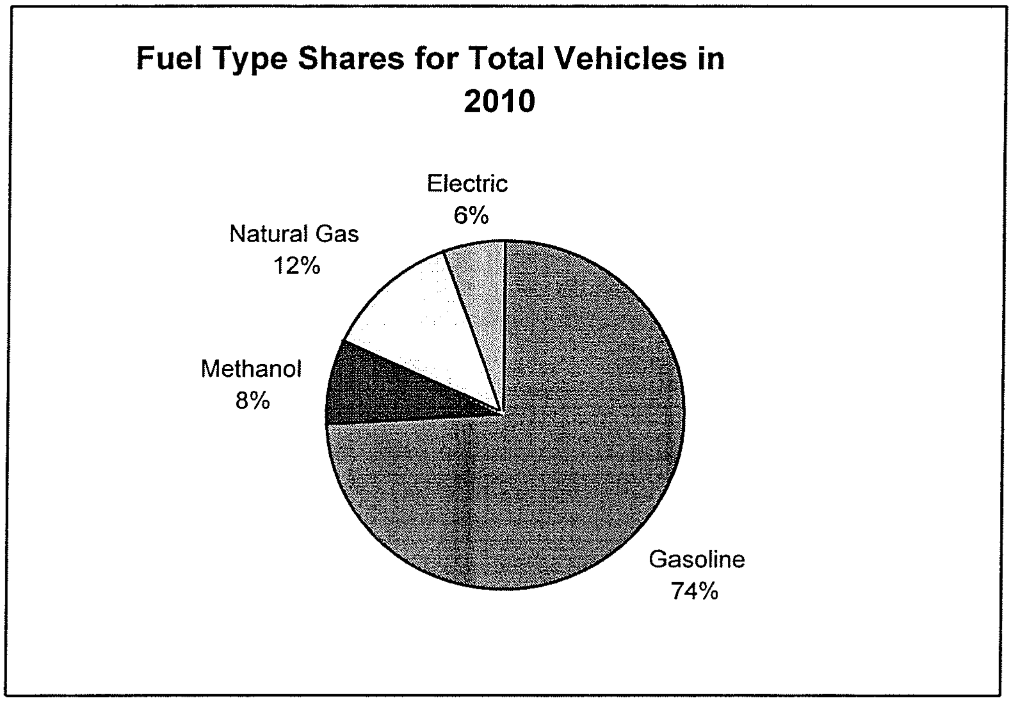


Figure 2.5



One important observation that may be of interest to those concerned with air quality is the cumulative effect of these vehicle purchases. In the year 2010, what will be the overall share of clean fuel vehicles on the road? This information is summarized in Figure 2.6. Note that, with relatively successful market penetration of AFV's during this period, the overall share of AFV's is approximately 25 percent in the year 2010.

Figure 2.6

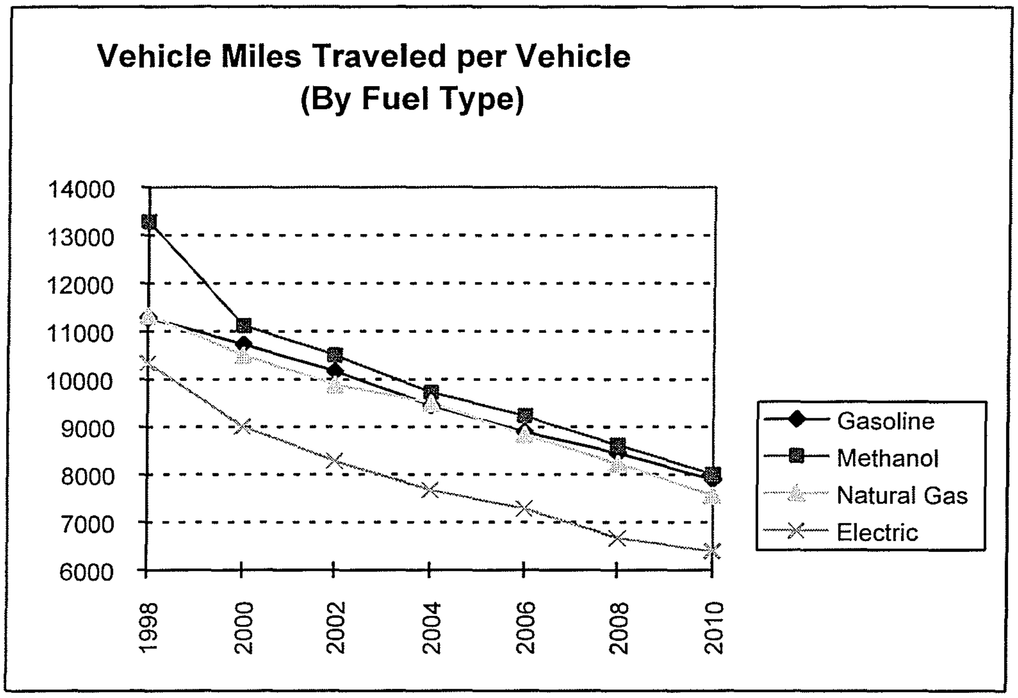


2.6.2 Vehicle Utilization

In addition to the number of vehicles and market penetration issues, the forecasting model also simulates annual vehicle miles traveled (VMT) by households at the individual vehicle level, and then aggregates these figures. VMT figures are important because the combination of numbers of vehicles, plus their VMT's, impact both fuel usage and air quality in the area. In particular, the number of electric vehicles, along with vehicle miles traveled, affects the electrical utility infrastructure (see next section).

An important feature of the ITS modeling system is that it predicts how vehicles will be used based on household demographics, total vehicle portfolios, and characteristics of the vehicle including range and operating cost. Vehicles will vary in their usage, as illustrated in Figure 2.7. Generally speaking, electric vehicles are driven fewer miles due to their limited range, with the remaining vehicles being similar to one another. There is a downward trend in the overall VMT per vehicle during the forecasting period, which is in contrast to the recent slightly increasing trend for gasoline vehicles. This prediction from the model is due primarily to the increase in the number of vehicles per household. The annual total VMT in California will continue to increase.

Figure 2.7

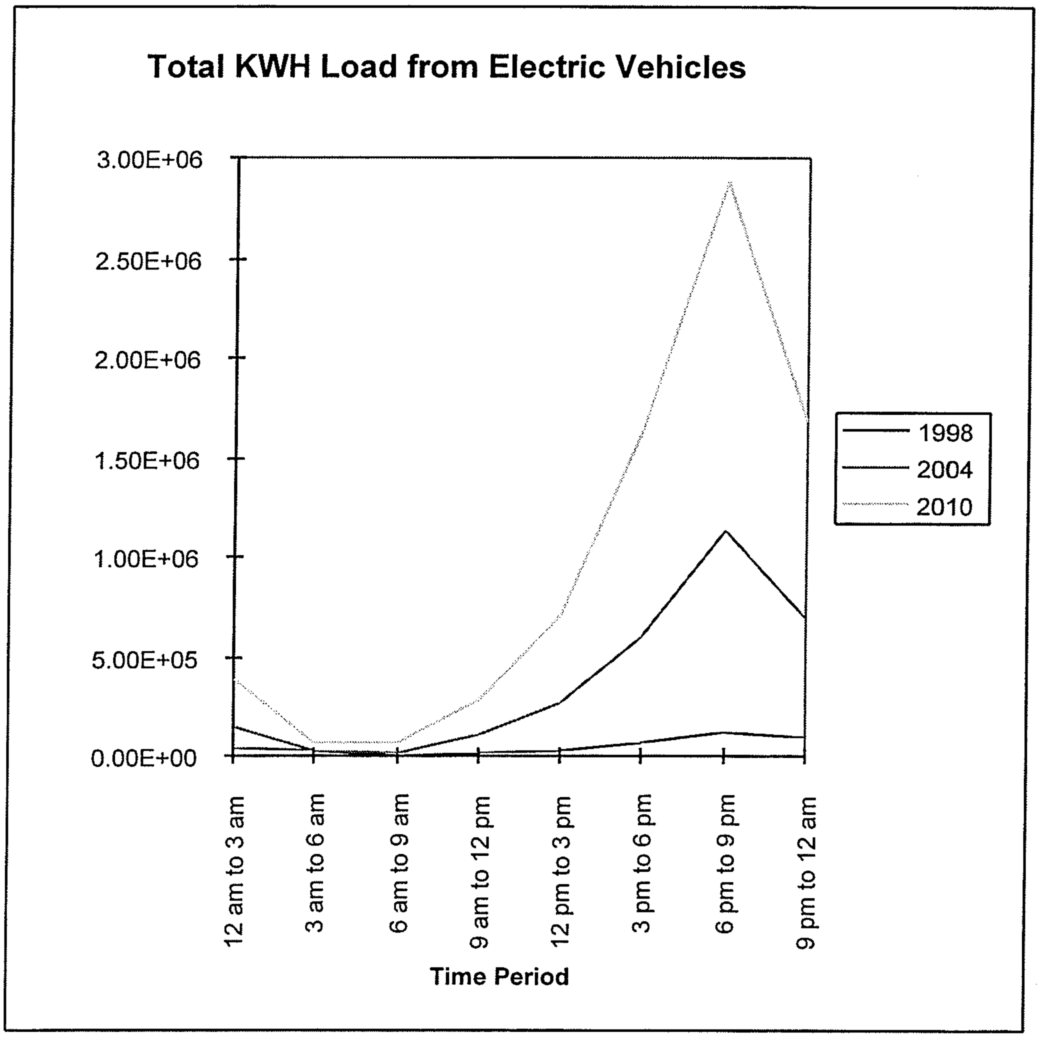


2.6.3 Kilowatt Load Due to Electric Vehicles

Finally, an important feature of the ITS model is that it provides estimates of electrical load due to the penetration of electric vehicles. The current version of the model assumes that EV's use 6 KV charging units, and that charging is uniform. Plug-in times for vehicles are based on data collected from Edison's vehicle trials research. See Figure 2.8 for the forecasted system-wide electrical load for the years 1998, 2004, and 2010. Load increases steadily along with the penetration of electric vehicles (despite the decrease in VMT per vehicle).

It should be noted that the load curves in Figure 2.8 represent unconstrained charging behavior with no load management. In the future, ITS can work with SCE to incorporate more sophisticated features and model enhancements, such as a charging unit characteristics file that can vary over time and from vehicle type to vehicle type, and also a means to examine the impact of load management techniques. Vehicle utilization questions from the Wave 2 Personal Vehicle survey will be used to improve the models for electric vehicle utilization.

Figure 2.8


Total KWH Load from Electric Vehicles
3.00E+06
2.50E+06
2.00E+06
1.50E+06
1.00E+06
5.00E+05
0.00E+00
12 am to 3 am
3 am to 6 am
6 am to 9 am
9 am to 12 pm
12 pm to 3 pm
3 pm to 6 pm
6 pm to 9 pm
9 pm to 12 am
Time Period
1998
2004
2010

3. PROGRAMMER INTERFACE GUIDE

3.1 Introduction

A microsimulation modeling system for forecasting vehicle transactions, fuel usage, and time-of-day recharge demand has been developed by researchers at the University of California's Institute of Transportation Studies (Irvine and Davis). The system integrates a collection of behavioral and demographic models to produce dynamic, multi-year forecasts for specified locations in California. This chapter addresses the portion of the system devoted to the personal/household vehicle market in the Southern California Edison (SCE) service territories. A separate module for commercial fleets is discussed in Chapter 12 of this Final Report.

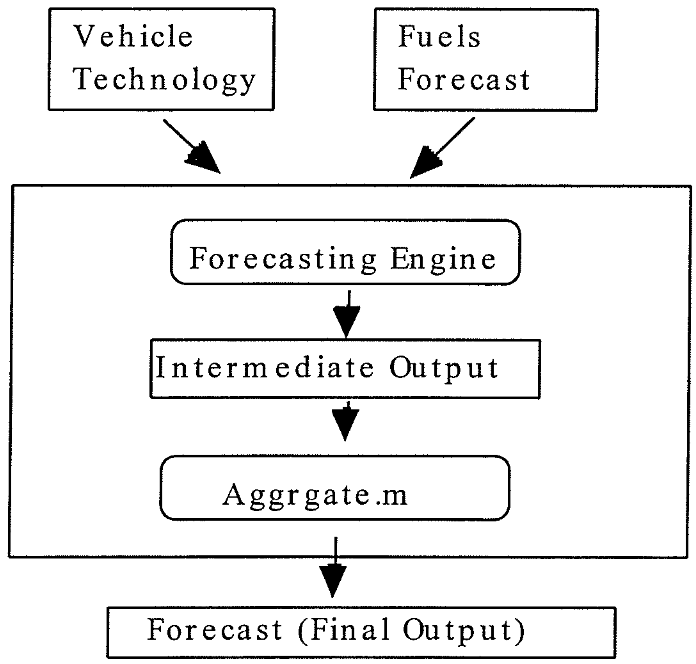
The system has been implemented as a "black box" so that it can be can be executed by other applications. For example, GIS application programs like ARC/INFO can be used to perform all user-interface and display functions, but can call the forecasting system to perform the necessary calculations and produce results. The purpose of this document is to provide the necessary information for programmers seeking to provide forecasting system access to their users. It is envisioned that most users will require a friendly user interface to operate the system, and that additional programming and user documentation will be required to augment the basic forecasting system. However, a knowledgeable user could take the information contained in this document and operate the system using a "user interface" as simple as a text editor or a spreadsheet (such as Excel, Lotus, Quattro, etc.). In either case, detailed technical documentation on the forecasting system is not required to perform these functions, and the information contained in this chapter is sufficient for this purpose. Accordingly, more detailed technical documentation is provided in other chapters of this Final Report.

3.2 Overview of Forecasting System

The forecasting system is implemented using MATLAB, a high-performance numerical computation language that is very portable and can be run on almost any platform. To facilitate portability, the system is designed so that all inputs and outputs are stored as ASCII data files. Input files will be set up external to the system in a manner to be determined by the programmer. Then, a microsimulation run can be performed, e.g., by launching a "batch" job (on DOS systems) or by starting a background process on UNIX systems. When the run is complete, the final results are contained in an ASCII output file (or files).

The current version of the system is not a completely black box: the system is designed so that the programmer may have access to a MATLAB program module, "aggrgate.m", that will allow modification of the contents and format of the final output file (or files). A simplified version of the system structure is shown in Figure 3.1. Files are denoted by boxes, and program modules are denoted by boxes with rounded corners.

Figure 3.1: Structure of Forecasting System.



The forecasting engine generates results by iteratively performing microsimulations for a sequence of six-month time periods. In other words, there are two microsimulations required per year. At the end of each year, intermediate output data are produced by the forecasting engine. However, these data are more detailed and in a different form than most users may wish to see. The steps of aggregating the intermediate output, re-formatting the results, and appending them to the final output file is performed by the MATLAB module "aggrgate.m." [1] (Once the aggregation step is complete, the intermediate output data may be erased or over-written to conserve disk space.) The microsimulation step for the next forecasting period then begins. At the conclusion of a completed run, the final output file will contain results for all years in the forecasting scenario.

This structure gives the programmer some flexibility and control over the output file contents, allowing changes in response to user requests, efficient management of disk space, and so forth. MATLAB interprets ASCII source code, so the programmer can directly modify the contents of the file aggrgate.m. Although this gives flexibility, such

[1] The module is called "aggrgate.m" rather than "aggregate.m" because module names (excluding extensions) are being limited to eight characters to maximize portability across platforms.

changes should obviously be made with extreme caution. Under no circumstances should the programmer make changes to the forecasting engine itself.

The remainder of this chapter contains relevant information on running the system, input file formats, the default output file format(s), and other interface-related issues.

3.3 Running the Forecasting System

This section describes how a programmer can execute commands to run the forecasting system. Examples will be oriented toward those using UNIX systems. Appropriate system-dependent modifications can be made by knowledgeable programmers using other platforms.

As indicated in Figure 3.1, there are two input files required by the system. The system produces a final output file (or files). To operate successfully, the system needs to know the names of all input and output files, and also the final year of the forecast. This information is passed to the system in the ASCII file "microsim.dat." In the current version of the system (designed to SCE specifications), the programmer creates a version of microsim.dat containing six data elements in comma-delimited format, as shown in Table 3.1. The contents of the first two (input) files must be correctly set up in advance of running the system: see sections 3.4 and 3.5 for the input file formats. Input and output files are stored in ASCII, comma-delimited format in this version of the system to satisfy SCE specifications; however, other ASCII formats (e.g., tab-delimited or space-delimited) can also be easily accommodated by MATLAB. It is assumed that the programmer is responsible for implementing the appropriate error checking on the input values as part of the user-interface. In particular, we recommend that users not be allowed to alter any information in the vehicle technology file prior to 1994. This information should be regarded to be "historical" and immune to change. Altering historical information could cause the system to crash, because the system must be able to locate all vehicles held by our current household sample in order to perform model calculations.

Once the system has been executed, the aggregated results are contained in three output files using the names included in data elements 3-5 of microsim.dat: the format of the output files is given in section 3.6 below. The programmer/user then takes responsibility for the analysis and use of the results. Because microsimulation is somewhat time-consuming, the final year is given as an input so that unnecessary computations may be avoided. For example, if forecasts through the year 1999 are the only ones currently of interest to the user, it would be a waste of time to perform calculations through the year 2010. The current system restricts the range of final years to 1994 through 2010, inclusive. If a year is given before 1994, the system will assume that the final year is 1994. If a year is given that exceeds 2010, the final year is set to 2010.

Table 3.1: Contents of the file "microsim.dat."

Data element 1:	Name of vehicle technology file (text string)
Data element 2:	Name of fuels forecast file (text string)
Data element 3:	Name of output/results file one (text string)
Data element 4:	Name of output/results file two (text string)
Data element 5:	Name of output/results file three(text string)
Data element 6:	The final year of the forecast (an integer between 1994 and 2010, inclusive)

Data elements are stored in comma-delimited format.
UNIX example:
~/tony/vehtech1.txt, ~/tony/fuels1.txt, ~/tony/results1.txt,
~/tony/results2.txt, ~/tony/results3.txt, 2003

IMPORTANT: For the system to work properly, MATLAB must be able to locate all the files that it needs to use. Specifically, the system must be able to locate: (1) the forecasting system program files and supporting data files, (2) microsim.dat, and (3) the input data files. In addition, the system should write the final output file to a directory or subdirectory where the programmer/user-interface expects to find it. In order to describe how this can be achieved, we must first give the relevant background information regarding how the system should be run.

We recommend that the system be run in the <u>background</u>, in a way that is transparent to the user. This is easy to achieve in UNIX by using a <u>script file</u>. The user can go about other business until the results are available. It is the programmer's responsibility to write script files, and to decide how to handle notification of the user when results are available. This is an important consideration since microsimulation forecasting is computationally intensive and cannot produce results in "real-time" as they would be, for example, in a spreadsheet re-calculation. The programmer must address this issue according to his or her own best judgment.

To implement the suggestion of running the system in the background, is helpful to first understand how MATLAB operates. In its normal operating mode, MATLAB runs in the <u>foreground</u> and expects to interact directly with the user. However, MATLAB can be automated with an appropriate script file that contains the necessary commands for performing a microsimulation run, and the programmer can execute the script file in the background.

Specifically, the script file should do two things: (1) change the default directory to the location of the forecasting system program files, and (2) run MATLAB. An example UNIX script file called "mscript" is given in Table 3.2. In this example, it is assumed that the programmer has stored all the forecasting system program files, and the microsim.dat file, in the subdirectory named “tony.” The script file performs the directory change before running MATLAB.

IMPORTANT: By convention, the modeling system assumes that the file microsim.dat is located in the same directory as the forecasting system program files. In addition, an input file called “sim.in” is located there. The forecasting system will not run properly if either of these two files is missing. The file microsim.dat is used by the programmer to pass information to the forecasting system. The file sim.in should not be altered by the programmer.

Table 3.2: Contents of "mscript".

```
cd ~/tony
matlab < sim.in > sim.out
exit
```

Under this approach, the programmer would only need to execute the following command under UNIX: "mscript&". This command executes the script file commands in the background. The contents of sim.in are directed to standard input for MATLAB, and contain the commands to run the system. Standard output is directed to “sim.out” so that it does not go to the screen. The contents of this file would come in handy if problems arise during the microsimulation run. In order to deal with more complex issues such as, for example, user notification, the programmer may need to add additional appropriate commands to the script file. Also, knowledgeable users could manipulate system search paths to alter the requirements described above.

Under the above convention, the only remaining issue is the location of the input and output files. Again, see Table 3.1. In Table 3.1 all file names are fully-specified, i.e., they include directory path information. This ensures that the exact location of files is unambiguous, and we recommend that the programmer follow this convention.[2]

3.4 Vehicle Technology File Format

The first input file contains data for all vehicle technologies required to perform the microsimulation. This includes both historical data on vehicles (for years 1976 through 1993), and all vehicles that are anticipated to be available in the future. One of the final deliverables of the ITS project is a "base case" vehicle technology file that provides a

[2] If file names are not fully-specified with the directory path information, then MATLAB will assume that files are located in the default directory (e.g., ~/tony in the example from Table 3.2). If it fails to find them there, it will also search along the default MATLAB search path. However, if the files are not located in either of these two locations, the system will produce an error message and then exit.

single forecasting scenario (see Chapter 2 in the ITS/SCE Final Report). Users may make changes to the "base case scenario" to produce their own forecasting scenarios. This is the preferred approach, since there is a large amount of data required to completely define a scenario.

The current version of the system is based on behavioral models produced using the Wave 1 Personal Vehicle Survey. These models are estimated using a system of vehicle descriptions that characterize vehicles in terms of body/size and fuel classes, as well as there "generic attributes" such as vintage, purchase price, range, acceleration, etc. See Table 3.3 for the file format. Sixteen vehicle classes (attribute 1) and five vehicle fuel types (attribute 2) are used to characterize vehicles. The vehicle class determines body type and size of the vehicle; the fuel class determines the type of fuel (or fuels) used by the vehicle. Codes are included in Table 3.3.

NOTE: ITS has produced two forecasting models, one for personal vehicles and one for commercial fleets. The input file formats have been structured so that both models can use the same input files. For example, vehicle classes 15 and 16 have been included for possible use by the commercial fleet model, but these classes are ignored by the personal vehicle model. In addition, the refueling time attributes 10 and 11 are used to include information that is needed by both models. Specifically, the personal vehicle model uses attribute 10 to store home refueling times for electric vehicles, but it ignores attributes 10 and 11 for CNG vehicles. However, the fleet model uses the values for CNG vehicles as indicated. During the course of the ITS research project it has become obvious that a future version of the system may require separate vehicle technology files for the personal and commercial models, since the characteristics of these two markets are likely to be different.

The vehicle technology file is stored in ASCII comma-delimited format, as previously discussed. Each row in the file corresponds to a model year/vehicle technology combination. There are 35 years (1976-2010) potentially covered by the system, 16 class codes, and 5 fuel types, giving a total of 35 x 16 x 5 = 2800 rows in the file. The number of rows is assumed to be fixed, even though all data may not be used in any particular forecasting scenario. Specifically, many vehicle technologies may not exist in the market, even though there are "place-holder" records for them. A key variable in this regard is the "Market Availability" variable. It can be toggled "on" ("1") or "off" ("0") by the user for individual records. In addition, if the number of models is zero, then the vehicle technology will not exist. However, as noted previously, the programmer (and user) are responsible for error checking the data. For example, allowing values of zero for purchase price, acceleration, etc., will produce nonsense results if the vehicle technology is introduced without carefully entering relevant values for all variables.

For additional details on the contents of vehicle technology files, see Chapter 2 and also Appendix C.

3.5 Fuels Forecast File Format

The second input file provides information related to fuels over the forecasting period. This includes fuel costs (for both service station and home refueling, if applicable), and the availability of fuel at service stations: see Table 3.4. Fuel availability is measured by an index from 0 to 1 (0 = no stations, 1 = the same number of stations as gasoline). The current base year of the system is 1993, meaning that the first microsimulation year is 1994. The maximum final forecast year is 2010, giving a maximum number of 17 years. There are four types of fuels in our models: gasoline, methanol, compressed natural gas, and electricity. There is one row per fuel per year, giving 17 x 4 = 68 rows in the file. As for the vehicle technology file, the number of records is fixed. However, records cannot be "turned on" or "turned off": reliable data should be available for all records.

As already noted, the file is stored in ASCII comma-delimited format. For a more detailed description of the base case fuels forecast file, see Chapter 2 and Appendix C.

Table 3.3: Vehicle Technology File Format

1. Vehicle Class Code	Integer, 1 to 16.
2. Vehicle Fuel Type Code	Integer, 1 to 5.
3. Vintage	Year (1976 to 2010)
4. Purchase Price (New)	Dollars
5. Fuel economy	MPG (for non-EVs) or, Miles per Kilowatt Hour
6. Acceleration Time (0 to 30 MPH)	seconds
7. Top Speed	MPH
8. Number of Models	
9. Service Station Refuel Time	Minutes
10. Refuel Time Two (= Home refuel time for personal electric vehicles, or slow-fill on-site refuel time for CNG fleet vehicles.)	Hours
11. Refuel Time Three (= On-site refuel time for electric fleet vehicles, or fast-fill on-site refuel time for CNG fleet vehicles.)	Minutes
12. Emissions Index	Fraction from 0 to 1. 1 = 1994 gasoline vehicle.
13. Range on Full Tank (Full Charge)	Miles
14. Luggage Space Index	Fraction from 0 to 1. 1 = comparable gasoline vehicle
15. Market Availability	1 = Vehicle is available, 0 = Not available.

Attribute 1. Vehicle Class Codes

1 - mini	9 - standard pickup
2 - subcompact	10 - minivan
3 - compact	11 - standard van
4 - intermediate	12 - small sport utility vehicle
5 - large	13 - large sport utility vehicle
6 - luxury	14 - mini sport utility vehicle
7 - sports car	15 - truck (10-14 KGVW)
8 - compact pickup	16 - shuttle bus

Attribute 2. Vehicle Fuel Type Codes

1 - gasoline	4 - dual fuel CNG
2 - methanol	5 - electric (dedicated)
3 - dedicated CNG	

Table 3.4: Fuels Forecast File Format

1. Year	1994 to 2010
2. Fuel Code	Integer, 1 to 4 (see below).
3. Fuel Cost One	Dollars per gallon at service station (for non-electric) or, Dollars per KWH (off-peak for electric)
4. Fuel Cost Two	On-peak for electric, Dollars per KWH
5. Fuel Availability Index (No. of stations relative to gasoline.)	Ratio from 0 to 1. 1 = gasoline service stations

Attribute 2. Fuel Codes

1 - gasoline 2 - methanol 3 - compressed natural gas (CNG) 4 - electricity

3.6 Output File Formats

This section contains the formats for the three output files currently produced by the forecasting system--see Tables 3.5a to 3.5c. Recall that these files are produced by aggregating results produced by the microsimulation. Aggregation of results is desirable to provide more useful formats for users, and also necessary to conserve disk space. As previously discussed, the design of the system will allow the programmer to modify the contents of the program file "aggrgate.m" to create different sets of results. However, the current version of this document does not contain the information to allow the programmer to do this. This will be made available in a later document. The three files are in ASCII comma-delimited format.

Table 3.5a: Output File One (16 data elements)

Output variable
1. District ID (key/index variable)
2. Year (key/index variable)
3. Total no. of vehicles of all types (in the district).
4. Total VMT of all vehicles (in the district).
5. Total number of gasoline vehicles.
6. Total VMT for gasoline vehicles (in 100,000's of miles).
7. Total fuel usage for gasoline vehicles (in 1000's of gallons).
8. Total number of methanol vehicles.
9. Total VMT for methanol vehicles (in 100,000's of miles).
10. Total fuel usage for methanol vehicles (in 1000's of gallons).
11. Total number of dedicated CNG vehicles.
12. Total VMT for dedicated CNG vehicles (in 100,000's of miles).
13. Total fuel usage for dedicated CNG vehicles (in 1000's of gallons).
14. Total number of dual fuel CNG vehicles.
15. Total VMT for dual fuel CNG vehicles (in 100,000's of miles).
16. Total fuel usage for dual fuel CNG vehicles (in 1000's of gallons).

Table 3.5b: Output File Two (24 data elements)

Output variable
1. District ID (key/index variable)
2. Year (key/index variable)
3. Total number of electric vehicles.
4. Total VMT for electric vehicles (in 100,000's of miles).
5. Total kilowatt hours for electric vehicles (in 1000's of KWH).
6. Total KW: 12 midnight to 3 AM.
7. Total KW: 3 AM to 6 AM.
8. Total KW: 6 AM to 9 AM.
9. Total KW: 9 AM to 12 noon.
10. Total KW: 12 noon to 3 PM.
11. Total KW: 3 PM to 6 PM.
12. Total KW: 6 PM to 9 PM.
13. Total KW: 9 PM to 12 midnight.
14. Percentage of all vehicles that are gasoline.
15. Percentage of all vehicles that are methanol.
16. Percentage of all vehicles that are ded-CNG.
17. Percentage of all vehicles that are DF-CNG.
18. Percentage of all vehicles that are electric.
19. Percentage of all vehicles that are AFVs.
20. Percentage of AFVs that are methanol (= 100 * total number of methanol vehicles /total number of AFVs).
21. Percentage of AFVs that are ded-CNG.
22. Percentage of AFVs that are DF-CNG.
23. Percentage of AFVs that are electric.
24. Total no. of new vehicles purchased.

Table 3.5c Output File Three (27 data elements)

Output variable
1. District ID (key/index variable)
2. Year (key/index variable)
3. Percentage of new vehicles that are AFVs.
4. Percentage of new vehicle purchases that are methanol.
5. Percentage of new vehicle purchases that are ded-CNG.
6. Percentage of new vehicle purchases that are DF-CNG.
7. Percentage of new vehicle purchases that are electric.
8. Total no. of used vehicles held by households.
9. Percentage of all used vehicles that are AFVs.
10. Percentage of newly purchased AFVs that are methanol (=100 * # of new methanol vehicles/# of new AFVs).
11. Percentage of newly purchased AFVs that are ded-CNG.
12. Percentage of newly purchased AFVs that are DF-CNG.
13. Percentage of newly purchased AFVs that are electric.
14. Total number of electric minicars.
15. Total number of electric subcompact cars.
16. Total number of electric compact cars.
17. Total number of electric intermediate cars.
18. Total number of electric large cars.
19. Total number of electric luxury cars.
20. Total number of electric sports cars.
21. Total number of electric compact pickups.
22. Total number of electric standard pickups.
23. Total number of electric minivans.
24. Total number of electric standard vans.
25. Total number of electric small sport utility vehicles.
26. Total number of electric large sport utility vehicles.
27. Total number of electric mini sport utility vehicles.

3.7 Intermediate Output File Format

This section describes the format of the intermediate output file. The current version of the system performs some accumulations for efficiency purposes, but leaves the final formatting and aggregation to aggrgate.m, as described previously.

In the intermediate ("temporary") output file, there is one record (row) per household in the microsimulation sample. Each record reports the expected (mean) values for vehicle holdings, VMT usage, fuel usage, and TOD recharging. These values are broken down according to the 70 possible vehicle technology classes (14 vehicle classes x 5 fuel types = 70 total classes for the personal vehicle model). In the case of vehicle holdings, counts are broken down by new vehicles and used vehicles. The ordering of vehicle class indexes is according to (I, J), where I = 1 to 14, J = 1 to 5, with the J index changing faster. In other words, the ordering is: (1,1), (1,2), (1,3), (1,4), (1,5), (2,1), (2,2)....(14,5).

Using this convention, the temporary file format is as follows:

Columns 1-70:	Expected counts for new vehicle purchases during the year.
Columns 71-140:	Expected counts of used vehicles held by households.
Columns 141-210:	Expected annual VMT for each of the 70 possible classes.
Columns 211-280:	Expected fuel usage for each of the 70 possible classes.
Columns 280-392:	Expected TOD recharging for the 14 possible EV classes.

In the last set of columns, column 280 contains KWH in period 1 for vehicle class 1, column 281 contains KWH in period 1 for vehicle class 2, etc. In the final output produced by the current version of aggrgate.m, we divide by 3 to get the KW used during each three hour period.

4. PROGRAM TECHNICAL DOCUMENTATION

4.1 Introduction

This chapter provides more details and technical documentation on the microsimulation modeling system for the personal vehicle market described in Chapter 3, "Programmer Interface Guide." Chapter 3 provides a useful overview of the system inputs and outputs, and how the system can be used to produce forecasts. Although there might be some duplication of information, we assume that readers of this chapter have already read Chapter 3 and are familiar with the information contained there.

Researchers at the University of California's Institute of Transportation Studies (Davis and Irvine) have developed microsimulation modeling systems for forecasting vehicle transactions, fuel usage, and time-of-day recharge demand in the commercial fleet and personal vehicle markets. This chapter focuses on the system for the personal vehicle market. (For commercial fleet documentation, see Chapter 12.) It integrates a collection of behavioral and demographic models to produce dynamic, multi-year forecasts for specified locations in California, namely, the Southern California Edison (SCE) service territories. The basic research performed to develop these models is described in various chapters of this final report. To produce a forecasting system, these models were implemented as modules and integrated into a microsimulation framework.

Modules are written in MATLAB, a fourth-generation programming language oriented toward high-performance numerical computation, and also in FORTRAN. One important feature of MATLAB is that it always uses source code, and is highly portable across many different platforms. As discussed in Chapter 3, the model system itself can be treated as a modular "black box" that uses ASCII data files for user inputs and outputs to maximize flexibility and portability. In some cases MATLAB execution time is slow, and it is necessary to implement subroutines in either FORTRAN or C. Modules are compiled and stored as executable library routines that can be called by MATLAB. In this system, many of the computationally intensive routines have been implemented as FORTRAN subroutines to speed the execution times.

This chapter gives a system overview containing the structure of the microsimulation algorithm, followed by more detailed descriptions of sub-modules used to perform demographic microsimulation and choice modeling calculations. For a high level of detail on the various model components, readers should consult the remaining chapters and the program module code that appears in Appendix A.

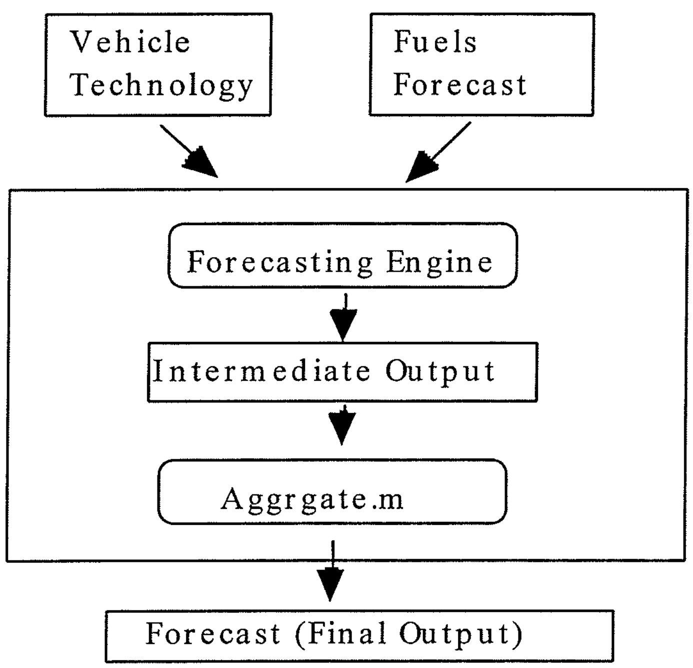
4.2 System Overview

Before proceeding to a more detailed view of the system, we present the "black box" view of the system from the perspective of a "user," as discussed in Chapter 3. See

Figure 4.1. The user has control over the contents of two ASCII files: vehicle technology, and fuels forecast. The contents of these files represent scenarios for which forecasts are desired.

The output of the current version of the forecasting system is stored in three comma-delimited ASCII files, as per SCE specifications--see Chapter 3. User-controlled input- and output-related information, including file names and the number of forecasting periods, are stored in 'microsim.dat'--again, see Chapter 3. The calculations that produce the final forecast numbers are performed by the MATLAB program module, 'aggrgate.m'. Ambitious programmers can alter the forecast result format by modifying aggrgate.m--see section 3.7. Additional modifications would also be required in the MATLAB program module 'foutput.m', which takes the forecasts stored in the temporary MATLAB data file 'output.mat' and creates the three ASCII output files prior to system termination.

Figure 4.1: Structure of Forecasting System



The forecasting engine generates results by performing a microsimulation for each of a sequence of six-month time periods. Vehicle transactions are simulated every six months, but aggregated results for vehicle counts, vehicle miles traveled, fuel usage, and (electric vehicle) time-of-day recharging are produced on a yearly basis. The output files contain one record per SCE district per year.

The structure of the microsimulation algorithm is given in Figure 4.2. The MATLAB program module 'simulate.m' is the main program module. It generally follows the structure of structure in Figure 4.2, calling a the sequence of sub-modules noted in square brackets ([ ])--see Appendix A for the current code used in simulate.m.

Figure 4.2: Steps in Microsimulation [simulate.m]

1. Perform initializations.
   Initialize global variables. [initglob.m]
   Read input file names, final year of forecast [readinpt.m]
   Load base-year (1993) household/vehicle information [initmdb.m]
   Set Year = Base_Year (=1993)

2. Year = Year + 0.5. Is Year > Final_Year?
   If YES, Go to 8.

3. "Age" the households by six-months (this includes reallocating vehicles, when necessary, for those households that split or merge). [agehh.m]

4. If Year = an integer, then:
   Update the vehicles that are available in this year's market.
   Update prices of used vehicles.
   [upvehmar.m]

5. Perform choice/usage/TOD recharge calculations. [choicemd.m]

6. If Year is an integer, then:
   Aggregate the results, and append them to the temporary output file.
   [aggrgate.m]

7. Go to step 2.

8. Read output file names and create ASCII files for the user.
   [readoutp.m]
   [foutput.m]

9. Microsimulation finished. Exit MATLAB.

4.3 Demographic Microsimulation

The models for demographic microsimulation are developed in Chapter 6. Because there is no causal connection in our models whereby vehicle holdings directly influence demographic changes (rather, we assume that the effect goes in the opposite direction), it is possible to logically separate these models so that all demographic microsimulation is performed off-line and stored in MATLAB data files. This saves substantial computing time. The modeling system contains a set of 35 data files corresponding to the 34 six-month forecasting periods for the years 1994 to 2010, plus the base year (1993). Files are named 'hhdemo##.mat,' where ## = 0 to 34. Files are structured to include pointers and transition flags so that households can be "aged." The main module call for aging households is to 'agehh.m,' which then calls other modules to perform various required calculations.

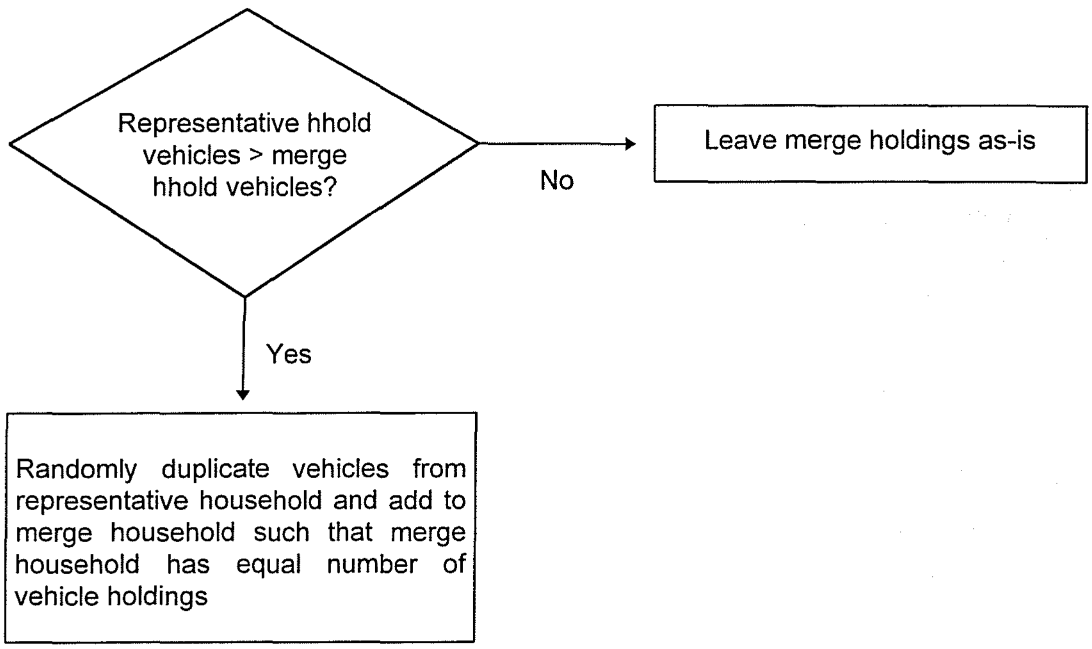
"Aging" households involves a number of issues other than people getting older. The demographic microsimulation model characterizes household changes according to 26 categories. These categories may be further consolidated into two main categories with respect to how household vehicle holdings are to be adjusted: merges (e.g., adults get married) and splits (e.g., adults get divorced or children leave the household). For example, when two single adults get married, they are each likely to bring a vehicle into the newly-created two-vehicle household. When an adult child leaves the household, he or she might take a vehicle. Rule-based algorithms for these two categories (merge and splits) have been developed to handle the disposition of vehicles when households undergo demographic transitions during the course of the microsimulation.

4.3.1 Merge Cases

In cases where households merge, such as through marriage, it is likely that the merged household will have more vehicles than either one of the single pre-merge components. However, the actual "merge" operation in the demographic microsimulation does *not actually merge two physical households from the database*. Rather, a household is simply "merged into," and the demographics section and household weights in the simulation database are adjusted to reflect the merge. Because the simulation does not actually merge two households, the vehicle adjustment is not as simple as consolidating two sets of vehicle holdings. Instead, the methodology we use is to find a household already existing in the database whose demographics match as closely as possible that of the merged household.

Figure 4.3 outlines the assignment rules for those merge categories that require adjustment.

Figure 4.3: Rules for Vehicle Adjustment in Merged Households



4.3.2 Split Cases

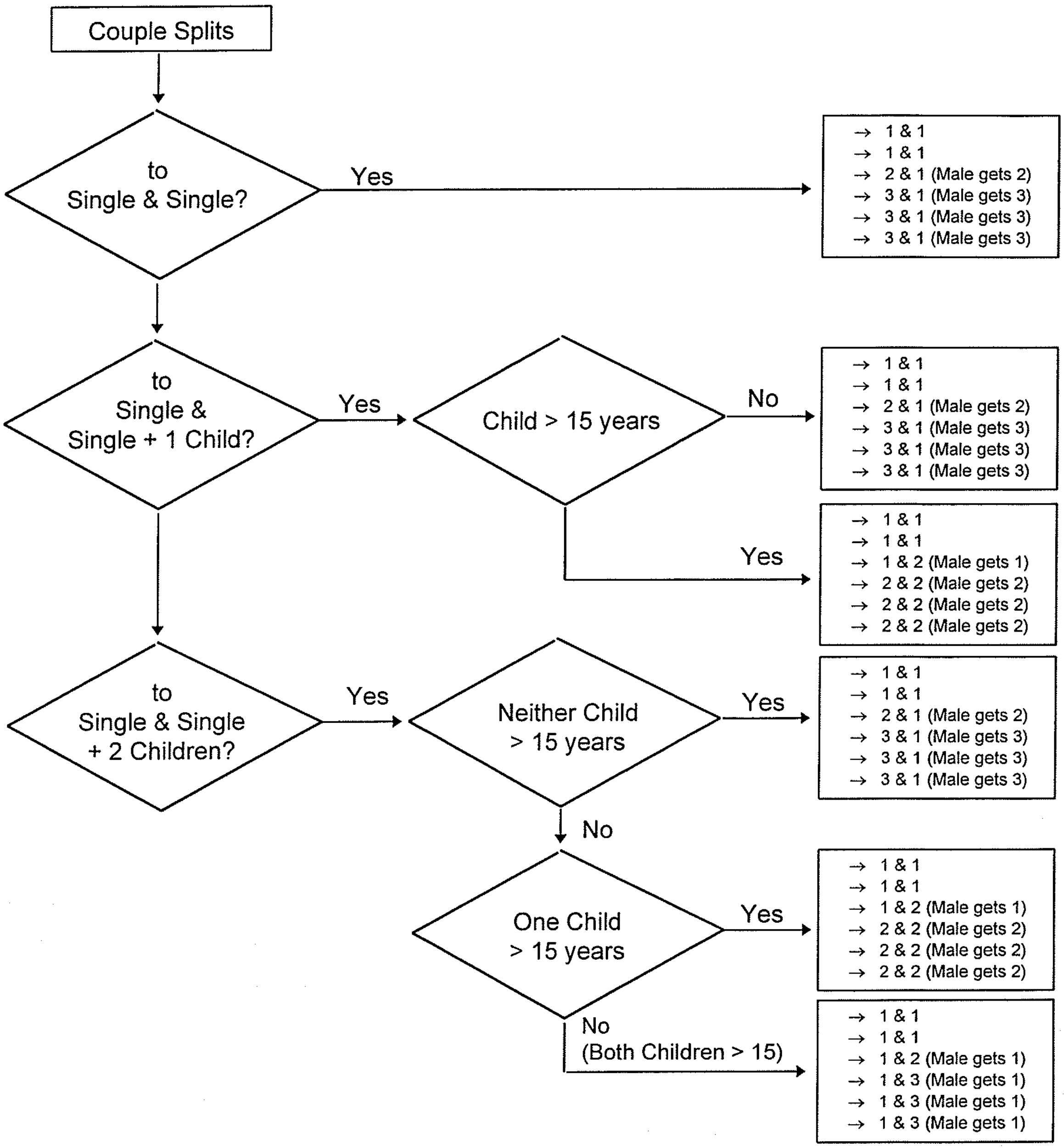
In cases where households split, such as from divorces or young adults leaving the household, it is likely that the two new split-off households will reassign the vehicle holdings from the original pre-split household. In these cases a set of hierarchical rules determines which household gets which vehicles.

Notes worth mentioning:

1. The "new" household resulting from the split holds the female(s) from the original household.
2. If a household had only one vehicle before, that vehicle is duplicated across both households.
3. Households coming from 5- or 6-car households will have at most only 3 cars after the split, since the database keeps at most four sets of held-vehicle attributes.

The flowchart in Figure 4.4 outlines the vehicle assignment rules for split cases.

Figure 4.4: Rules for Vehicle Adjustment When Households Split



4.4 Choice/Utilization/Recharge Model Calculation Framework

The module that manages all choice modeling calculations is 'choicemd.m'--see Appendix A. This module implements the current structure for the Wave 1 modeling system, in which there are separate modules for one-, two-, and three-or-more-vehicle households. Specifically, the module partitions the 'mother data base' of households into three sub-databases and calls the appropriate modules. If the variable Year is not an integer, then choicemd.m calls three routines [cnuf1c.m, cnuf2c.m, and cnuf3c.m] that: (1) compute choice probabilities, and (2) simulate household vehicle transactions.

If the Year is an integer, then choicemd.m calls an alternate set of three routines [cnuf1a.m, cnuf2a.m, and cnuf3a.m] that perform much more detailed calculations. In addition to computing choice probabilities and simulating transactions, these modules also compute vehicle miles traveled (VMT), fuel usage, and (for electric vehicles) time-of-day recharging for all vehicles in all choice set scenarios. This includes both vehicles that could be purchased as well as vehicles currently held by the households. Disaggregated results are accumulated by computing choice probability-weighted expected values for all quantities of interest and storing them in the arrays 'accum#' (for choice set vehicles) and 'haccum# (for held vehicles), where # = 1, 2, or 3. After all three routines have been called, disaggregated results for choice set and held vehicles are merged together according to vehicle type by calling the routine 'acc_held.m.' These results are stored in temporary files that are accessed by 'aggrgate.m' when it is called by simulate.m--see section 4.2.

4.5 Choice Model Calculation Details

The structure and details associated with the vehicle choice models and vehicle utilization models calculations are included in Chapters 7 through 9, and also in Appendix A. This section provides appropriate background material, but adding too much detail would make the section redundant with the other material contained elsewhere in this Final Report. All choice model and utilization calculations are conducted by the module 'choicemd.m.' This module performs database manipulations, and calls lower level modules described below.

In the current version of the model system, choice probabilities are partitioned into three conditionally independent probability calculations: (1) transaction occurrence, (2) vehicle disposal/ next intended body-type and vintage group, and (3) transaction type/vehicle type. Due to the complexity and modularity of the research efforts that were integrated into the final forecasting model, terminology may occasionally be inconsistent between the model estimation discussions that appear in the other chapters and the forecasting discussion that appears here. For example, hazard models were estimated to produce a transactions timing model (described in Chapter 7), but the model requirement for forecasting is the probability that a household will engage in a transaction during a six-month period. Additional mathematical derivations are performed and coded into the forecasting system to get 'transaction occurrence probabilities.' We now review the three types of probabilities, and how they are integrated.

The term 'transaction probability' refers to the probability that a transaction of <u>any type</u> will occur during the current six-month period: the two possible outcomes are 'does transact' and 'does not transact'. One may think of these two outcomes as 'decision branches'. Transaction probabilities are computed by the M-file function 'ccalc_px'. The function is generic, but there is a different parameter vector passed to the function depending on whether the household has one, two, or three-plus vehicles.

Assuming that a household is on the 'does transact' branch, the next level of the model subdivides the household's course of action into options consisting of: (1) all disposal opportunities for currently held vehicles (with the exception of 1-vehicle households), and (2) all possible purchase <u>intentions</u> for combinations of body-type and vintage group. In the personal vehicle model there are 14 body-types and 10 vintage groups, so there are 140 'branches' for 'next intended vehicle type' (14 body-types times 10 vintage groups), plus the number of held vehicles. This model is described in Chapter 8, and is based on the stated intention responses for the next intended vehicle transaction from the Wave 1 survey. The current model therefore assumes a market structure in which competition within a body-type category is more intense than across body-type categories. The probabilities for the second level of the model, by definition, sum to one. Therefore, the probability that a household will, for example, enter the market to make a purchase with the prior intent of buying a sub-compact car is obtained by multiplying the 'does transact' probability times the branch probability for 'new sub-compact car'. The module that is called to compute these 'level 2' probabilities is 'nu_model.m'. The call is generic, but there are different lower-level sub-modules for one-, two-, and three-plus-vehicle households.

The <u>actual</u> transaction type choice is modeled in the third (and final level) of the model. At the endpoint of each of the (non-dispose) level 2 branches, the system computes a set of transaction type/vehicle type probabilities. This model is also described in Chapter 8. The vehicle choices (not to be confused with transaction choices) available on this branch are limited to all vehicles in the associated vintage group. For example, even if a household <u>intends</u> to buy a new sub-compact car, all new vehicle types (e.g., new minicar, new sub-compact, new compact, etc.) are available. This is consistent with the Wave 1 survey, in which some switching was observed at the time the SP choice was made. (Note, however, that the model takes into account that there is a high probability of not switching body-types once an intention has been established in the level 2 model.)

The transaction type options are generated by considering all possible replacements of any of the currently held vehicles, and also all possible 'add' transactions from the vehicle choice set. The conditional transaction type probabilities are computed for each branch of the level 2 model by the routines chmv1.m, chmv2.m, and chmv3.m for one-, two-, and three-plus-vehicle households, respectively. The overall set of probabilities for all possible choice outcomes is obtained by appropriate multiplication of the three types of probabilities. In MATLAB, calculations are handled in a matrix format where the number of rows corresponds to the number of households, and the number of columns generally corresponds to the number of choice options. Probabilities are computed using appropriate MATLAB matrix multiplication notation. Given these probabilities, transactions are readily simulated by the Monte Carlo approach. The aforementioned calculations are performed in the modules with the leading characters 'cnuf'.

In the current version of the system, the different types of calculations are performed at six-month versus twelve-month intervals, as previously indicated. In mid-year situations, we compute choice probabilities and simulate transactions only. This is relatively efficient, since we can simulate branch choices sequentially and only perform calculations for those households engaging in the particular transaction. Furthermore, the VMT calculations, which are also quite time consuming, are avoided at the mid-year point. These calculations are performed by calls to cnuf1c.m, cnuf2c.m, and cnuf3c.m for one-, two-, and three-plus-vehicle households, respectively.

In the end-of-year calculations, a full set of choice probabilities is computed for all households. Subsequent to the transaction choice simulation, household weights are also multiplied into the choice probabilities for purposes of computing weighted expected vehicle counts, VMT's, etc. These calculations are performed by calls to cnuf1a.m, cnuf2a.m, and cnuf3a.m for one-, two-, and three-plus-vehicle households, respectively.

To review, end-of-year calculations for a given set of households are as follows:

1. Compute the probabilities of transaction occurrence.
2. Compute probabilities for all 'disposal/ intended body-type and vintage' branches.
3. For each branch on the 'disposal/body-type and vintage' tree:
    Construct all possible transaction type outcomes.
    Compute probabilities for all transaction type outcomes.
    Compute VMT's for all vehicles in each possible transaction type outcome.
    Compute fuel usage for all vehicles in all transaction type outcomes.
    Compute recharge profiles for electric vehicles in all outcomes.
    Using simulation, produce a transaction realization and store it.
    Compute weighted unconditional probabilities for all outcomes.
    Compute expected values for vehicle counts, VMT's, fuel usage, and
        recharging; store values in accum# and haccum#.
4. Update the temporary mother database with new transactions.
5. Return to 'choicemd.m.'

Details of these calculations may be seen by examining the source code in the modules cnuf1a.m, cnuf2a.m, and cnuf3a.m.

4.6 Vehicle Miles Traveled

VMT (vehicle miles traveled) figures are computed at the household level for all vehicles in each potential transaction outcome, as described in the previous section. Chapter 9 gives a detailed account of the models and approaches used to compute these figures, which were based on both stated preference and revealed preference vehicle usage data from the Wave 1 survey. VMT calculations have been implemented in a variety of MATLAB and FORTRAN Modules that start with the three letters 'cvmt.' See Appendix A and Chapter 9 for details.

4.7 Time-of-Day Recharging

An important feature of the system is that it provides aggregate time-of-day recharging profiles from electric vehicles that will facilitate load management activities of SCE. Recharging profiles are computed by indicating KWH in each of 8 three-hour periods: 12 a.m. to 3 a.m., 3 a.m. to 6 a.m., 6 a.m. to 9 a.m., 9 a.m. to 12 p.m., 12 p.m. to 3 p.m., 3 p.m. to 6 p.m., and 6 p.m. to 9 p.m. Recharging profiles are computed for each individual electric vehicle that appears in the microsimulation, and then aggregating using weighted probabilities.

The calculation is performed by the M-file function 'tod_func.' The specific function call is: [tod] = tod_func(vmt,plug_time,ev_range,fuel_eff,chrg_rating), where the names of the variables are self-explanatory. Using VMT and FUEL_EFF, the routine estimates the amount of battery discharge during the day. It is then assumed that the vehicle is plugged in at the PLUG_TIME, and that the battery is fully recharged by a recharging unit with a charger rating of CHRG_RATING, which is assumed to be 6 KW. Recharging is assume to occur uniformly until recharging is complete. Each household has its own PLUG_TIME, which is randomly drawn from a distribution of plug-in times obtained from the SCE vehicle trials. All electric vehicles associated with the household are assumed to use the household plug-in time, which is adequate for simulation purposes. The routine that assigns plug-times is 'plug.m'.

5. HOUSEHOLD SAMPLE WEIGHTING

The personal vehicle household sample was constructed as a geographically stratified random sample of residential telephone numbers in the study area. Weights, or expansion factors, are needed so that the survey sample will accurately reflect the underlying residential population. The first set of weights were constructed assuming that the respondents to our survey (7,099 observations who completed the CATI interview) are a completely random subsample of those whose telephone were chosen by the survey company. These original weights, described in Section 5.1, correct for the geographic stratificiation and the presence of multiple telephone lines in many households.

Once these original weights were constructed, we then compared our weighted survey population to the 1990 Census. If our original assumption that the respondents were a completely random sample was correct, then the weighted distributions from our sample should match the Census figures. Since these distributions did not match, we then constructed correction factors to correct for the biased response mechanism in our sample. Section 5.2 describes the construction of these correction factors. Finally, Section 5.3 describes the construction of additional correction factors to expand the forecasting sample (1336 observations with complete data in the SCE region) to the original sample who completed the CATI interview.

5.1 Original weight construction procedure

The idea behind our sampling weight construction is to estimate the probability that a person from our survey would be selected in a simple random sample of the study area. The original weights were calculated using population counts from the Census. The survey responses were then compared to population characteristics of the study region.

Accurate population and household counts for each district are needed to generate weights. The original weights are constructed using the 1990 Census of Population and Housing Summary Tape File 3 (STF3). The STF3 files contain such data geographically located down to the census tract level. These numbers in the census tracts are aggregated to the district level.

The census definition of household does not exactly match the survey definition of household. Specifically, the census definition of household includes roommates whereas our survey doesn't. As an alternative, the census identifies the number of families in a census tract, consisting only of related individuals. Since neither definition matches survey definition of household exactly, we used the household file to find the distribution of number of phone, and the family file for other characteristics such as income.

Originally, two weights are calculated for the survey data. The first weight applies to households and is based on the following two criteria. One, the sum of household weight for all households in a district must equal the total number of households for that district. Two, weights must be adjusted, but only based on the total number of phone numbers in the household. Since the total household count is based on the census definition of household, the weights do not need to be adjusted for the number of households sharing the phone. The second weight applies to individuals within a household and is based on the following two criteria. First, the sum of individual weight for all individuals in a district must equal the total population for that district. Second, the weights must be adjusted based on both the number of phone lines for the household that the individual belongs to and the number of other households sharing that phone.

5.2 Correction factors and new weight construction procedure

The weights described in Section 5.1 were used to estimate the joint distribution of sex, age, family structure, and income. These joint distributions were compared to those generated from the total number of persons (13,381 observations) in most of the Metropolitan Statistical Areas (MSA) in California from the March, 1993 Current Population Survey (CPS). The San Diego MSA and non-MSA area are excluded. This comparison indicated that our original assumption of a completely random survey response mechanism was incorrect. We therefore used a statistical matching procedure described below to generate correction factors for the original weights. The resulting new weights are constructed so that the weighted distributions from our survey sample will match the corresponding weighted distributions from the CPS.

Since the sampling weights are given for each observation in the CPS, we need to match each respondent in our survey to a group of respondents in the CPS. We do this by creating 'bins' in both the CPS and our survey sample which have roughly the same sex, family structure, age, and income. We then assume the original weights of the respondents in our survey have the same sampling weights as those in the corresponding CPS bin. The weights which result from this process have the property that the weighted distribution of the variables used in the matching process should be similar in both the CPS and our survey. If these weights are not same, we can adjust the original survey weights to CPS weights proportionally. These proportional factors are called the correction factors.

The bin numbers are generated as follows. First, numerical values are assigned to each category of sex, with or without children under 6 years old, number of persons in the family, age, and family income. The categories are as follows: Sex: male - 1, female - 2; Children under 6 years old: with - 1, without - 2; Total persons: 1 to 2 - 1, 3 to 4 - 2, 5 or more - 3; Age: 24 years old and younger - 1, 25 to 34 - 2, 35 to 44 - 3, 45 to 54 - 4, 55 to 64 - 5, 65 and older - 6; Family income: less than $15,000 - 1, $15,000 to $29,999 - 2, $30,000 to $44,999 - 3, $45,000 to $59,999 - 4, $60,000 to $74,999 - 5, $75,000 to $99,999 - 6, $100,000 to 124,999 - 7, $125,000 and more - 8. Bin numbers are generated using: Bin Number = Sex*10,000 + Child6*1,000 + Totper*100 + Age*10 + Family income. The bin numbers for the survey sample (7,099 observations who completed the CATI interview) are made following the same procedure as above.

After constructing bin numbers for each set of data, survey bin numbers are matched to census bin numbers. If there are survey bin numbers which do not appear in the census bins, the survey income category numbers are moved one up or down. If the bin numbers still do not match, the age category numbers are moved one up or one down, until all the observations are matched to census bins. Some times categories have to be moved more than one category. Once all survey respondents have been matched to a census bin, the survey respondents are given a weight which is a random draw for the set of weights in the census bin. These weights are then compared to the original survey weights and the original survey weights are rescaled so that the sum of the weights in each survey bin matches the sum of the weights in the corresponding census bin. This ensures that the total weights are the same for the same census and census bins.

Specifically, the procedure to generate correction factor and new sample weights using bin matching method are as follows.

a. Generate bins.

b. Generate total census weights by bin. (Sum of weights by bin. There are 477 bins in the CPS.)

c. Generate total survey weights by bin. (There are 426 bins in the survey.)

d. Merge census and the survey.

e. If there are bins do not match in the survey, shift the last two categories and generate new bin file. Re-generate total survey weight by newly shifted bin.

f. Merge census and the survey again.

g. In this case, all the survey bins are shown in the census bins. However, some census bins are not matched to the survey. If there are bins in the census that do not show in the survey data, do the shifting procedure for those bins and make a new census bin file. (Having those zero cells will lower the total population count.)

h. Re-generate total census weights by newly shifted bin.

i. Merge census and survey again. (This time there will be no non-matching or omitted bins. Total number of the final bins is 412.)

j. The merged file will have sum of census weights and sum of survey weights by bin. Correction factors for person and household can be calculated by dividing census weights by survey weights.

k. Keep bins and correction factors in a new file.

l. Merge this new file with original survey file by bin.

m. Newly corrected weights can be calculated by multiplying correction factors to the person and household weights.

5.3 Correction factors for final forecasting sample

There were 3268 respondents who completed the CATI interview in the SCE region. Only 1336 of these gave complete responses to all of the questions on the CATI and mailout survey, so we need to create correction factors which expand this forecasting sample back to the original 3268 respondents. Since we have all of the information in the CATI interview for all 3268 respondents, we use this information to fit a binomial logit model of the probability of being included in the final sample. The required correction factor is then given by the inverse of this probability.

The following table gives the results from maximum likelihood estimation of the logit model for the probability of being included in the final sample. Variables labelled "reg1" - "reg38" are dummy variables corresponding to the SCE districts in increasing numerical order. Positive coefficients imply that higher values of the variable are associated with higher probabilities of being included. The results show that smaller households with a small number of vehicles with older and more educated household heads are more likely to be included in our final sample.

Variable	Coefficient	Std. Error	T-Statistic
Household Size	-.0951606	.0348219	-2.733
Number of Drivers	.0894708	.0703844	1.271
Number of Vehicles	-.3828852	.0606167	-6.316
More than One Phone Line	.1520634	.1018953	1.492
Age	.0698713	.0135811	5.145
Age Squared	-.0006328	.0001471	-4.303
High School Diploma	-.4622199	.1269785	-3.640
Some College	-.1173398	.1066513	-1.100
reg1	-.9882091	.4012055	-2.463
reg2	-1.152056	.4358241	-2.643
reg3	-1.316482	.4333865	-3.038
reg4	-.9081138	.4292614	-2.116
reg5	-1.762673	.4283616	-4.115
reg6	-1.039232	.4054386	-2.563
reg7	-.7359529	.412531	-1.784
reg8	-1.19369	.4388503	-2.720
reg9	-1.281017	.4163197	-3.077
reg10	-.7208134	.4226489	-1.705
reg11	-1.051012	.4286923	-2.452
reg12	-1.111869	.4356718	-2.552
reg13	-.899027	.386861	-2.324
reg14	-1.100213	.3923384	-2.804
reg15	-1.186022	.4302771	-2.756
reg16	-1.661676	.4466177	-3.721
reg17	-1.17926	.4221077	-2.794
reg18	-1.313144	.4015746	-3.270
reg19	-.822961	.4075493	-2.019
reg20	-1.154075	.4217865	-2.736
reg21	-1.250886	.4110408	-3.043
reg22	-.367866	.4956199	-0.742
reg23	-1.221309	.391582	-3.119
reg24	-1.096373	.4168419	-2.630
reg25	-1.179104	.4355062	-2.707
reg26	-.7727779	.4349712	-1.777
reg27	-1.275703	.4028917	-3.166
reg28	-.854016	.4010691	-2.129
reg29	-1.316356	.4287773	-3.070
reg30	-1.267089	.3671601	-3.451
reg31	-1.304643	.3764132	-3.466
reg32	-1.013904	.407541	-2.488
reg33	-1.650926	.4342778	-3.802
reg34	-.5864893	.4492932	-1.305
reg35	-.800665	.4102427	-1.952
reg36	-1.471628	.4286362	-3.433
reg37	-1.316635	.4413008	-2.984
reg38	-1.288857	.3964517	-3.251

The inverse of the inclusion probabilities from the above binomial logit model were then used as correction factors for the weights coming from the procedure described in Section 5.2. The resulting weights were then rescaled so that the weighted forecasting sample exactly matched the population in each SCE district. This last reweighting reduced the correspondence between the weighted forecasting sample and the CPS.

6. HOUSEHOLD DEMOGRAPHIC MICROSIMULATION

6.1 Introduction

For vehicle choice and usage models, changes in household structure affect the outcome of the decision-making process. The original sample of households will certainly undergo a series of demographic changes during the period of the microsimulation. People will marry, divorce and separate, have children, and so forth. The purpose of this work is to develop demographic models that will be able to simulate these changes. They are dynamic in nature, and can be used by other microsimulation systems that model decision-making at the household level. They extend previous work in three ways: 1) by using continuous time hazard models, 2) by allowing for inter-dependencies across the various types of changes that a household may undergo, and 3) by including several explanatory variables.

These demographic models are interesting in their own right. They provide insight into different demographic patterns across socioeconomic groups. For example, I find that single black women are more likely to have a child out-of-wedlock than their white counterparts, all else being equal. And the differences are quite dramatic. They are also less likely to marry. White women with a higher education and income levels are less likely to have a child out-of wedlock and less likely to marry early than their white counterparts with less education and income. Differences also exist between first and second (or higher) marriages. Holding all other factors constants (such as age, race, income, education, and gender), individuals that have previously been married are likely to remarry sooner than individuals who have not married for the first time. These are just a few of the important differences. Several more are given in the estimation results in section 6.6.1, and illustrated by the survivor curves in section 6.6.1.1.

6.2 Changing Demographics in the United States

Households and families in the United States have undergone radical demographic changes over the past 30 to 40 years. In the 1950's, most people lived either with their parents or in college housing (supported by their parents) until they married. After marriage, women typically stayed home and raised children while their husbands went to work. Since then women have entered the labor force in large numbers. Women often experience a period of living on their own, working and independently supporting themselves. In 1965, 38.1% of all white women (regardless of their marital status) and 48.6% of nonwhite women were employed. By 1984 these numbers had grown to 53.3% and 55.% respectively (Blau and Ferber, 1986). The increases are even more dramatic when we break down the figures by marital status.

Table 6.1 Labor Force Participation Rates for Women		
	1966	1984
Never married	40.8	63.3
Married, husband present	35.4	52.8
Other (married at one time)	39.5	44.9
Married, husband absent	n.a.	61.1
Widowed	n.a.	20.4
Divorced	n.a.	74.3

Source: US. Department of Labor, Bureau of Labor Statistics, Special Labor Force Report, no. 2163, Table B-5, p. 16 and Bureau of Labor Statistics data reported in Bureau of National Affairs, Daily Labor Report, no. 145 (July 27, 1984), p. B-3.

At the same time that women were entering the labor force, families was undergoing tremendous change. Divorce was becoming more prevalent, increasing numbers of children were born to unwed mothers, many were postponing childbearing or choosing to forego the experience altogether, and cohabitation became more common. The following subsections present and discuss data on each of these trends.

6.2.1 Divorce and Separation

Compared to other countries, the United States has a high divorce rate. In 1976, 5 out of every 1000 people divorced. Other countries with relatively high divorce rates in 1976 included Australia at 4.3 (per 1000), USSR at 3.4, Sweden at 2.7, Denmark at 2.5, Canada at 2.2, Egypt at 2.2, and Finland at 2.1. The median time between first marriage and divorce was approximately 7 years in the United States, between divorce and remarriage was 3 years, and between remarriage and second divorce was 5 years for those that passed through each phase. (Glick and Norton, 1977). Since most data sources do not distinguish between divorce and separation, marital disruption will refer to an occurrence of either event.

Using life table estimates based on the 1985 Current Population Survey, Martin and Bumpass (1989) have projected that nearly two-thirds of all marriages in the late 1980's will end in divorce or separation. Others argue that the rate is lower. While the marital disruption rate rose dramatically from the late 1960's to 1970's, it has declined slightly during the 1980's so that roughly half of all current marriages with partners in their thirties are likely to dissolve (Glick, 1990).

The disruption rate for first marriages varies by several factors. As shown in Table 2, disruption rates within the first 5 years of marriage have increased across all categories from the period 1970-1974 to the period 1980-1985. Those who marry at earlier ages, have less education, have children before marriage, or are of African American ethnicity are more likely to experience marital disruption within the first 5 years. Even after accounting for differences in education, employment status, and premarital births, African Americans are still at higher risk.

Table 6.2: Proportion of First Marriages Disrupted Within 5 Years

	1970-74	1980-85
Age at Marriage		
14-19	0.23	0.31
20-22	0.14	0.26
23-29	0.11	0.15
30+	0.14	0.14
Education		
0-11 years	0.21	0.33
High school graduate	0.18	0.26
Some college (or beyond)	0.16	0.16
Kids before marriage		
0	0.17	0.21
1+	0.22	0.36
Race		
White	0.17	0.22
African American	0.24	0.36
Hispanic	0.15	0.24

Source: Martin, T. C. and L. L. Bumpass (1989), Table 1, p. 41.

Others find similar results (Glick and Norton, 1977; Spanier and Glick, 1981; Bennet, et al., 1989; Heaton and Jacobson, 1994). While the previous table did not separately identify education beyond college, Glick and Norton (1977) find that women with advanced education (17 or more years of schooling) are more likely than women with high school or college degrees to end their marriage.

Using two-state hazard models and data from the 1982 and 1988 National Surveys of Family Growth , Heaton and Jacobson (1994) find that about half of black marriages and a third of white marriages will have ended within 15 years of marriage. Age at marriage has a large negative effect on divorce for whites (e.g. marrying at a young age increases the chance of marital disruption), but virtually no effect for blacks. Racial differences persist even after accounting for differences in mother's education, religion, region, and age at marriage.

Along another vein, families with sons are more likely to stay together than families with daughters. "Sons reduce the risk of marital disruption by 9% more than do daughters. The differences hold across marriage cohorts, racial groups, and categories of mother's education."[1] Fathers may be more involved in the upbringing of their sons and therefore more committed to the marriage. In addition, families with only 1 child are more likely break apart as compared to families with 2 or more children (Morgan, et al., Figure 6.2, p. 118).

Marital disruption rates for second marriages follow similar patterns as first marriages. Table 3 shows that rates in general are higher for second marriages as compared to first marriages (see previous table). Nonetheless, similar factors are associated with higher rates of separation and divorce. Those who marry at earlier ages, are less educated, and are African American are more likely to divorce or separate within the first 5 years. Children before marriage (presumably from the previous marriage) are not as detrimental as children born out-of-wedlock before the first marriage.

Table 6.3: Proportion of Second Marriages Disrupted Within 5 Years

	1970-74	1980-85
Age at Second Marriage		
14-19	0.26	0.40
20-22	0.15	0.26
23-29	0.17	0.27
30+	0.13	0.14
Education		
0-11 years	0.17	0.36
High school graduate	0.17	0.26
Some college (or beyond)	0.20	0.22
Kids before marriage		
0	0.16	0.24
1+	0.18	0.28
Race		
White	0.18	0.26
African American	0.21	0.43
Hispanic	0.10	0.28

Source: Martin, T. C. and L. L. Bumpass (1989), Table 3, p. 45.

There are important areas of conflict within families which are not captured by simple measures such as race, educational background, age at marriage, and so forth. Some argue that marriages entail a balance of power between husband and wife. Events which tip the balance without the consent of both parties causes stress, potentially ending in divorce. One common event which may cause such stress occurs when the wife works outside the home. Husbands may view such an arrangement as either

[1] Morgan, S. P., Lye, D. N., and G. A. Condran (1988), p. 110.

positive or negative. Pyke (1994) argues that it may be viewed negatively when the "husband suffers low occupational status or chronic unemployment" in which case the husband will "devalue their wife's market work and view it as a burden (p. 75)." In addition, husbands may feel inadequate as the breadwinner in the family when the wife earns more (Stanley, et al. 1986). Men in general may find it difficult to relate to women who earn more (Bane1976), and empathy is often an important component of marital stability.

6.2.2 Marriage

The United States appears to have similar marriage rates when compared to other industrialized nations. In 1976, the marriage rate in the United States (defined as the number of marriages divided by the population in thousands) was 9.9. Corresponding rates were 8.1 in Australia, 8.7 in Canada, 10.0 in Egypt, 8.5 in Israel, 7.8 in Japan, and 10.1 in the USSR (Glick, 1977).

But singles today are more likely to postpone marriage. For example, in 1960 approximately 11% of women in their early thirties were unmarried. By 1987 that figure had increased 2 1/2 times to 27% . It is estimated that around 10% of young adults in the 1980's were likely to remain unmarried throughout their entire life (Glick, 1990). An even more dramatic comparison can be made between the turn of the century and current time.

> "In 1890, the median age of the wife at marriage was 22 years and the median age when her husband died was only 53 years. (There was) a fifty-fifty chance that the marriage would actually end before the last child left home. Because of the much longer length of life today, the corresponding age of the wife at dissolution of the marriage is now 68 years" (Glick, 1990).

Marriages used to last about 30 years, but today may last around 50 years assuming the couple does not divorce.

Within the United States, marriage rates vary by race. Heaton and Jacobson (1994) use two-state hazard models to examine first marriages. They find a dramatic difference between the marriage patterns of whites and African Americans. Fewer and fewer African Americans are choosing to marry. For example, 85% of white women will marry between the ages of 19 and 25, whereas less than 65% of African American women will be married by the age of 30. Bennett, et al. (1989) find similar differences, and argue that African American women are faced with a smaller pool of available mates. Several factors contribute to this fact including a smaller male to female birth ratio and higher death and incarceration rates for young black males.
The effects of education varies depending upon race. For whites, higher education (college and beyond) is negatively associated with the chance of ever marrying, but the

association is weak. On the other hand, higher education is positive and significantly associated with the probability of ever marrying for blacks (Bennet, et al., 1989). In general, increased education is associated with delayed marriage (Heaton and Jacobson, 1994).

Remarriage rates follow slightly different patterns. First, young adults who have divorced are much more likely to remarry when compared to similar cohorts who have yet to marry (Glick, 1990). This may to due to sampling bias since those who have already married probably have a higher propensity to marry in general.

Remarriage rates vary by education and income as shown by the figures for 1980 in Table 4. Increased education is associated with lower remarriage rates. The differences are more pronounced for women, with rates dropping from 61% for high school dropouts to 42% for women with graduate work. Both men and women with college degrees are less likely to remarry because they face a small pool of eligible mates as a result of the fact that college-educated people are more likely to have stable first marriages. Men with higher income and women with lower income are most likely to remarry (not necessarily to each other).

Table 6.4: Remarriage Rates in 1980

	Men	Women
Education		
Some high school or less	0.67	0.61
High school graduate	0.67	0.56
Some college	0.65	0.50
College graduate	0.61	0.44
Graduate work or degree	0.59	0.42
Income		
< 8,000	0.54	0.58
8,000 - 15,000	0.65	0.45
15,000 - 25,000	0.69	0.38
> 25,000	0.70	0.39

Source: Glick, P. C. and S. Lin (1987), Tables 2 and 3, pp. 168-171.

Bumpass, Sweet, and Martin (1990) use proportional hazard models and data from the 1980 and 1985 Current Population Survey to identify differences in remarriage rates across different groups. They find that remarriage rates are 26% lower for women aged 30-39 at separation (63% lower for those over 40) as compared to those under 25. Women who were 22 year or older when they married for the first time have a 38% lower remarriage rate than those who married at a younger age for the first time. Women with children from their previous marriage have a 25% lower remarriage rate than those without children, supporting Becker's notion that children are capital specific to the marriage in which they were conceived (Becker, et al., 1977). Finally remarriage is much less common among blacks. Controlling for all other variables, they have a remarriage rate that is only 1/4 of their white counterparts.

6.2.3 Fertility and Child Rearing

The most dramatic changes with respect to fertility is the increasing numbers of children born to unwed mothers. In 1960 only 5% of births occurred to unwed mothers (2% of white births and 22% of black births). By 1987 25% of all births occurred to unmarried women (17% of white births and 62% of black births).[2] Coupled with high divorce and separation rates, this leads to the fact that 23% of families with children less than 18 years old were headed by a single parent in 1988.[3]

A second change over the past 40 years has been the reduction in family size. During the baby boom of the mid-1950's, the number of children born per woman was about 3.8. From 1970 on, the number has remained stable at 1.8 per woman.[4]

Race is strongly associated with the timing of childbearing. Black males and females are likely to be sexually active at earlier ages and more accepting of sexual activity at younger ages when compared to white cohorts (Cherlin, 1992; Moore and Steif, 1991). As a result, there is 25 point difference in the percentage of those who have given birth by age 21 between blacks and whites.

White women from 2-parent families with highly educated mothers are more likely to deter child birth. For black women, growing up in a 2-parent family has a much less significant impact while their mother's education is important in delaying child birth.[5]

6.2.4 Cohabitation

While more and more singles are delaying marriage, the incidence of cohabitation has been increasing. Nearly one third of all young unmarried adults (ages 18-35) will live together (Thornton, 1988).

Cohabitation is a substitute for marriage, and often actually precedes marriage. In the mid-1980's, nearly half of all first marriages were preceded by a cohabitation period (Bumpass, 1990). As the following table shows, the percent married has dropped dramatically from 1970 to 1985. At the same time, cohabitation rates have been

---

[2] Glick, P. C. (1990), p. 139.
[3] Ibid, p. 141.
[4] Ibid, p. 140.
[5] Heaton and Jacobson (1994). The authors look only at the mother's educational influence on their daughter's fertility instead of the daughter's eduation.

increasing so that the percentage of adults in any form of union (marriage or cohabiting) has remained relatively stable.

Table 6.5: Marriage and Cohabitation Proportions

	1970 % Ever Married	1985 % Ever Married	1985 % Ever in Union
Before Age 20			
Total	27	14	23
Males	18	5	13
Females	35	22	33
Whites	28	15	25
Blacks	24	6	15
< High school	43	30	49
High school graduate	42	17	28
College	14	5	10
Before Age 25			
Total	72	55	69
Males	65	43	59
Females	79	66	78
Whites	76	58	71
Blacks	61	37	61
< High school	72	58	76
High school graduate	80	62	74
College	66	48	62

Source: Bumpass, Sweet and Cherlin (1991), Table 1, p 916.

As might be expected, cohabitation is a less stable than marriage. The median length of cohabitation is 1.5 years (Bumpass, 1990). As far as differences among groups, there is a higher reported rate of cohabitation among men than women. Blacks are also more likely to cohabitate. In general, Blacks are more likely to live in all types of 'non-traditional' family arrangements (Heaton and Jacobson, 1994).

6.3 Demographic Modeling Strategies

Merz (1991) identifies two main types of dynamic microsimulation: longitudinal and cross-sectional. Both can be carried out at the household or individual level. Demographers typically focus on individual transitions while maintaining household records in accordance with individual-level changes. When actually implementing this strategy, the analyst must provide algorithms to maintain the household records.
Longitudinal microsimulation creates synthetic microunits (in this case, households) and forecasts their life cycle. Synthetic microunits are randomly assigned core characteristics based on the joint distribution of these features from a population

sample. They are then assigned other non-core characteristics based on cross-tabulations of core and non-core features from the population sample. The synthetic households are weighted by the number of households that they represent in the population, and the weights are adjusted each period based on analyst-supplied forecasts of population growth and migration. The longitudinal technique has been used by Hensher et al. (1992) in their study of automobile demand and by Cowing and McFadden (1984) in their study of residential energy demand.

As an alternative method, cross-sectional microsimulation ages an actual population sample using empirically based transition probabilities. The number and composition of households changes from one period to the next, and the evolving demographics affect future behavior and transitions. Generally cross-sectional microsimulation is more difficult to apply than synthetic microunits, but the technique has several advantages. The analyst does not need to reweight the sample each period based on exogenous forecasts of population changes. In addition, a great deal of detail is maintained including the distributional impact of the policy under study. There are several ways to implement cross-sectional microsimulation including step-by-step submodules, multistate demography, and hazard models.

Step-by-step submodules have been used in several simulation programs including the Dynamic Simulation of Income Model (DYNASIM, 1976 or DYNASIM2, 1983 from the Urban Institute), MicroHaus (Gothenburg School of Economics, Sweden), and to a certain extent MIDAS (Goulis and Kitamura). Depending on the level of analysis, demographic changes occur as the individual or household sequentially passes through several submodules. For example, an individual becomes one year older, may marry, may divorce, may have a child, may become unemployed, and so on. Separate submodules are used for each demographic change, and the order the submodules is fixed. In other words, the individual first ages, then is subjected to the chance of marriage, then divorce and so forth through all the demographic processes. Often these systems do not account for interdependencies between transitions. As a result, changing the order of the submodules may change the outcome of the overall simulation.

Multistate demography (Land and Rogers, 1982; Rima and Van Wissen, 1987) determines the rate of movement between several analyst-defined states. A state describes the composition of and/or position within the household (e.g. head of a married couple, head of a family with 2 children, etc.). Movements from one state to another encompass several submodules from the previous step-by-step technique. The analyst must define several states which are typically at both the individual and household level. Assuming 'I' states at the individual level and 'H' states at the household level, the analyst determines movement rates across individual states (a matrix of I x I rates). These movement rates are then combined with fertility rates to determine movements across the household states (another H x H matrix). Some individual movements result in reclassification of the household, while others have no effect. The way in which individual movements affect households can become rather complex, and ground-rules must be determined in the design phase for handling actual household reclassifications (or movements across states). For example, a divorce may maintain the original household record after removing the husband and create a new household record containing only the husband.

The main disadvantage of multistate demography is the data requirements. If age is an important factor affecting individual movement rates, the analyst may define several age categories and the I x I matrix of movement rates must be determined for *each* age category. Extending the breakdown to include race and employment status would be unmanageable. It is unlikely that data would exist to fill in all the cells of the matrices of movement rates. If race and employment status are important explanatory variables of the those movement rates, simpler multistate demographic models which exclude those variables would produce suspect results. Nonetheless, multistate demography has theoretical grounding in the field of demographics and may be more consistent than tacking together several submodules.

An alternative to multistate demography uses hazard models. Hazard models measure the time until an individual or household undergoes some demographic change. In comparison to multistate demography, a hazard model may include several variables as determinants of then the change will occur instead of creating several transition matrices for each value of the variable. Therefore characteristics such as gender, ethnicity and income can easily be included in addition just age. For this reason I have chosen to use hazard models.

Previous work in this area has used discrete time hazard models (Davies, 1992). Davies also assumes that movement from the initial type of household to another is independent from the chance of movement from that initial type to any other household. This can be a very restrictive assumption, especially when important explanatory variables are excluded from the specification and estimation of the models.

Our models will extend previous work in two ways: 1) by using continuous time hazard models and 2) by allowing for inter-dependencies across the movements from one household type to the various other possible types. The advantage of continuos time hazard models is that the hazard rate (defined as the instantaneous probability of movement from one household type to another given that the household has not made such a movement yet) can vary over any time interval. With discrete choice models, the hazard rate is assumed to be constant over the discrete time intervals. I will also allow for interdependencies between movements from the one type of household and other forms (referred to as unobserved heterogeneity below). In most cases, the movements rates are found to be independent.

6.4 Individual or Household Level Models?

What exactly is meant by the term 'household'? Households include individuals who live alone, families, cohabiting adults with or without children, and other extended families (for example, older parents who have moved in with their adult son or daughter's family). When defining a household, the litmus test is whether the group engages in shared consumption. Ermisch (1988, page 23) provides an economic definition of a household as "a unit which combines the time of its members and purchased goods and services in the production of outputs, at least some of which are shared among its members." The Census Bureau considers a group of individuals to be a household if they live in the same dwelling and share meals, and the alternative-fuel vehicle survey has adopted a similar definition (with the inclusion of sharing household expenses). Roommates would not typically be considered a household unit.

A longitudinal definition of a household can be rather ambiguous. What happens to a household when a couple divorces? Which person is a continuation of the original household? Duncan and Hill (1985) persuasively argue that simply restricting the analysis to households that remain intact will lead to biased results. For this reason, they argue that the analysis should be done at the individual level with the unit of measurement at the household level. In other words, individuals are described in terms of the type of household to which they belong. Individuals move from one type of household to another, and the households must be maintained in accordance with those moves.

While this modeling strategy makes sense, it implies an additional level of accounting (e.g. programming for the microsimulation) to maintain the households. I have instead chosen to directly model at the household level. Households will be defined in terms of the head of the household. The head is the man in a couple, the single parent, or the single individual living alone. Therefore when a household splits into two, the portion containing the original household head is the continuation of the original household. The other portion is considered a new household. They are not dropped from the analysis. If they were dropped or excluded, the estimation results would clearly be biased. Duncan and Hill's modeling strategy is a natural extension to the results in this paper, and I plan to estimate such a model in the future and compare the final simulation results between both techniques.

6.5 Hazard Models

Hazard (or duration) models are used to model the time until an event or transition occurs. Most of the work in this field has been done by researchers in medicine and industrial engineering. Typical applications include models of drug effectiveness where the event of interest is curing the disease or death, and studies of machine reliability. Economists have applied the techniques to the study of unemployment duration (Meyer, 1990; Lancaster, 1979; Flinn and Heckman, 1982). More recently, duration models have been used in transportation analysis to study the timing of automobile purchases (Hensher and Mannering, 1994; Hensher, 1994; Jong, 1993) and accidents (Chang and Jovanis, 1990). They have also been applied to demographic processes

(Heaton and Jacobson, 1994). The literature contains quite a bit of terminology, and the same concepts are often referred to differently in each field. As a point of reference, we will define the basic concepts as referred to in this paper.

A "state" describes the household's (or individual's) current status. For example, states might be single, married with 1 child, employed, unemployed, and so forth. The movement from one state to another is called a "transition" (or exit). The terms "spell" and "episode" are used to mean the total amount of time spent in a specific state before a transition occurs. If the household never makes a transition during the period of observation, the spell is said to be "right-censored". If the household was in the current state before the period of observation (and it is unknown when they entered their current state), then the spell is "left-censored".

The simplest hazard models describe a situation of only two states and one episode. A transition from the first state to the second only occurs once. The second state is referred to as an absorbing state such as death. More complex models include multiple states and multiple spells. Competing risk hazard models describe the situation where an spell can end in many different ways.

In this paper we will model demographic transitions at the household level using the nine possible states shown in Table 6. Roommates are not considered a household type: instead, each person would be considered a single adult.

Table 6.6: Household States

S	Single adult
C	Couple (including cohabiting adults together at least 1 year)
C1	Couple with 1 child
C2	Couple with 2 children
C3+	Couple with 3 or more children
S1	Single adult with 1 child
S2	Single adult with 2 children
S3+	Single adult with 3 or more children
O	Other households (extended families and first-year cohabiting adults)

The models will be estimated using the Panel Study of Income Dynamics (PSID), and we have used that data as a guide for determining how to break down the household types. Just under 5% of the PSID sample exists in the 'Other' state in any given year. Typical households described by this state include parents living with their children's family, adult brothers and sisters (and sometimes their children) living together, and first-year cohabiting adults. The following table shows the breakdown of other households. The percentages for 'Single' are for households that would have been coded as 'Single' or 'Single with children' if the additional household members were excluded. Likewise, the percentages for 'Couple' are for households that would have been coded as 'Couple' or 'Couple with children' if the additional household members were excluded.

Table 6.7: Breakdown of 'Other' Household Types
(Proportion by type if additional household members were excluded)

Additional household members:	'Single'	'Couple'	All Types
Brother or Sister	19.7	18.1	19.1
Parents	21.6	37.1	27.6
Spouse of child	5.8	2.4	4.5
First-year cohabitator	21.7	0	13.2
Other relative	11.4	30.1	18.7
Other non-relative	19.7	12.3	16.8

Based on PSID data from 1979 to 1980 broken down into 6-month intervals. 'Single' refers to households that would have been coded as single or single with children if the additional household members had been excluded. 'Couple' refers to households that would have been coded as couple or couple with children if the additional household members had been excluded.

Although additional states could be added for each such subtype, I have combined them all because there are few households of each subtype. Including further breakdowns could be a future extension to this work.

The PSID makes the following distinction for cohabiting adults: after one year, unmarried couples are treated as though they were married. According to Bumpass (1990), the average length of cohabitation is relatively short (the median is 1.5 years), and almost 50% of first marriages in the mid-1980's were preceded by a cohabitation period. It seems plausible that after one year most cohabiting couples are in a stable, "marriage-like" relationship, or at least well on their way to being legally married. Therefore, I have maintained the PSID convention of separating cohabiting adults, and including longer-term relationships (over 1 year) in the 'Couple' category. Recent cohabitators are included in the 'Other' category.
Additional states could also be added for larger families. The grouping of household with 3 or more children could be split into finer categories (e.g. 3 children, 4 children, 5+ children). Since there are few households in each finer categories, I have chosen to group them together.

Death can be included in several ways. 'Single' households may die by dropping them from the sample of households. Simple death tables tabulated by age can be used to simulate this event. In addition, transitions from 'couple' to 'single' may also be caused by death of the spouse. Simple death tables again can be used to determine whether to create the splitoff household caused by the divorce, or to assume that the spouse died.

Given these nine categories, households move from one type to another over the lifetime of household members. These movements will modeled with hazard models.

6.5.1 Basic Two-State Hazard Models

Before describing the more complex competing risk models, I will begin with some of the basic concepts for a simple single-episode hazard model with only two states.[6] T is defined as a continuous non-negative random variable. T = 0 when the household enters a given state, and represents the amount of time spent in that state (also referred to as the duration of stay).

The probability that the household exits the current state during some small time period (*dt*) given that it has been in current state for an amount of time equal to *t* is represented by $P(t \leq \mathrm{T} < t + dt \mid \mathrm{T} \geq t)$. Dividing this probability by *dt* gives the average probability of leaving per unit of time conditioned upon remaining in the state until *t*. As *dt* goes to zero, we are left with the hazard function, *h(t)* given in 1.1, which represents the instantaneous rate of leaving at t conditioned upon remaining in the state until t.

$$h(t) = \lim_{dt \to 0} \frac{P(t \leq \mathrm{T} \leq t + dt \mid \mathrm{T} \geq t)}{dt} \tag{1.1}$$

The hazard function can also be expressed in terms of the distribution and density functions of T, $F(t)$ and $f(t)$ respectively. $F(t)$ is defined as P(T < *t*) as opposed to the usual P(T ≤ *t*). In addition, $F(0) = 0$ and $f(t) = \frac{d}{dt}F(t)$. This representation of the hazard function is shown in equation 1.2.[7]

$$h(t) = \frac{f(t)}{1 - F(t)} \tag{1.2}$$

---

[6] Most of this section is based on Lancaster (1990), Chapter 2, Sections 2.1-2.3.

[7] Equation 1.2 can be derived by using the law conditional probability. $P(t \leq \mathrm{T} \leq t + dt \mid \mathrm{T} \geq t)$ = $P(t \leq \mathrm{T} \leq t + dt, \mathrm{T} \geq t) / P(\mathrm{T} \geq t)$ which equals $P(t \leq \mathrm{T} \leq t + dt) / P(\mathrm{T} \geq t)$ since T ≥ *t* is a subset of $(t \leq \mathrm{T} \leq t + dt)$. Using the distribution and density functions, $P(t \leq \mathrm{T} \leq t + dt) / P(\mathrm{T} \geq t)$ = $F(t + dt) - F(t) / \; - F(t)$. Finally dividing by *dt* as *dt* goes to zero gives $h(t) = \lim_{dt \to 0} (F(t + dt) - F(t)) / (1 - F(t))\, dt = F'(t) / (1 - F(t))$. The hazard function, $h(t) = f(t) / (1 - F(t))$.

Again, the hazard function is defined as the conditional density of the transition out of a state given the length of time already spent in that state.

The survivor function is defined as the probability that the household remains in the current state at least until time t:

$$S(t) = 1 - F(t) = \frac{f(t)}{h(t)} \tag{2}$$

Based on these equations, a given F() determines the hazard and survival functions. Additionally if we know the hazard function, we can determine the distribution and density functions for the failure time since the distribution function is simply the solution to the differential equation in 1.1.[8]

$$F(t) = 1 - \exp\left(- \int_0^t h(u)\, du \right) \tag{3}$$

$$f(t) = h(t) \exp\left(- \int_0^t h(u)\, du \right)$$

What does the hazard function look like and how should it be specified? There are two approaches, namely parametric and nonparametric specification. I will focus on the parametric specification. The appropriate functional form varies across applications, and is related to the concept of duration dependence. Duration dependence describes how the failure rate changes (or remains the same) as time passes. In the simplest case, the failure rate is constant implying that the probability of exit is independent of the length of the duration. Typically this case is modeled by defining f(t) as an exponential distribution so that $f(t) = \lambda\exp(-\lambda t)$ and $h(t) = \lambda$. In practical applications, a constant hazard rate is often too limiting. Instead positive or negative duration dependence may be more appropriate. Positive duration dependence exists when households are more likely to exit their current state as time passes, $\partial h/\partial t > 0$. Negative duration dependence exists when households are less likely to exit their current state as time passes, $\partial h/\partial t < 0$. Several functional forms for the hazard function

---

8 The solution to the differential equation in 1.1 is derived as follows. First note that $f(t) = -\frac{d}{dt}[1 - F(t)]$. The hazard function can be rewritten as $h(t) = -\frac{d}{dt}[1 - F(t)] \cdot \frac{1}{1 - F(t)}$. Integrating both sides gives

$$\int_0^t h(u)du = - \int_0^t d[1 - F(u)] \cdot \frac{1}{1 - F(u)}$$, which evaluates to

$\int_0^t h(u)du = -\ln[1 - F(u)] \,\Big|_0^t = -\ln[1 - F(t)]$ since F(0)=0. Rearranging leads to the first line of equation 3, $F(t) = 1 - \exp(-\int_0^t h(u)du)$.

capture either positive or negative duration dependence such as the Weibull or Gompertz models.

Davies (1992) shows that the duration dependence changes for many demographic transitions. For example, a single person might have positive duration dependence in their twenties and thirties, constant duration dependence for a period, and then negative duration dependence (e.g. past a certain point, if they are single, they will most likely never marry or remarry). In the early twenties, the person might be postponing marriage because of educational considerations but as they reach late twenties and early thirties they might want to marry in order to have children. As the "biological clock" ticks away, women experience positive duration dependence. Past a certain point, they are no longer marrying to have children, at which point constant duration dependence might kick in. Finally, the pool of available mates becomes smaller and smaller so that she may experience negative duration dependence as she ages. This reasoning is meant only to be suggestive. In such cases, the hazard function can be modeled as following a quadratic or cubic form.[9]

Flinn and Heckman (1982) suggest the following flexible parametric function form for the hazard function. The exponential form guarantees that the hazard function is nonnegative.

$$h(t) = \exp(\gamma_0 + \sum_{k=1}^{K} \gamma_k \left[\frac{t^{\lambda_k} - 1}{\lambda_k} \right]) \tag{4}$$

The term within brackets is a Box-Cox transformation of time. As λ_k approaches 0, this transformation approaches ln(t). The values for K and λ_k determine the specific functional form as given in Table 8. Constant, increasing, decreasing and varying duration dependence are all included in the flexible form depending upon the specific parameters values for K, λ_ks, and γ_ks.

[9] The software package (CTM) which I have used allows for exponential, Weibull, Gompertz, quadratic and cubic forms of the hazard function. Other potential options include log-logistic and log-normal forms. These are not currently available in CTM.

Table 8: Hazard Functions from the Flexible Parametric Form		
Exponential (K=0)	$h(t)=\delta, \quad \delta=\exp(\gamma_0)$	(4.1)
Weibull (K=1, λ_1=0)	$h(t)=\delta\, t^{\gamma_1}, \quad \delta=\exp(\gamma_0)$	(4.2)
Gompertz (K=1, λ_1=1)	$h(t)=\delta\, \exp(\gamma_1 t), \quad \delta=\exp(\gamma_0-\gamma_1)$	(4.3)
Quadratic (K=2, λ_1=1, λ_2=2)	$h(t)=\delta\, \exp(\gamma_1 t+\frac{\gamma_2}{2}t^2), \quad \delta=\exp(\gamma_0-\gamma_1-\frac{\gamma_2}{2})$	(4.4)
Cubic (K=3, λ_1=1λ_2=2λ_3=3)	$h(t)=\delta\, \exp(\gamma_1 t+\frac{\gamma_2}{2}t^2+\frac{\gamma_3}{3}t^3), \quad \delta=\exp(\gamma_0-\gamma_1-\frac{\gamma_2}{2}-\frac{\gamma_3}{3})$	(4.5)

Household characteristics such as income, ethnicity, and age of the household members are likely to influence the rate of transitions. These characteristics are referred to as covariates, of which there are four major types. The first type do not vary over time, such as race, gender, and indicators of previous demographic status. These are referred to as time-invariant. The second type may vary over time, but their complete path is known before hand. For example, age is deterministic once the age at entry to the current state is known. These are referred to as "defined external covariates" (Kalbfleisch and Prentice, 1980) and can also be treated as time-invariant. The third type varies over time, but has an unknown future path. If we can determine the value of the covariate based on past history of the covariate, and the covariate can be described by a process that is unrelated to the duration of stay in current state, then it is referred to as an "exogenous" covariate (Lancaster, 1990). All of these three types can easily be incorporated into the hazard function as shown in equation 5. I assume that the covariates in this study are of these three types.

$$h(t,x)=\exp(\gamma_0+x_t\beta+\sum_{k=1}^{K}\gamma_k\left[\frac{t^{\lambda_k}-1}{\lambda_k}\right]) \tag{5}$$

x_t is a vector of exogenous covariates and β is an additional vector of parameters to be estimated. All of the five specific functional forms in Table 8 can be expanded to include covariates by replacingγ_0 by $\gamma_0+x_t\beta$.

The final covariate type is referred to as "endogenous" (Lancaster, 1990). In this case, the fact that the current state has not been left by time *t* + *dt* helps predict the covariate value from time *t* to *t* + *dt*. Endogenous covariates complicate the models, and "raise some rather subtle issues not all of which have been fully clarified in the literature" (Lancaster, 1990, page 23). I will assume that all covariate used in the models are not endogenous.

The previous models assume that the hazard function and survival distribution are homogeneous over the population of households. This assumption will almost certainly be invalid when important explanatory variables are excluded from the model, or when the transition times or covariates are imprecisely measured. In either case, the problem can be corrected by including an unobserved heterogeneity term in the hazard function.

Given these basic concepts, the demographic application that I am interested in requires a more complex model. Households can transition among a number of states, not just two, and multiple episodes are observed over the lifetime. A typical household may begin as a single person, transition into a couple, then to a couple with a child or possibly several children, and finally end as a couple. The variations are numerous. The most common transitions will be modeled using competing risk hazard models.

6.5.2 Competing Risk Hazard Models

6.5.2.1 Multistate/Single Episode

Again I begin with the simpler case of single-episode models.[10] The hazard function and the distribution and density functions of T are still defined as before. But now, the household may leave the current state to several possible destinations which leads to conditional analogs to the hazard, distribution and density functions.

Assuming that the household begins in state i (i=1,2,..N), it can then move to one of N - 1 other states represented by j. Let D_j be a dummy variable that indicates whether state j was entered upon transition (e.g. D_j =1 if state j was entered, D_j = 0 otherwise). The "transition intensity" (sometimes referred to as a state-specific hazard function) represents the instantaneous rate of leaving state i to state j at time t conditioned upon remaining in state i until t. It is given by the following equation:

$$h_{ij}(t) = \lim_{dt \to 0} \frac{P(t \le T \le t + dt, D_j = 1 \mid T \ge t)}{dt} \qquad (6)$$

The usual hazard function represents the instantaneous rate of leaving state i to any given state (conditioned upon remaining in state i until t), and is simply the sum of the state-specific transition intensities.

$$h_i(t) = \sum_{j=1, j \ne i}^{N} h_{ij}(t) \qquad (7)$$

Another important concept is the marginal probability of a destination; in other words, the probability that when the household exits the current state, they move to state j. This marginal probability is represented by π_{ij}. Recall that the survival function S(*t*) = [1 - F(*t*)]. Therefore the marginal probability of a destination is given by:[11]

---

[10] Most of this section is based on Lancaster, 1990, Chapter 5 and Flinn and Heckman, 1992.

[11] This can be derived by noting that S(*t*) $h_{ij}(t)dt$ = P(survival to *t*) x P(exit to state j in the interval *t* + d*t* | survial to *t*). Integrating over all *t* gives the empirical counterpart of the fraction of households that ever transition to state j.

$$\pi_{ij} = \int_0^\infty S(u) h_{ij}(u) du \tag{8}$$

The sum of these marginal probabilities over all N-1 possible destination states equal 1, $\sum_{j=1, j \neq i}^{N} \pi_{ij} = $.

Finally, the conditional distribution of T (e.g. conditioned on transitioning to state j) given the starting state is i is represented by $F_{ij}(t)$. In other words, $F_{ij}(t)$ represents the probability that the household departs state i before time t given that when the departure occurs, it is to state j. Therefore $\pi_{ij} F_{ij}(t)$ is the probability that the household departs before time t *and* that they depart to state j. Finally the original unconditional distribution of T for the starting state i is given by the sum of this product over all possible j, $F_i(t) = \sum_{j=1, j \neq i}^{N} \pi_{ij} F_{ij}(t)$.

We only observe that the household left the initial state at some time t, and entered one of the j possible states. Using the dummy variables (D_j) and T, the likelihood function is given as follows:

$$\exp\left(- \int_0^t h_{ij}(u) du \right) \prod_{j=1, j \neq i}^{N} h_{ij}(t)^{D_j} \tag{9}$$

Another way of formulating this same model uses latent exit times from state i to the other N-1 possible states. This is what the terminology "competing risks" refers to. By assuming that these latent exit times are independent, the joint density of the those latent exit times is also given by equation 9 above. This assumption makes the estimation procedure much easier, but at may be unrealistic for some demographic transitions. For example, the transition from the state of couple with one child to the state of couple with two children may be related to the hazard rate for moving from the state of couple with one child to single with one child. In other words, the failure time for the first transition may be related to the latent failure time for second type of transition. Knowing that you are unlikely to divorce might influence your decision to have another child, just as knowing that you are on the verge of divorce affects your decision to have another child (Lillard, 1993). Unobserved heterogeneity may be used to address this issue by allowing for interdependencies among groups of transitions as described below.

Given these basic concepts and definitions, the flexible parametric form for the hazard function in equation 5 can be extended to deal with the multistate case.

$$h_{ij}(t,x) = \exp(\gamma_0 + x_t\beta_{ij} + \sum_{k=1}^{K}\gamma_{kij}\left[\frac{t^{\lambda_{kij}}-1}{\lambda_{kij}}\right]) \qquad (10)$$

This hazard function includes time-varying covariates and terms for duration dependence. An unobserved heterogeneity component, $_{ij}$, can also be added. Flinn and Heckman (1982) suggest adding it to the term within the exponent.

$$h_{ij}(t,x) = \exp(\gamma_0 + x_t\beta_{ij} + \sum_{k=1}^{K}\gamma_{kij}\left[\frac{t^{\lambda_{kij}}-1}{\lambda_{kij}}\right] + V_{ij}) \qquad (11)$$

Unobserved heterogeneity is important when important explanatory variables have been excluded from the model. In this situation, groups of transitions are inherently inter-related, and the unobserved heterogeneity captures the interdependence.

Unobserved heterogeneity term can be specified in several ways. For ease of estimation, Flinn and Heckman (1982) suggest simplifying the unobserved heterogeneity components by assuming that they are constant within spells but vary across spells, so that unobserved heterogeneity across spells is restricted to a one-factor error specification.

$$_{ij} = C_{ij}\, V \qquad (12)$$

The software which I have used to estimate the models (CTM) uses this form, and estimates the parameter C_{ij}. Some assumption must be made about the distribution of V. For example, V may be normally distributed with mean 0 and variance of 1, V~N(0,1). Therefore the variance of V_{ij} is allowed to change across spells. CTM allows several other assumptions about the unobserved term. V may be follow a lognormal, exponential, gamma, or even nonparametric distribution.

Unobserved heterogeneity may be used to the fact that the latent exit times are not truly independent. The independent competing risk model with unobserved heterogeneity can be estimated on subsets of the states (e.g. those states that are thought to be more closely related). A significant unobserved heterogeneity parameter (C_{ij}) implies that important variables have been excluded, and the transitions are inter-related.
The fullest hazard model is the multistate and multi-episode formulation. Multi-episodic data allows for variation in hazard functions depending on which episode is being modeled. For example, a second divorce may be different from the first.

6.5.2.2 Multistate/Multi-episode

Equations 6 through 12 can be expanded to a multi-episode model. Over the period of study, the household experiences several transitions across the possible states. Assume that the household is in state i for its m^{th} spell. It has been in this state for a length of time equal to t_m and entered the state at calendar time τ_m. The hazard function is given by:

$$h_{i_m j_m}(t_m, x) = \exp(g_0 + x_{t_m + t_m} b_{ij}^m + \sum_{k=1}^{K} g_{kij}\left[\frac{t_m^{\ l_{kij}} - 1}{l_{kij}}\right] + V_{ij}) \tag{13}$$

where the coefficient vector β_{ij}^m can vary across episodes.

The unobserved heterogeneity can again be simplified by assuming that $_{ij}$ = C_{ij}V where several distributional assumptions can be made about V such as V ~ N(0,1). The derivation of the likelihood function for this model can be found in Flinn and Heckman (1982, 1983).

6.6 The Panel Study of Income Dynamics

As was mentioned previously, the models will be estimated using data from the Panel Study of Income Dynamics (or PSID). The PSID began in 1968, and has surveyed the same sample of households every year. When children move out or families split apart, every effort was made to track both the original and the new splitoff household. Response rates are quite high, and range from 97 to 98.5% each year.

The PSID was initially conducted to study poverty. The full data set contains a large over sampling of low-income households, and can be broken into two subsamples: (1) the Social Research Council (SRC) sample of approximately 3000 household which was randomly drawn from the population of the 48 contiguous states, and (2) the non-random SEO sample of approximately 2000 low-income households selected from respondents to the Survey of Economic Opportunity. For the purposes of this demographic modeling, I have only used the random SRC sample for the years 1980 through 1989. The SRC subsample contained between 3500 and 4000 households each year. New households enter the sample as they split off from existing households, and others leave the sample because of nonresponse. Overall, the sample contains information for approximately 4600 households. In some cases (such as splitoff households), the data is available for only a subset of the full ten years.
The PSID collects information about movements into and out of households each month. I have set the observation period at every 6 months, based on the assumption that only 1 transition is likely to occur in that time frame. In other words, I have determined household type for each household in the PSID every 6 months. For the 4600 households, almost 10,000 spells occurred in the 10 year period. Table 9 shows a 9x9 transition matrix illustrating all the possible transitions and frequency counts. The headings on the left side of each row indicate the beginning state, and the headings across the top of the columns indicate the ending state. Highlighted cells indicate transitions that were estimated. For example, the transition from 'Couple with 2 children' to 'Single' is included, but the transition from 'Single' to 'Couple with 2 children' is not. I do not plan to estimate several cells, particularly those with less than 50

observations. Transitions with over 50 observations which currently are excluded may be included in future estimations (mainly household types described as 'Other').

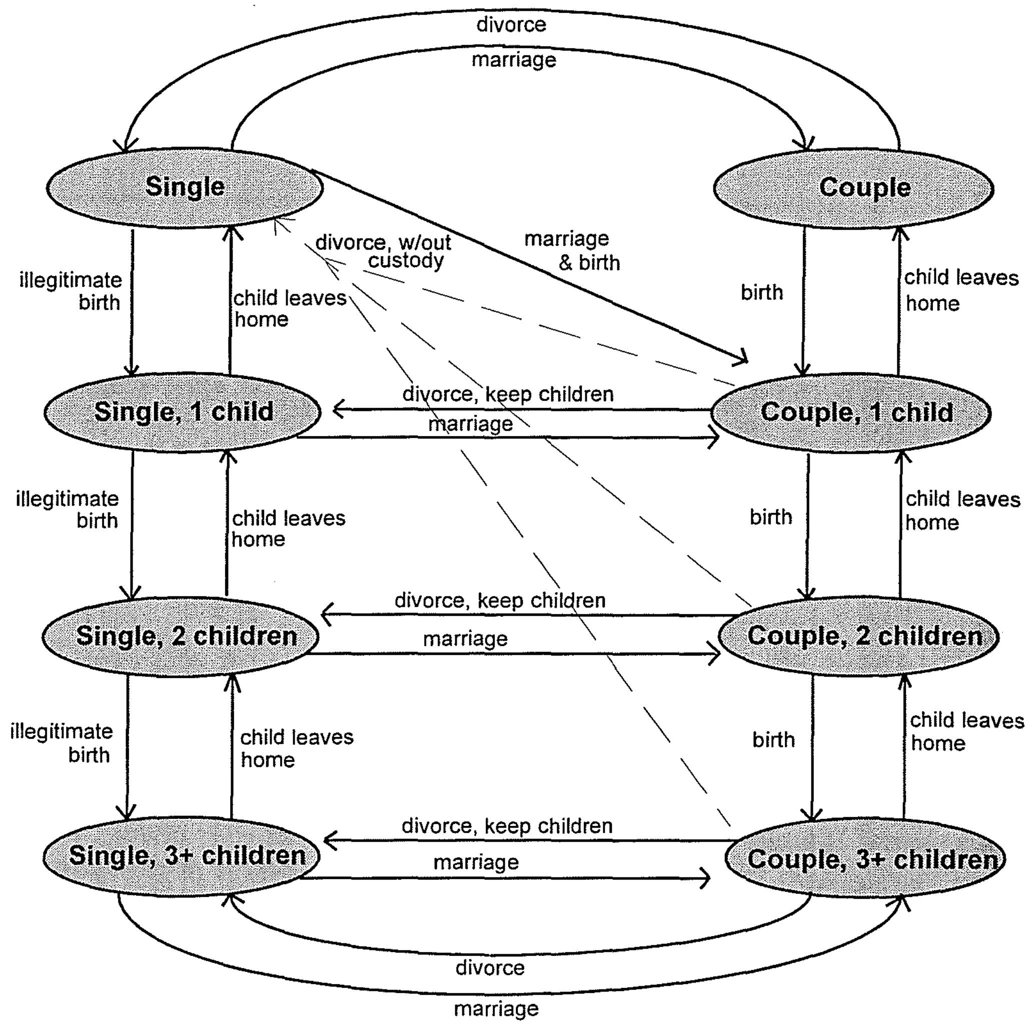
Table 6.9: Household Transition Matrix and Frequency Counts

States	S	C	C1	C2	C3+	S1	S2	S3+	O
S	1192	428	67	31	15	119	8	4	187
C	363	1063	634	17	5	5	0	0	59
C1	78	311	591	671	17	115	8	0	49
C2	64	47	303	674	295	18	89	3	50
C3+	23	6	31	187	412	8	4	52	39
S1	119	12	95	13	8	185	55	0	35
S2	18	2	9	52	9	64	110	34	22
S3+	4	3	3	5	24	9	44	59	9
O	156	116	54	46	38	23	16	10	193

It is important to note that just because a transition is not directly modeled, households may still make those moves, just not in a 6 month time frame. For example, a 'Single' household may move to 'Couple with 2 children' by first transitioning to 'Couple with 1 child' and then moving from 'Couple with 1 child' to 'Couple with 2 children'. At a minimum, this path would take 1 year.

Figure 1 shows the possible movements among all household states except 'Other'. Movements labeled "marriage" also include long-term cohabitation arrangements. Couples with children can dissolve in two ways, either the household head has custody of the children (the solid lines) or the he leaves the relationship without custody (the dashed lines). In the first case, the remaining adult forms a new single household. In the second case, the remaining adult forms a new household consisting of a single with children.

Figure 6.1: Flow Chart of Demographic Changes



The PSID is rich with possible explanatory variables, many of which have been identified in the demographic literature as being significant. But I must be able to forecast all explanatory variables. Therefore I have limited the covariates to race, education level, employment status, age and gender of adult household members, age of children, and household income. For transitions beginning from a 'couple' state, I have included information about both the husband and wife.

6.6.1 Race

In 1980, the PSID random sample contains a breakdown of 88.4% white households, 9.4% black households, and 2.2% of other households (mainly Asian, with some Native American). In 1989 the breakdown has shifted slightly to include 90.3% white households, 8.4% black households, and 1.3% other races. Since blacks are more

likely to postpone marriage, bear children out-of-wedlock, and experience martial disruption, and live in extended families, I expect that a large percentage of black households would be classified as 'single with children' or 'other'.
Table 10 shows racial breakdown for each household type for the years 1980 and 1988.

Table 6.10: Race by Household Type

1980	All	C	C1	C2	C3+	S	S1	S2	S3+	O
% White	88.4	93.9	87.9	91.4	86.6	90.1	74.1	73.3	57.1	74.5
% Black	9.4	5.3	9.2	5.5	10.8	8.6	23.0	22.1	40.8	19.8
% Other	2.2	0.8	2.9	3.1	2.6	1.3	2.9	4.6	2.1	5.7
Total households	3579	900	555	514	344	839	135	86	49	157
1989										
% White	90.3	94.4	93.5	93.9	88.7	90.2	76.2	68.5	44.2	87.0
% Black	8.4	4.6	4.8	5.1	9.4	8.6	23.1	29.2	51.9	11.7
% Other	1.3	1.0	1.7	1.0	1.9	1.2	0.7	2.3	3.9	1.3
Total households	3802	945	525	607	371	904	147	89	52	162

S - Single, C - Couple, C1 - Couple with 1 child, C2 - Couple with 2 children, C3 +- Couple with 3 or more children, S1 - Single with 1 child, S2 - Single with 2 children, S3+ - Single with 3 or more children, O - Other household types

As expected, more blacks are categorized as 'single with 1 child', 'single with 2 children', and 'single with 3+ children' than the black proportion of the population as a whole. In the most extreme case, 50% of all PSID households consisting of single of 3 or more children are headed by a black person in 1989. Only 44.2% of such households are headed by a white person in that same year while whites make up 90.3% of the total sample.

Other the other hand, whites make up a larger relative percentage of couples, couples with 1 child, and couples with 2 children. For example, 94.4% of couples are headed by a white person in 1989 while whites make up 90.3% of the total sample in that year.

6.6.2 Education

The demographic literature also provided insight into the relationship between education and household transitions. Those with less education are more likely to divorce (with the exception of women with advanced graduate work), are more likely to remarry for women and less likely to remarry for men, and are less likely to delay child birth. Table 11 shows the breakdown of education for each household type and for the sample as a whole.
Overall, education levels were rising from 1980 to 1989. In 1980 very few people had continued past college. By 1989, 10.6% had continued on to graduate work. Likewise

the percentages for those without a high school diploma had fallen sharply from 25.3% in 1980 to 16.2% in 1989.

For all of the single with children categories, the percentage of household heads with low education levels is higher than for the population as a whole. For example, 37.3% of households which are single with 3 or more children are headed by a person who has not graduated from high school. The corresponding percentage for the population as whole is 16.2%.

Men with a higher level of education (college or graduate work) are more likely to be married with children. In 1989 37.9% of couples with 1 child, 40.4% of couples with 2 children, and 35.2% of couples with 3 or more children contain a husband with some college education. These percentages are higher than the population as a whole which contains 34.4% of people with some college education.

In general, men within couples appear to have more schooling than their wives at the higher categories of education. On the other hand, women in couples are more likely to have finished high school than their husbands at the lower categories of education.

6.6.3 Other Covariates

I have included several other covariates besides race and education. These covariates include household income, employment status, age, gender of singles, number of children in various age categories, and some indicators of previous states occupied. Covariates were included only when appropriate (e.g. number of children was not included for transitions out of the 'single', 'couple', or 'other' states).

Household income for the years 1979 through 1989 was converted to 1989 dollars, and the income for the entire year was divided by 2 to represent 6-month earnings. But using income for the current year that the transition might have occurred would lead to biased results. For example, a single person who marries would necessarily have higher income during that year if his or her spouse worked. So instead of using income from the current year, I have used lagged income from the prior year. Employment status was coded as either employed, unemployed, or out of the workforce (which includes homemakers). The age categories for children included the number of children less than 6 years old, number of children between 6 and 18 years old, and number of children over 18 years.

Table 11: Education by Household Type

		C		C1		C2		C3+						
1980 %	All	Husb	Wife	Husb	Wife	Husb	Wife	Husb	Wife	S	S1	S2	S3+	O
Some HS	25.3	30.3	23.9	19.4	17.2	20.0	19.3	23.5	24.8	31.4	41.5	41.2	58.3	32.1
HS grad	40.6	33.3	43.8	41.7	54.5	37.8	51.4	36.7	51.6	31.1	37.8	38.8	31.3	34.6
College	33.4	35.8	31.6	38.1	29.8	41.7	28.8	39.3	22.7	36.8	20.7	20.0	10.4	33.3
Graduate	0.7	0.6	0.7	0.8	1.5	0.5	0.5	0.5	0.9	0.7	0.0	0.0	0.0	0.0
1989 %														
Some HS	16.2	19.3	15.3	15.2	12.0	10.1	10.1	12.5	11.3	24.1	20.4	28.1	37.3	23.5
HS grad	38.8	34.9	40.0	38.1	47.6	36.1	44.1	40.1	46.1	31.3	44.2	36.0	33.3	37.7
College	34.4	31.5	34.9	37.9	32.6	40.4	35.6	35.2	35.0	33.9	28.6	25.8	25.5	28.4
Graduate	10.6	14.3	9.8	8.8	7.8	13.4	10.2	12.2	7.6	10.8	6.8	10.1	3.9	10.5

S - Single, C - Couple, C1 - Couple with 1 child, C2 - Couple with 2 children, C3 +- Couple with 3 or more children, S1 - Single with 1 child, S2 - Single with 2 children, S3+ - Single with 3 or more children, O - Other household types. For couple household types, ages are shown for both husbands (husb) and wifes.

Multi-episode modeling was used by included indicators of previous states occupied. A flag for previously married was set if the household was in any of the couple states during some past period of observation. Another flag was set if the household previously existed in a state of single with children. I did not use data or information prior to 1979 in setting these flags.

6.7 Estimation Results

6.7.1 Without Unobserved Heterogeneity

I have estimated models for each highlighted transition in Table 9 assuming independent Weibull hazard functions without unobserved heterogeneity. The results are summarized in the following tables. Tables 12.1 though 12.5 include transitions which begin in either the 'single' or 'other' states (e.g. states S, S1,S2,S3+, and O). Tables 12.6 though 12.9 show the results for transitions which begin in 'couple' states (e.g. C, C1, C2, and C3+). All tables include general covariates describing the household such as income, while tables 12.1-12.5 include covariates describing the single household head and tables 12.6-12.9 include covariates describing the both adult members in the household referred to as husband and wife.

For dummy variables, a negative coefficient implies that the group will remain in the current state longer than the reference group. In other words, it will take them longer to transition. Race is one such dummy variable. The coefficient for blacks in the model of transitions from 'Single' to 'Couple' is negative (-0.777 in Table 12.1 with a t-statistic of 2.79). This implies that blacks will remain in the single state longer than whites (the control group) before getting married. Blacks are likely to marry at a later age supporting the results of Heaton and Jacobson (1994).

A positive coefficient for a dummy variable implies that the group will transition faster than the reference group. Previous research finds that blacks are more likely to have a child out-of-wedlock, and at an earlier age than whites. My results support those findings. The coefficient for blacks in the the model of transititioning from 'Single' to 'Single with 1 child' is positive and significant (1.424 in Table 12.1 with a t-statistic of 6.52).

For continuous variables, a negative coefficient implies that as the variable increases, the household is likely to remain in the current state longer. Likewise, a positive coefficient implies that the household is likely to transition sooner at larger values of the variable. For example, at higher income levels (e.g. larger values for the log of income), singles are more likely to marry sooner (the coefficient is 0.196 in Table 12.1 with a significant t-statistic).

Table 6.12.1: Estimation Results for Independent Hazard Functions
Weibull Duration Dependence without Unobserved Heterogeneity

	Single → Couple		Single → Couple, 1 child		Single → Single, 1 child	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Constant intercept (γ_0)	-3.509	5.32	-6.857	3.03	-4.812	4.20
Gamma_1 (γ)	0.275	4.08	0.142	0.71	-0.087	0.73
Log (lagged household income)	0.196	3.29	0.079	0.54	0.125	0.94
Race (vs. white) [a]						
- Black	-0.777	2.79	-0.334	0.71	1.424	6.52
- Other	-0.286	1.68	-0.493	1.04	-0.146	0.40
Single's education (vs. < high school)						
- High school graduate	-0.054	0.28	-0.866	2.74	-0.261	1.05
- College	-0.019	0.10	-1.551	4.26	-0.755	2.81
- Graduate work or degree	0.111	0.48	-2.315	2.88	-1.303	2.56
Single's employment (vs. employed)						
- Unemployed	0.058	0.27	0.178	0.45	0.644	2.17
- Out of work force	0.166	0.78	-0.016	0.03	0.516	1.79
Age of single	-0.055	2.30	0.155	1.29	-0.030	0.96
(Age of single)^2	0.000	0.10	-0.003	1.51	0.000	0.23
Gender of single (female vs. male)	-0.200	2.01	-1.651	3.51	0.599	2.70
# of kids < 6 years old	----	----	----	----	----	----
# of kids between 6 & 18 years old	----	----	----	----	----	----
# of kids >= 18 years old	----	----	----	----	----	----
Previous marriage	0.369	3.16	0.926	3.51	----	----
Previously a single parent	----	----	----	----	----	----
- Log likelihood($\hat{\gamma}_0$, $\hat{\gamma}$, $\hat{\beta}$)	1770.81		359.07		636.61	
- Log likelihood($\hat{\gamma}_0$, $\hat{\gamma}$)	1974.82		431.68		696.71	

[a] Race is only reported for the head of the household. For single households, this poses no uncertainty. For couples, the head may be male or female. In --% of couples, the man has been coded as the head of the household.

Table 6.12.2: Estimation Results for Independent Hazard Functions

Weibull Duration Dependence without Unobserved Heterogeneity

	Single → Other		Single, 1 child → Single		Single, 1 child → Couple, 1 child	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Constant intercept (γ_0)	-4.919	6.54	-6.525	3.41	-0.240	0.13
Gamma_1 (γ)	0.007	0.08	0.165	1.32	0.259	1.59
Log (lagged household income)	0.181	2.94	0.028	0.17	0.065	0.63
Race (vs. white) [a]						
- Black	0.089	0.32	-0.106	0.41	-0.778	2.81
- Other	-0.161	0.63	-0.353	0.91	-0.053	0.15
Single's education (vs. < high school)						
- High school graduate	-0.264	1.15	-0.759	3.27	0.476	1.70
- College	-0.379	1.70	-0.741	2.90	0.680	1.96
- Graduate work or degree	-0.507	1.44	-0.740	1.69	0.887	1.72
Single's employment (vs. employed)						
- Unemployed	0.585	2.36	0.742	1.53	0.082	0.24
- Out of work force	0.377	1.38	-0.134	0.54	0.134	0.47
Age of single	-0.002	0.08	0.156	3.50	-0.047	0.45
(Age of single)^2	-0.000	1.37	-0.001	3.66	-0.001	0.86
Gender of single (female vs. male)	-0.172	1.15	-1.003	4.37	-1.201	4.08
# of kids < 6 years old	----	----	----	----	0.035	0.11
# of kids between 6 & 18 years old	----	----	----	----	----	----
# of kids >= 18 years old	----	----	0.828	2.92	----	----
Previous marriage	----	----	----	----	-0.200	1.01
Previously a single parent	----	----	----	----	----	----
- Log likelihood($\hat{\gamma}_0$, $\hat{\gamma}$, $\hat{\beta}$)	960.34		428.71		333.33	
- Log likelihood($\hat{\gamma}_0$, $\hat{\gamma}$)	1014.52		488.24		411.44	

[a] Race is only reported for the head of the household. For single households, this poses no uncertainty. For couples, the head may be male or female. In --% of couples, the man has been coded as the head of the household.

Table 6.12.3: Estimation Results for Independent Hazard Functions
Weibull Duration Dependence without Unobserved Heterogeneity

	Single, 1 child → Single, 2 children		Single, 2 children → Couple, 2 kids		Single, 2 children → Single, 1 child	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Constant intercept (γ_0)	-4.105	2.24	-7.152	1.64	-10.875	3.75
Gamma_1 (γ)	0.039	0.20	0.474	2.23	0.275	1.56
Log (lagged household income)	-0.139	1.09	0.487	1.97	0.162	0.73
Race (vs. white) [a]						
- Black	0.834	2.55	-1.486	2.17	-0.121	0.33
- Other	0.357	0.72	0.000	0.00	-0.307	0.68
Single's education (vs. < high school)						
- High school graduate	-0.094	0.25	0.247	0.71	0.286	0.70
- College	0.138	0.34	0.057	0.14	-0.211	0.44
- Graduate work or degree	-0.672	0.54	---	---	-0.654	0.86
Single's employment (vs. employed)						
- Unemployed	0.422	0.73	0.238	0.43	0.299	0.57
- Out of work force	0.556	1.54	0.477	1.27	-0.314	0.72
Age of single	0.077	1.27	0.063	0.26	0.225	2.54
(Age of single)^2	-0.001	1.25	-0.002	0.60	-0.002	2.54
Gender of single (female vs. male)	-0.506	1.12	-0.498	0.97	-0.134	0.31
# of kids < 6 years old	----	----	----	----	----	----
# of kids between 6 & 18 years old	----	----	-0.242	0.87	-0.178	0.39
# of kids >= 18 years old	-0.787	1.87	-0.647	0.99	0.697	1.40
Previous marriage	----	----	0.239	0.71	----	----
Previously a single parent	----	----	----	----	----	----
- Log likelihood($\hat{\gamma}_0$, $\hat{\gamma}$, $\hat{\beta}$)	256.27		202.30		231.97	
- Log likelihood($\hat{\gamma}_0$, $\hat{\gamma}$)	268.44		232.23		271.78	

[a] Race is only reported for the head of the household. For single households, this poses no uncertainty. For couples, the head may be male or female. In --% of couples, the man has been coded as the head of the household.

Table 6.12.4: Estimation Results for Independent Hazard Functions
Weibull Duration Dependence without Unobserved Heterogeneity

	Single, 2 children → Single, 3+ kids		Single, 3+ kids → Single, 2 children		Other → Single	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Constant intercept (γ_0)	0.256	0.09	-4.092	1.15	1.999	2.55
Gamma_1 (γ)	0.159	0.62	0.116	0.52	0.363	3.29
Log (lagged household income)	-0.278	1.72	-0.044	0.15	-0.130	1.76
Race (vs. white) [a]						
- Black	1.446	2.71	0.061	0.14	-0.317	-1.27
- Other	0.973	1.43	0.184	0.30	0.272	1.18
Single's education (vs. < high school)						
- High school graduate	-0.053	0.09	-1.001	2.10	0.486	2.03
- College	0.275	0.50	-0.833	1.68	0.830	3.59
- Graduate work or degree	----	----	----	----	0.713	1.95
Single's employment (vs. employed)						
- Unemployed	-0.587	0.71	0.310	0.46	0.381	1.21
- Out of work force	0.258	0.49	0.330	0.73	0.262	1.06
Age of single	-0.080	0.54	0.210	1.97	-0.177	6.86
(Age of single)^2	0.000	0.17	-0.002	0.15	0.001	5.19
Gender of single (female vs. male)	----	----	----	----	0.182	1.13
# of kids < 6 years old	----	----	-1.233	3.69	----	----
# of kids between 6 & 18 years old	-0.456	1.07	-1.062	3.57	----	----
# of kids >= 18 years old	0.347	0.67	-0.562	1.96	----	----
Previous marriage	----	----	----	----	----	----
Previously a single parent	----	----	----	----	----	----
- Log likelihood($\hat{\gamma}_0$, $\hat{\gamma}$, $\hat{\beta}$)	134.87		156.17		534.99	
- Log likelihood($\hat{\gamma}_0$, $\hat{\gamma}$)	166.76		180.26		606.20	

[a] Race is only reported for the head of the household. For single households, this poses no uncertainty. For couples, the head may be male or female. In --% of couples, the man has been coded as the head of the household.

Table 6.12.5: Estimation Results for Independent Hazard Functions
Weibull Duration Dependence without Unobserved Heterogeneity

	Other → Couple		Other → Couple, 1 child	
	Estimate	t-stat	Estimate	t-stat
Constant intercept (γ_0)	1.422	1.51	-1.501	0.67
Gamma_1 (γ)	0.023	0.15	0.065	0.29
Log (lagged household income)	-0.174	2.09	-0.004	0.02
Race (vs. white) [a]				
- Black	-1.337	2.88	-1.160	1.86
- Other	-0.165	0.59	-0.359	0.78
Single's education (vs. < high school)				
- High school graduate	0.405	1.54	-0.461	1.32
- College	0.423	1.52	-0.472	1.25
- Graduate work or degree	0.614	1.65	----	----
Single's employment (vs. employed)				
- Unemployed	0.026	0.07	-0.241	0.38
- Out of work force	-0.109	0.32	-0.612	1.00
Age of single	-0.113	3.17	-0.036	0.47
(Age of single)^2	0.001	2.31	-0.000	0.08
Gender of single (female vs. male)	-0.630	2.83	-0.686	1.74
# of kids < 6 years old	----	----	----	----
# of kids between 6 & 18 years old	----	----	----	----
# of kids >= 18 years old	----	----	----	----
Previous marriage	----	----	----	----
Previously a single parent	----	----	----	----
- Log likelihood($\hat{\gamma}_0$, $\hat{\gamma}$, $\hat{\beta}$)	442.22		247.86	
- Log likelihood($\hat{\gamma}_0$, $\hat{\gamma}$)	481.55		265.82	

[a] Race is only reported for the head of the household. For single households, this poses no uncertainty. For couples, the head may be male or female. In --% of couples, the man has been coded as the head of the household.

Table 6.12.6: Estimation Results for Independent Hazard Functions

Weibull Duration Dependence without Unobserved Heterogeneity

	Couple → Single		Couple → Couple, 1 child		Couple, 1 child → Single	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Constant intercept (γ_0)	0.799	1.03	-1.696	2.89	-0.776	0.31
Gamma_1 (γ)	0.274	3.47	0.334	6.24	0.374	1.97
Log (lagged household income)	-0.169	2.03	0.085	1.37	-0.068	0.34
Race (vs. white) [a]						
- Black	0.214	0.90	0.100	0.53	0.267	0.53
- Other	-0.250	1.26	-0.168	1.09	-0.648	1.26
Husband's education [b]						
- High school graduate	-0.203	1.27	-0.156	1.18	-0.575	1.69
- College	-0.323	1.83	0.037	0.26	-1.088	2.64
- Graduate work or degree	-0.236	1.11	-0.137	0.68	-0.987	1.46
Husband's employment						
- Unemployed	0.257	0.91	-0.009	0.05	1.499	4.19
- Out of work force	0.771	3.87	-0.781	3.53	1.090	2.11
Age of husband	-0.069	2.13	-0.099	2.95	0.026	0.23
(Age of husband)^2	0.001	1.87	0.001	1.62	-0.001	0.72
Wife's education						
- High school graduate	-0.105	0.59	0.048	0.34	-0.206	0.55
- College	0.069	0.34	0.052	0.32	-0.078	0.16
- Graduate work or degree	0.783	3.51	-0.005	0.02	1.442	2.63
Wife's employment						
- Unemployed	0.492	1.66	0.920	5.34	-0.558	0.72
- Out of work force	-0.058	0.36	1.386	15.34	-0.508	1.69
Age of wife	-0.080	2.32	0.015	0.39	-0.123	1.00
(Age of wife)^2	0.001	2.35	-0.001	1.37	0.001	0.82
# of kids < 6 years old	----	----	----	----	-0.571	1.49
# of kids between 6 & 18 years old	----	----	----	----	----	----
# of kids >= 18 years old	----	----	----	----	----	----
Previous marriage	----	----	----	----	----	----
Previously a single parent	----	----	----	----	0.155	0.39
- Log likelihood($\hat{\gamma}_0$, $\hat{\gamma}$, $\hat{\beta}$)	1518.48		2175.21		394.27	
- Log likelihood($\hat{\gamma}_0$, $\hat{\gamma}$)	1751.98		2708.82		458.44	

[a] Race is only reported for the head of the household. For single households, this poses no uncertainty. For couples, the head may be male or female. In --% of couples, the man has been coded as the head of the household.

[b] For all couples (even cohabiting adults), I refer to the male as the husband and the female as the wife.

Table 6.12.7: Estimation Results for Independent Hazard Functions
Weibull Duration Dependence without Unobserved Heterogeneity

	Couple, 1 child → Couple		Couple, 1 child → Couple, 2 children		Couple, 1 child → Single, 1 child	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Constant intercept (γ_0)	-10.798	8.33	-4.046	5.07	-0.747	0.43
Gamma_1 (γ)	0.176	2.44	0.539	9.62	0.330	2.28
Log (lagged household income)	-0.017	0.31	0.049	0.84	-0.138	0.78
Race (vs. white) [a]						
- Black	-0.003	0.01	0.172	1.07	0.565	1.67
- Other	0.120	0.63	-0.211	1.53	-0.622	1.36
Husband's education [b]						
- High school graduate	0.127	0.80	-0.042	0.31	0.192	0.70
- College	0.163	0.88	-0.030	0.21	-0.284	0.80
- Graduate work or degree	0.349	1.28	0.087	0.41	0.778	2.07
Husband's employment						
- Unemployed	0.305	0.89	0.161	0.87	0.751	1.82
- Out of work force	-0.217	1.05	-0.260	0.98	1.975	6.82
Age of husband	0.116	1.64	-0.084	1.98	-0.068	0.70
(Age of husband)^2	-0.001	1.24	0.001	1.58	-0.000	0.01
Wife's education						
- High school graduate	-0.277	1.77	-0.185	1.37	-0.187	0.66
- College	-0.039	0.19	-0.021	0.14	-0.098	0.28
- Graduate work or degree	-0.458	1.40	0.182	0.78	-0.154	0.29
Wife's employment						
- Unemployed	-0.246	0.52	0.687	3.66	-0.304	0.46
- Out of work force	0.005	0.39	0.657	7.76	-0.687	2.44
Age of wife	0.105	1.64	0.101	1.96	-0.056	0.60
(Age of wife)^2	-0.001	1.76	-0.002	2.51	0.001	0.98
# of kids < 6 years old	----	----	1.006	8.38	-0.035	0.10
# of kids between 6 & 18 years old	----	----	----	----	----	----
# of kids >= 18 years old	2.391	10.76	----	----	----	----
Previous marriage	----	----	----	----	----	----
Previously a single parent	----	----	----	----	0.985	3.58
- Log likelihood($\hat{\gamma}_0$, $\hat{\gamma}$, $\hat{\beta}$)	1070.89		2203.47		533.49	
- Log likelihood($\hat{\gamma}_0$, $\hat{\gamma}$)	1384.40		2498.12		630.33	

[a] Race is only reported for the head of the household. For single households, this poses no uncertainty. For couples, the head may be male or female. In --% of couples, the man has been coded as the head of the household.
[b] For all couples (even cohabiting adults), I refer to the male as the husband and the female as the wife.

Table 6.12.8: Estimation Results for Independent Hazard Functions

Weibull Duration Dependence without Unobserved Heterogeneity

	Couple, 2 children → Single		Couple, 2 children → Couple, 1 child		Couple, 2 children → Couple, 3+ kids	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Constant intercept (γ_0)	3.439	1.11	-13.826	9.35	-0.368	0.27
Gamma_1 (γ)	0.308	1.42	0.066	0.91	0.197	2.26
Log (lagged household income)	-0.203	2.09	-0.035	0.49	-0.055	0.58
Race (vs. white) [a]						
- Black	0.270	0.42	-0.291	1.15	0.519	2.11
- Other	-0.357	0.62	0.125	0.65	0.145	0.79
Husband's education [b]						
- High school graduate	-0.146	0.30	-0.043	0.26	0.035	0.19
- College	0.413	0.82	-0.314	1.54	0.086	0.42
- Graduate work or degree	0.661	0.93	-0.268	0.92	-0.056	0.18
Husband's employment						
- Unemployed	0.467	0.95	0.698	2.60	0.201	0.82
- Out of work force	-0.690	0.44	-0.202	0.78	0.220	0.58
Age of husband	-0.046	0.23	0.138	1.94	-0.034	0.52
(Age of husband)^2	0.001	0.20	-0.001	1.46	0.000	0.40
Wife's education						
- High school graduate	0.145	0.26	-0.215	1.37	0.009	0.05
- College	0.022	0.33	-0.224	1.00	0.300	1.31
- Graduate work or degree	0.663	0.82	-0.070	0.21	0.304	0.90
Wife's employment						
- Unemployed	0.356	0.59	-0.029	0.06	0.702	2.21
- Out of work force	-1.075	2.54	0.055	0.41	0.981	7.30
Age of wife	-0.294	1.53	0.224	2.76	-0.144	2.01
(Age of wife)^2	0.003	1.17	-0.003	2.87	0.001	1.67
# of kids < 6 years old	-0.039	0.57	----	----	0.154	0.65
# of kids between 6 & 18 years old	0.025	0.05	0.799	2.83	0.050	0.24
# of kids >= 18 years old	----	----	1.803	6.25	----	----
Previous marriage	----	----	----	----	----	----
Previously a single parent	-0.029	0.05	----	----	----	----
- Log likelihood($\hat{\gamma}_0$, $\hat{\gamma}$, $\hat{\beta}$)	357.99		1070.40		1273.39	
- Log likelihood($\hat{\gamma}_0$, $\hat{\gamma}$)	393.18		1392.87		1366.12	

[a] Race is only reported for the head of the household. For single households, this poses no uncertainty. For couples, the head may be male or female. In --% of couples, the man has been coded as the head of the household.

[b] For all couples (even cohabiting adults), I refer to the male as the husband and the female as the wife.

Table 6.12.9: Estimation Results for Independent Hazard Functions

Weibull Duration Dependence without Unobserved Heterogeneity

	Couple, 2 children → Single, 2 kids		Couple 3+ kids → Couple, 2 children		Couple, 3+ kids → Single, 3+ kids	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Constant intercept (γ_0)	3.872	1.28	-7.970	2.74	1.510	0.37
Gamma_1 (γ)	0.438	2.60	0.057	0.57	0.319	1.15
Log (lagged household income)	-0.083	0.35	-0.080	0.93	0.053	0.20
Race (vs. white) [a]						
- Black	0.365	0.79	-0.171	0.45	0.692	1.33
- Other	-0.041	0.12	0.047	0.18	-0.075	0.15
Husband's education [b]						
- High school graduate	0.356	0.98	-0.345	1.58	-0.496	1.10
- College	0.303	0.75	-0.388	1.53	0.806	1.55
- Graduate work or degree	1.504	3.23	-1.133	2.99	0.273	0.33
Husband's employment						
- Unemployed	0.871	1.97	-0.041	0.11	0.868	1.54
- Out of work force	2.262	5.83	0.077	0.21	2.850	5.55
Age of husband	-0.037	0.30	0.216	2.11	0.366	1.99
(Age of husband)^2	0.000	0.22	-0.002	1.69	-0.004	2.02
Wife's education						
- High school graduate	-0.085	0.24	-0.122	0.49	0.030	0.07
- College	-0.434	0.96	-0.111	0.38	-0.404	0.59
- Graduate work or degree	-0.719	1.04	-0.337	0.62	0.241	0.27
Wife's employment						
- Unemployed	1.386	4.29	0.788	1.78	1.022	1.40
- Out of work force	-0.255	0.83	0.058	0.33	-0.227	0.56
Age of wife	-0.315	2.15	0.200	1.78	-0.707	3.46
(Age of wife)^2	0.003	1.81	-0.002	1.76	0.007	3.22
# of kids < 6 years old	-0.966	2.48	-1.720	4.86	-0.019	0.05
# of kids between 6 & 18 years old	-0.370	1.32	-1.572	5.39	-0.179	0.54
# of kids >= 18 years old	----	----	-0.960	3.46	0.156	0.41
Previous marriage	----	----	----	----	----	----
Previously a single parent	0.464	1.42	----	----	0.980	2.13
- Log likelihood($\hat{\gamma}_0$, $\hat{\gamma}$, $\hat{\beta}$)	431.35		674.47		238.70	
- Log likelihood($\hat{\gamma}_0$, $\hat{\gamma}$)	515.30		850.04		303.47	

[a] Race is only reported for the head of the household. For single households, this poses no uncertainty. For couples, the head may be male or female. In --% of couples, the man has been coded as the head of the household.

[b] For all couples (even cohabiting adults), I refer to the male as the husband and the female as the wife.

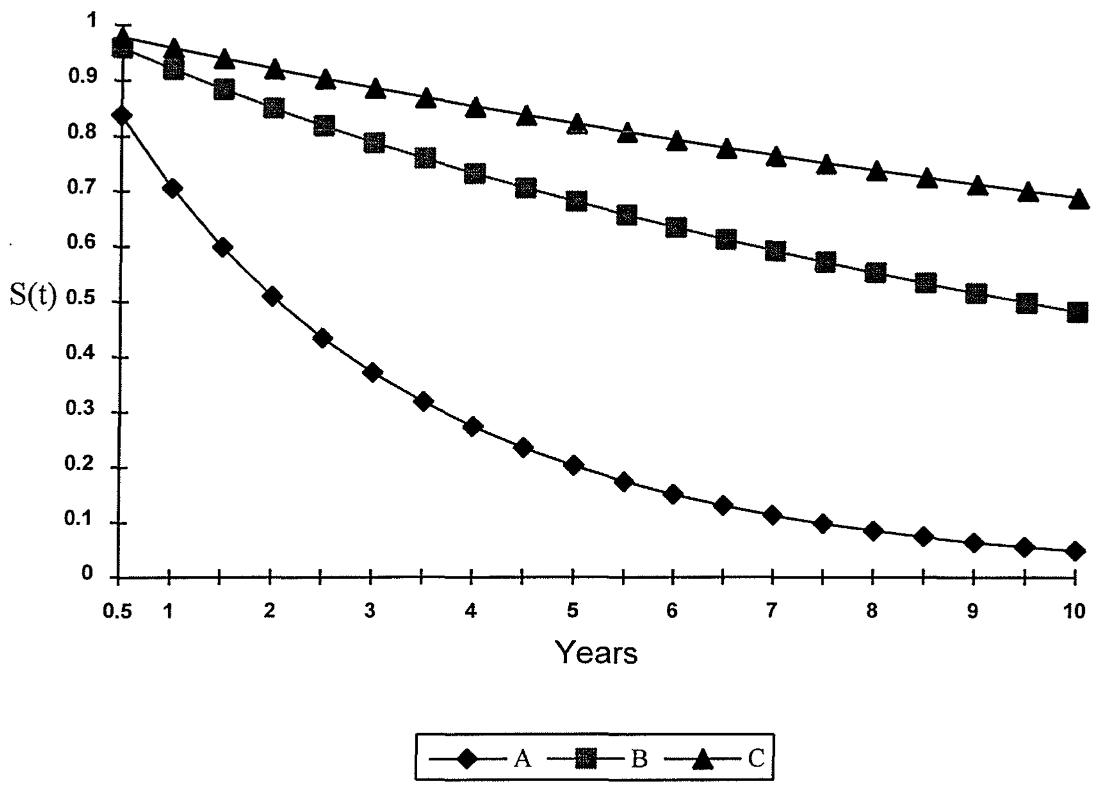
Are these models better than simple Weibull models without covariates? The likelihood ratio test answers this question by testing the null hypothesis that the coefficients on the covariates are zero, H_0: $\hat{\beta} = 0$. For all transitions except from 'Single with 1 child' to 'Single with 2 children', the likelihood ratio test rejects the null hypothesis that the coefficients are zero (see Appendix A for further details). The full models with covariates describe the data better than simple models without covariates.

6.7.1.1 Survivor Curves for Select Transitions

It is often easier to interpret the estimation results by plotting survivor functions. Survivor curves give the probability that the household remains in the beginning state at some time *t* given that the household has not made a transition before *t*. So, for the transition from single to couple, the survivor curve gives the probability that the person has not yet married.

Survivor curves depends on the covariate values, some of which may be time-varying. To aid in the presentation, I have selected a few representative households and transitions. The following figures show survivor curves for the transitions from 'Single' to 'Single with 1 child', from 'Single' to 'Couple', from 'Couple with 2 children' to 'Single with 2 children', and from 'Couple' to 'Couple with 1 child'. Only a few representative households were chosen, but others may be added later.

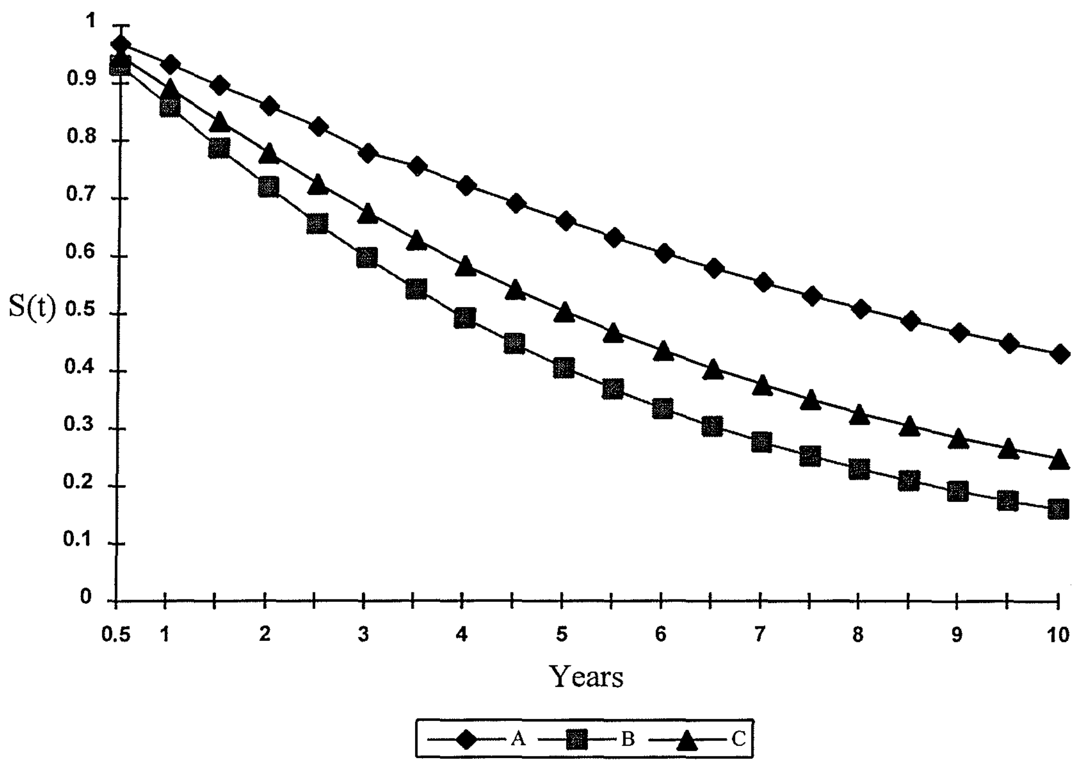
Fig 2. Survival Curves for Single to Single with 1 child



A = Black female who dropped out of high school, is employed, earns \$15,000/year (\$7500 per 6 months), and was age 18 in year 1.
B = White female who dropped out of high school, is employed, earns \$15,000/year, and was age 18 in year 1.
C = White female who is a college graduate, is employed, earns \$30,000/year, and age 22 in year 1.

As this figure shows, black women who drop out of high school are most likely to have a child without being married. Their survival curve slopes down the quickest, implying that they transition from single to single with a child sooner. For both cateogies of white women, the chance of survival is much higher, e.g. they are less likely to have a child before marriage. But differences still exist between highly educated, higher earning white women and their less educated, poorer counterparts. Those who dropped out of high school and earn less are likely to transition sooner.

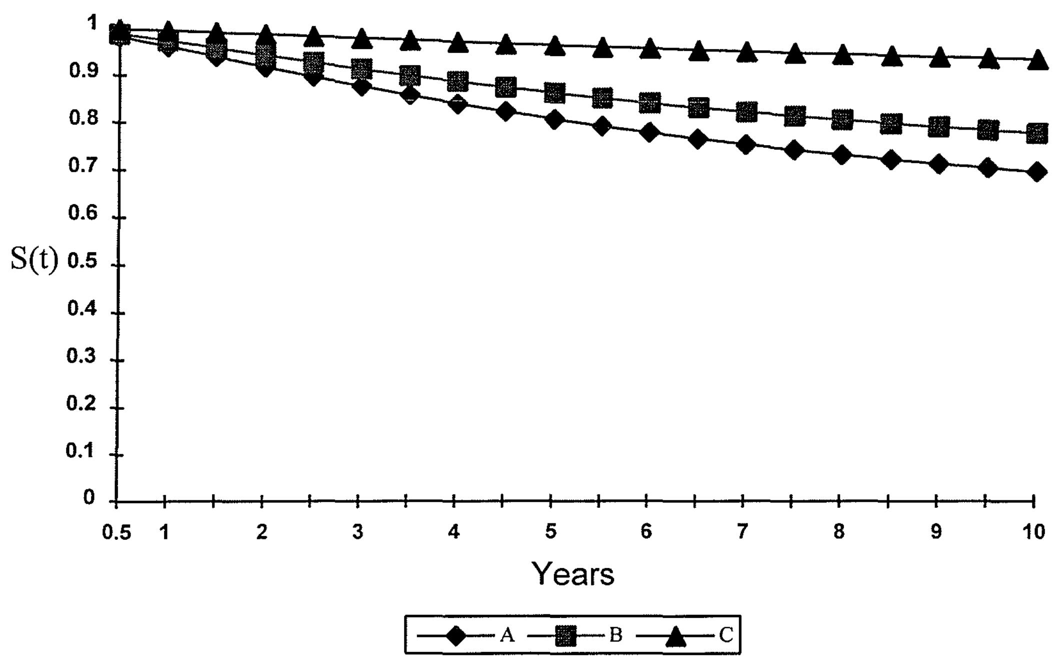
Fig. 3 Survival Curves for Single to Couple



A = Black female who dropped out of high school, is employed, earns $15,000/year ($7500 per 6 months), was age 18 in year 1, and was not previously married.
B = White female who dropped out of high school, is employed, earns $15,000/year, was age 18 in year 1, and was not previously married.
C = White female who is a college graduate, is employed, earns $30,000/year, was age 22 in year 1, and wasnot previously married.

The survival curves for these three groups of women are much more similar than the previous figure. White women who drop out of high school are the most likely to marry at an earlier age. The next most likely to marry early are black women with little education. Finally, white women with college degrees are the most likely to postpone marriage. Keep in mind that I have included long-term cohabitation relationship in the married category.

Fig. 4 Survival Curves for Couple with 2 children to Single with 2 children



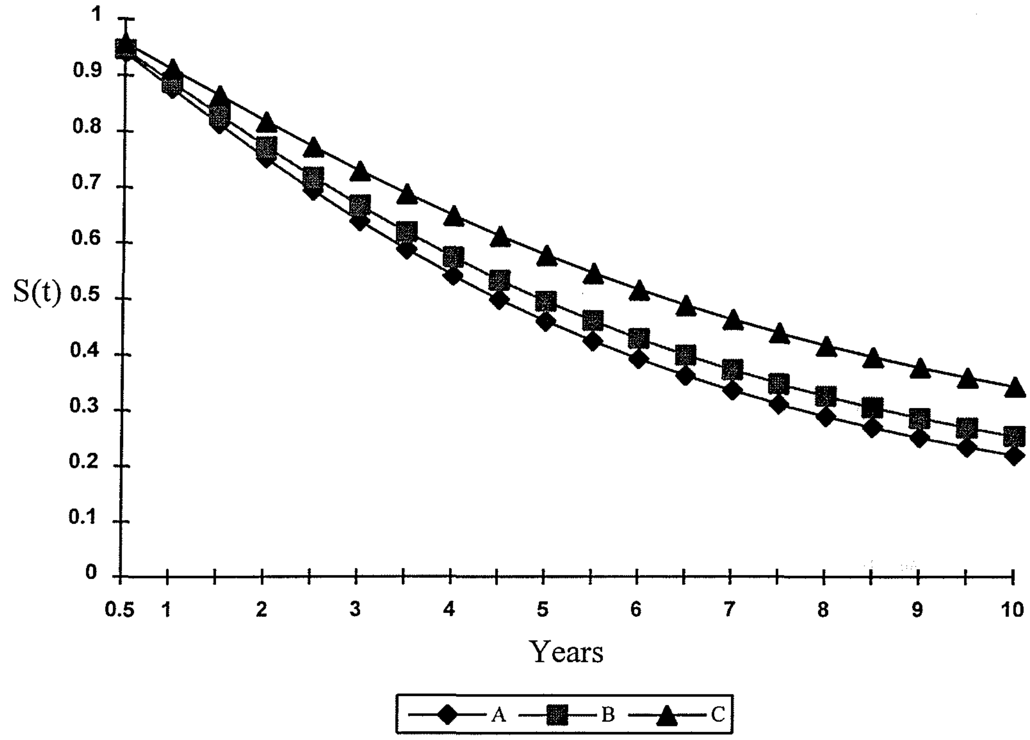
A=Black household, husband and wife both employed, household income of $15,000/year, both husband and wife dropped out of high school and were age 25 in year 1. Both children between the ages of 6 and 18. Children were not born before the marriage.

B=White household, husband and wife both employed, household income of $15,000/year, both husband and wife dropped out of high school and were age 25 in year 1. Both children between the ages of 6 and 18. Children were not born before the marriage.

C=White household, husband employed, wife is out of the work force, household income of $40,000/year, both husband and wife went to college and were age 30 in year 1. Both children between the ages of 6 and 18. Children were not born before the marriage.

This figure shows that blacks are likely to transition from 'Couple with 2 children' to 'Single with 2 children' before whites. The household most likely to remain in the stable marriage consists of a housewife (e.g. she is out of the workforce) and a sole breadwinning husband who makes a relatively high income ($40,000/year).

Fig. 5 Survival Curves for Couple to Couple, 1 child



A=Black household, husband and wife are both employed, household income of $15,000/year, both husband and wife dropped out of high school and were age 25 in year 1.
B=White household, husband and wife are both employed, household income of $15,000/year, both husband and wife dropped out of high school and were age 25 in year 1.
C=White household, husband and wife are both employed, household income of $40,000/year, both husband and wife went to college and were age 30 in year 1.

This figure shows that the timing of child birth is similar for these three categories of households. Differences across racial groups are not as dramatics as some of the previous figures. Wealthier households are more likely to postpone child birth. What this figure does not show is that households which consists of a stay-at-home wife are likely to have a child before households with working women.

6.7.2 Unobserved Heterogeneity

Next we have grouped transitions according to possible inter-relations in the hazard functions. As mentioned earlier, the decision to have another (or a first) child may be predicated by the chance of divorce, and vice verse. These types of inter-relationships may be captured with unobserved heterogeneity which will be estimated within the following groups shown in Table 13.

Table 6.13: Groups of Transitions for Unobserved Heterogeneity Estimation

Group 1:	Group 2:
Single → Couple	Couple, 1 child → Single
Single → Couple, 1 child	Couple, 1 child → Couple
Single → Single, 1 child	Couple, 1 child → Couple, 2 children
Single → Other	Couple, 1 child → Single, 1 child
Couple → Single	Couple, 2 children → Single
Couple → Couple, 1 child	Couple, 2 children → Couple, 1 child
Other → Single	Couple, 2 children → Single, 2 children
Other → Couple	
Other → Couple, 1 child	
Group 3:	Group 4:
Single, 1 child → Single	Couple, 2 children → Couple, 3+ children
Single, 1 child → Couple, 1 child	Couple, 3+ children → Couple, 2 children
Single, 1 child → Single, 2 children	Couple, 3+ children → Single, 3+ children
Single, 2 children → Couple, 2 children	Single, 2 children → Single, 3+ children
Single, 2 children → Single, 1 child	Single, 3+ children → Single, 2 children

Estimation results for all groups indicate that unobserved heterogeneity is insignificant. When important explanatory variables were excluded (for example, race), unobserved heterogeneity was estimated to be significant. Therefore, the full models with all covariates are complete in the sense that no significant unobserved heterogeneity remains.

6.8 Implementing the Demographic Microsimulation

The estimated coefficients from Tables 12.1 - 12.9 are used to simulate movements from one household type to another over the 16 years from 1993-2010. A "snapshot" of households is generated each 6-month period for a total of 32 snapshots during the simulation. Each snapshot contains the following information about each household:

- Household id (less than 5000 if the household is a member of the original survey
  sample, over 5000 if the household is created during the simulation process)
- Household type (one of the nine possible types)
- Time that the household has been classified as the current household type
- Household income

- Indicator of whether the household is or has previously consisted of a couple
- Indicator of whether the household is or has previously consisted of a single parent
- Indicator of whether the household married last period
- Indicator of whether the household divorced last period
- Number of adults and children in the household
- Age and gender of household members (including children)
- Employment, education and race for the husband or single head of the household
- Employment and education for the wife in a couple household
- Household weight
- Geographic location of the household (SCE district or PG&E planning area)
- Detailed indicators of the type of household change that occurred last period

All households that responded to the stated-preference survey have been categorized into one of the nine household types. The current household type determines how the household may change (e.g. which of the remaining eight types the household be reclassified into). In many cases, no change occurs during the 6-month period and the household remains as the same type. The following list shows the possible movements.

Current Type	Possible New Types (may change over 1 period)
Single	Remain as Single
	Single with 1 child
	Couple
	Couple with 1 child
	Other
Single with 1 child	Remain as Single with 1 child
	Single
	Single with 2 children
	Couple with 1 child
Single with 2 children	Remain as Single with 2 children
	Single with 1 child
	Single with 3+ children
	Couple with 2 children
Single with 3+ children	Remain as Single with 3+ children
	Single with 2 children
Couple	Remain as Couple
	Single
	Couple with 1 child

Current Type	Possible New Types (may change over 1 period)
Couple with 1 child	Remain as Couple with 1 child
	Single
	Single with 1 child
	Couple
	Couple with 2 children
Couple with 2 children	Remain as Couple with 2 children
	Single
	Single with 2 children
	Couple with 1 child
	Couple with 3+ children
Couple with 3+ children	Remain as Couple with 3+ children
	Single with 3+ children
	Couple with 2 children
Other	Remain as Other
	Single
	Couple
	Couple with one child

Detailed flags indicate exactly which type of change occurred. They are as follows:

0	No change
1	Single to Single with 1 child
2	Single to Couple
3	Single to Couple with 1 child
4	Single to Other
5	Couple splits to Single & Single
6	Couple to Couple with 1 child
7	Couple with 1 child splits into Single & Single with 1 child
8	Couple with 1 child splits to Couple & Single (e.g. child leaves home)
9	Couple with 1 child to Couple with 2 children
10	Couple with 2 children to Single & Single with 2 children
11	Couple with 2 children to Couple with 1 child & Single (e.g. child leaves home)
12	Couple with 2 children to Couple with 3 children
13	Couple with 3+ children to Single & Single with 3+ children
14	Couple with 3+ children add another child
15	Couple with 3+ children lose a child (may reclassify as Couple with 2 children if only 3 children to begin with)
16	Single with 1 child to Single & Single (e.g. child leaves home)
17	Single with 1 child to Couple with 1 child
18	Single with 1 child to Single with 2 children
19	Single with 2 children to Single with 1 child & Single (e.g. child leaves home)
20	Single with 2 children to Couple with 2 children
21	Single with 2 children to Single with 3 children

22 Single with 3+ children lose a child (may reclassify as Single with 2 children if only 3 children to begin with)
23 Single with 3+ children add another child
24 Other to Single
25 Other to Couple (found a mate for the 'other' single)
26 Other to Couple with 1 child (found a mate and child)
27 Single with 1 child to Single (child has died)
28 Single with 2 children to Single with 1 child (child has died)
29 Single with 3 children to Single with 2 children (child has died)
30 Couple with 1 child to Couple (child has died)
31 Couple with 2 children to Couple with 1 child (child has died)
32 Couple with 3 children to Couple with 2 children (child has died)
33 Couple to Single (husband or wife died)
34 Couple with 1 child to Single with 1 child (husband or wife died)
35 Couple with 2 children to Single with 2 children (husband or wife died)
36 Couple with 3 children to Single with 3 children (husband of wife died)

For a given household, the programs determine the probability of each type of change. For a 'Single' household, the programs determine the probability of moving to 'Single with 1 child' (PStoS1), the probability of moving to 'Couple' (PStoC), the probability of moving to 'Couple with 1 child' (PStoC1), and the probability of moving to 'Other' (PStoO). The probability of no change (e.g. remaining as a single household) is equal to 1 - (PStoS1 + PStoC + PStoC1 + PStoO). The probabilities are derived by integrating the appropriate hazard function. For example, if $h_{StoS1}(x,t)$ is the hazard function for a movement from 'Single' to 'Single with 1 child', the probability of making such a change from time t0 to t1 is given by:

$$\text{PStoS1} = 1 - \exp(-\int_{t0}^{t1} h_{\text{StoS1}}(x,u)du) \qquad (14)$$

Please see the preceding documentation for additional details about the specification of the hazard function, estimated coefficients, and explanatory variables.

The simulation occurs by drawing a random number (r.n.) between 0 and 1 from a Uniform distribution. Continuing with the example of a 'Single' household, household changes are determined based on the following scheme.

IF r.n. < PStoS1
- Add a newborn child to the household
- Reclassify as 'Single with 1 child'

ELSE IF r.n. >= PStoS1 & r.n. < PStoS1+PStoC
- Find a mate for the single person
- Reclassify as 'Couple'

ELSE IF r.n. >= PStoS1+PStoC & r.n. < PStoS1+PStoC+PStoC1

- Find a mate and child for the single person
- Reclassify as 'Couple with 1 child'

ELSE IF r.n. >= PStoS1+PStoC+PStoC1 & r.n.< PStoS1+PStoC+PStoC1+PStoO
- Reclassify as 'Other'

ELSE
- Remain as a single household with no changes

Similar logic holds for all the other the household types, with the IF-THEN-ELSE logic depending upon the possible types of changes that a given household can undergo.

6.8.1 Finding a Mate

When a household change occurs that requires a new mate (or a new mate with children), the programs searches the current sample of households for a couple-household with similar characteristics to the single person undergoing the marriage. This procedure occurs for the following household changes:

- Single to Couple
- Single to Couple with 1 child
- Single with 1 child to Couple with 1 child
- Single with 2 children to Couple with 2 children
- Other to Couple
- Other to Couple with 1 child

If the single person is a man, a couple household with a matching husband is found. If the single person is a woman, the match is made to the characteristics of the wife in the couple. The routine looks for a "close" match, and then gradually broadens the scope of the search until a match is found. The matching criteria was set as follows:

1st try	Match on age plus or minus 5 years, same race, education and employment status. If the single household contained children, the matching couple household should also have children in the same age categories.
2nd try	Match on age plus or minus 10 years, same race, education and employment status. Again, if the single household contained children, the matching couple household should also have children in the same age categories.
3rd try	Match on age plus or minus 10 years, same race, and education. Again, if the single household contained children, the matching couple household should also have children in the same age categories.
4th try	Match on age plus or minus 10 years and same race. Again, if the single household contained children, the matching couple household should also have children in the same age categories.

5th try	Match on age plus or minus 10 years. No restriction on children's age.
Final	Spouse is given the same characteristics as single person. New child is assumed to be a newborn infant with equal chance of boy or girl.

When a matching couple is found, the spouse's characteristics (age, education, employment status) are duplicated for this newly married household.

6.8.2 Divorce or Separation

When a household change occurs that implies the removal of an adult member, the programs first check to determine if either the husband or wife has died (see the following section on mortality). If no one has died, the programs create a new household which contains the split-off from the original household. This procedure occurs for the following changes:

- Couple to Single
- Couple with 1 child to Single
- Couple with 1 child to Single with 1 child
- Couple with 2 children to Single
- Couple with 2 children to Single with 2 children
- Couple with 3+ children to Single with 3+ children

The PSID used the head of the household (the husband in a couple) as the continuation of the original household. Wives were treated as split-off households . For that reason, movements from Couple with 1 child to Single, Couple with 2 children to Single, and Couple with 3+ children to Single lead to the creation of a new household that contains the wife and children from the original household. The husband remains in the original household as a single. For movements from Couple to Single, Couple with 1 child to Single with 1 child, Couple with 2 children to Single with 2 children, and Couple with 3+ children to Single with 3+ children, the programs create a new single household that only contains the wife from the original household. The husband remains in the original household with the children when appropriate.

6.8.3 Children leaving home

When a household change occurs that implies the removal of a child, the programs first check to determine if a child has died (see the following section on mortality). If no children have died, the programs assume that the oldest child has left the home *if* that eldest child is at least 16 years old. A new Single household is created which contains the oldest child (who is over age 16) from the original household. This procedure occurs for the following changes:

- Single with 1 child to Single
- Single with 2 children to Single with 1 child
- Single with 3+ children to Single with 2 children
- Couple with 1 child to Couple
- Couple with 2 children to Couple with 1 child
- Couple with 3+ children to Couple with 2 children

Education for this split-off child is set to education level of his/her father, or if no father exists, to the education level of his/her mother (e.g. the head of the household from which the child came). New split-off children are assumed to be employed.

In addition, children may leave a household with 3+ children which will not result in a reclassification to a household with 2 children. For example, the household may have 5 children. When the eldest child leaves, 4 children remain so that the household is still classified as having 3+ children. The rate at which the 4th or higher child leaves the household is the same as the rate of movement from a household with 3 children to a household with 2 children.

6.8.4 Mortality

Certain types of household changes may have been caused by death (either of a spouse or child). The following breakdown shows which household changes may have been caused by death.

Death of an adult spouse:
- Couple to Single
- Couple with 1 child to Single with 1 child
- Couple with 2 children to Single with 2 children
- Couple with 3 children to Single with 3+ children

Death of a child:
- Single with 1 child to Single
- Single with 2 children to Single with 1 child
- Single with 3 children to Single with 2 children
- Couple with 1 child to Couple
- Couple with 2 children to Couple with 1 child
- Couple with 3 children to Couple with 2 children

Additionally, singles may die and should be removed from the sample.

In all these cases, the probability of death is taken from 1989 Mortality Tables from the Vital Statistics of the United States, Vol. II Sec. 6, Life Tables. These tables provide mortality rates by age, race, and gender.

6.8.5 Weighting

Weights are adjusted when couples split into two households, when children leave home and start their own households, or when the spouse in a household dies.

When couples split apart, the original household weight was divided by 2 for each of the households (e.g. the continuation of the original household and the split-off household for the newly single woman). This occurred for the following household changes:

- Couple to Single
- Couple with one child to Single
- Couple with one child to Single with one child
- Couple with two children to Single
- Couple with two children to Single with two children
- Couple with three + children to Single with three + children

When children leave home, a matching single household was found and that household's weight was replicated for the split-off household being created. Matches were made on the same criteria as outlined in section 6.7.1 (Finding a Mate). These matches were made for the following household changes:

- Single with 1 child to Single
- Single with 2 children to Single with 1 child
- Single with 3 children to Single with 2 children
- Couple with 1 child to Couple
- Couple with 2 children to Couple with 1 child
- Couple with 3 children to Couple with 2 children

Finally when a spouse dies, the original household's weight is divided by 2 (similar to the case when couples split apart). This may occur for the following household changes:

- Couple to Single
- Couple with one child to Single
- Couple with one child to Single with one child
- Couple with two children to Single
- Couple with two children to Single with two children
- Couple with three + children to Single with three + children

After the demographic simulation was complete, population totals were calculated by adding up the household weight times the number of household members (for example, a couple with a household weight of 100 would contribute 200 to the population total). The population totals were then compared to external population forecasts for Southern California. These external forecasts project a population growth rate of 1.59% per year (U.S. Census Bureau Population Projections from 1993 to 2020). This growth rate included increases from immigration which was not included in the demographic microsimulation. As a result, the population growth rate from the microsimulation model is below 1.59%. Therefore, the households weights were re-scaled so that the population from the simulated sample also grew at 1.59% per year.

6.8.6 Additional Models - Employment

Every 6-month period, the employment status for all major adults are updated using employment-change equations. There are 3 employment categories: employed, unemployed, and out of the workforce. Conditioned upon the starting employment status, changes in employment are modeled using multinomial logit specifications. There are a series of employment-change equations for different types of people: one set for husbands, one set for wives, and a third set for single household heads. 'Other' households are treated as single households, and employment changes are only modeled for the head of the household.

For employed individuals, the dependent variable (which indicates the type of employment change that might occur in the next 6 months) takes on the following values: 0 - no change, 1 - from employed to unemployed, and 2 -from employed to out-of-the-workforce. For unemployed husbands, the dependent variable takes on the following values: 0 - no change, 1 - from unemployed to employed, and 2 - from unemployed to out of the workforce. for husbands that are out of the workforce, the dependent variable takes on the following values: 0 - no change, 1 - from out of the workforce to employed, and 2 - from out of the workforce to unemployed.

Employment changes were simulated by comparing the "utility" (V_{ij}) for each type of possible change. In this multinomial logit specification, V_{ij} consists of a deterministic component and a random component as follows:

$$V_{ij} = \beta_{ij}X + e_{ij}, \qquad \text{where } e_{ij} \text{ follows a lognormal distribution} \tag{15}$$

The simulation proceeded by calculating the deterministic component, drawing a random component from a lognormal distribution, and then choosing the type of employment change with the highest "utility". The deterministic component was calculated using the estimates for β_{ij} as given in the following tables.

Table 6.14.1: Employment Change Estimates for Husbands

Employed → Unemployed		Employed → Out-of-workforce	
High school	-0.6277	High school	-0.2639
College	-1.1549	College	-0.4261
Graduate school	-1.7665	Graduate school	-0.1030
Black	0.5184	Black	0.6358
Age	-0.2883	Age	-0.1517
Age^2	0.0000	Age^2	0.0024
Has children	-0.0262	Has children	-0.4287
Intercept	-1.6246	Intercept	-1.6111
Unemployed → Employed		**Unemployed → Out-of-workforce**	
High school	0.4462	High school	-0.2379
College	0.7727	College	0.2423
Graduate school	0.7449	Graduate school	1.3496
Black	0.2332	Black	0.6062
Age	-0.0020	Age	-0.0838
Age^2	0.0001	Age^2	0.0016
Has children	0.2964	Has children	0.0508
Intercept	-0.0595	Intercept	-0.4439
Out-of-workforce → Employed		**Out-of-workforce → Unemployed**	
High school	0.5635	High school	-0.3987
College	0.5267	College	-0.2968
Graduate school	1.2259	Graduate school	0.0320
Black	-0.2885	Black	0.5584
Age	-0.0896	Age	-0.0702
Age^2	0.0000	Age^2	-0.0005
Has children	-0.2382	Has children	0.0000
Intercept	2.5474	Intercept	0.8423

Table 6.14.2: Employment Change Estimates for Wives			
Employed → Unemployed		**Employed → Out-of-workforce**	
High school	-0.0500	High school	-0.3546
College	-0.8085	College	-0.3113
Graduate school	-0.7736	Graduate school	-0.4423
Black	-0.5812	Black	-0.1943
Age	0.4624	Age	-0.2727
Age^2	-0.1087	Age^2	0.0032
Has children	0.0001	Has children	0.1773
Intercept	-0.0500	Intercept	3.4157
Unemployed → Employed		**Unemployed → Out-of-workforce**	
High school	0.6950	High school	0.3761
College	0.5245	College	-0.0143
Graduate school	1.0192	Graduate school	0.6480
Black	-0.3565	Black	0.3921
Age	-0.1825	Age	-0.1947
Age^2	0.0022	Age^2	0.0025
Has children	0.2790	Has children	0.1645
Intercept	4.0848	Intercept	3.8056
Out-of-workforce → Employed		**Out-of-workforce → Unemployed**	
High school	0.2974	High school	-0.4486
College	0.3531	College	-0.5012
Graduate school	0.6680	Graduate school	-0.1565
Black	0.3577	Black	1.0304
Age	0.1255	Age	0.0622
Age^2	-0.0022	Age^2	-0.0016
Has children	-0.3041	Has children	-0.6537
Intercept	-2.9121	Intercept	-2.7640

Table 6.14.3: Employment Change Estimates for Single Household Heads

Employed → Unemployed		Employed → Out-of-workforce	
High school	-0.3526	High school	-0.1689
College	-0.8929	College	-0.1738
Graduate school	-1.0512	Graduate school	-0.4699
Black	-0.0066	Black	-0.0465
Divorce last period	0.3423	Divorce last period	0.8059
Age	-0.1018	Age	-0.2218
Age^2	0.0008	Age^2	0.0028
Gender (female)	-0.0974	Gender (female)	0.4454
Has children	0.3186	Has children	0.5294
Intercept	0.0164	Intercept	0.2478
Unemployed → Employed		**Unemployed → Out-of-workforce**	
High school	0.8774	High school	0.1184
College	1.3186	College	0.0779
Graduate school	0.2143	Graduate school	-0.3582
Black	-1.5031	Black	-0.5001
Divorce last period	-0.2330	Divorce last period	0.1491
Age	-0.0302	Age	-0.1390
Age^2	0.0002	Age^2	0.0019
Gender (female)	0.1876	Gender (female)	1.0376
Has children	0.1237	Has children	0.3016
Intercept	0.9492	Intercept	1.0052
Out-of-workforce → Employed		**Out-of-workforce → Unemployed**	
High school	0.5637	High school	-0.2072
College	1.3262	College	-0.2165
Graduate school	0.3656	Graduate school	0.2981
Black	-0.3311	Black	0.9202
Divorce last period	0.9799	Divorce last period	-0.6575
Age	0.0363	Age	-0.0130
Age^2	-0.0010	Age^2	-0.0007
Gender (female)	-0.2016	Gender (female)	-0.4638
Has children	-0.2007	Has children	-0.5115
Intercept	-1.2758	Intercept	-0.0657

6.8.7 Additional Models - Income

Household income is updated each year using models that forecast the change in the log of income. There are two such income change models, one for single-headed household (including 'Other' households) and one for couple households. The general form of the model is given by:

$$\text{Log}(\text{income}_t) - \text{Log}(\text{income}_{t-1}) = \beta X_t + \varepsilon_t \tag{16}$$

The error term is assumed to be normally distributed with mean zero and variance as estimated. The estimates for β and the variance component of the error term are given in the following tables.

Table 6.15.1: Income Change Estimates
Single Households

Log(last year's income)	-0.4083
High School	0.1353
College	0.2731
Graduate school	0.3513
Currently employed	0.3069
Entered the workforce during the year	-0.2812
Black	-0.2104
Other racial classification (non-black, non-white)	-0.0313
Divorced during the year	-0.2042
Age	0.0093
Age^2	-0.0001
Gender (female)	-0.1182
Black x Employed	0.0698
Children in the household	0.1913
Intercept	3.4733
Standard deviation of the error term	0.4381

Table 6.15.2: Income Change Estimates Couple Households	
Log(last year's income)	-0.4586
High School - husband	0.1101
College - husband	0.1991
Graduate school - husband	0.2710
High School - wife	0.0894
College - wife	0.1525
Graduate school - wife	0.1819
Currently employed - husband	0.2471
Currently employed - wife	0.1032
Black	-0.1771
Other racial classification (non-black, non-white)	-0.0042
Married during the year	0.0722
Age of husband	0.0165
(Age of husband)^2	-0.0002
Age of wife	0.0085
(Age of wife)^2	-0.0001
Black x Employed husband	0.0649
Black x Employed wife	0.0575
Children in the household	0.0133
Intercept	3.7354
Standard deviation of the error term	0.4480

7. HOUSEHOLD TRANSACTIONS TIMING

7.1 Introduction

This chapter presents the results of a series of analyses that examined the characteristics of households' vehicle holding durations and stated intentions about the next transaction. Alternative distribution functions are evaluated for their ability to represent the distributions of vehicle holding durations in the first-wave sample of the survey. The distributions examined include: negative exponential, log-normal, log-logistic, generalized gamma, and Weibull. Based on the statistical results, the Weibull distribution is selected as the most suitable distribution and used for further analysis. The explanatory variables used in the duration models include those describing the attributes of the vehicle, the household vehicle fleet, the primary user of the vehicle, and the household. A similar set of explanatory variables is used to develop models that describe the timing of the next intended vehicle transaction.

The models presented in this report are exploratory in nature and their usefulness as predictive models is limited because the analysis is based solely on the data available from the first-wave of the panel survey. Although retrospective measurements are available for a few variables, longitudinal measurements of explanatory variables, or "covariates," are in general not available for the analysis of the history of vehicle transactions and holding durations which can be reconstructed (albeit incomplete due to reporting errors) from the survey results. Therefore in some of the models past behavior is explained in terms of the currently measured explanatory variables. This reduces the models' explanatory capability and limits their theoretical coherence. It is anticipated that the problem will be largely resolved once longitudinal observations become available from the second wave of the panel survey.

Another note which is due here concerns the fact that the types of explanatory variables used in the models are not restricted to ones whose future values can be forecast using the demographic forecasting micro-simulation model system being developed as part of the project. As noted above, the intention of the model development effort here is to explore the nature of vehicle transaction and holding duration decision, for which only a small number of studies can be found in the literature. In particular, the analysis aims at determining the variables that influence vehicle transaction and holding duration behavior in California. For this reason, all variables that are theoretically expected to influence the behavior or those that show significant effects in statistical analyses are included in the models. In this sense the models are mostly descriptive but not predictive. A set of predictive transaction timing models that are used in the micro-simulation is briefly described later in this section.

Section 7 is organized as follows. Sections 7.2 through 7.5 are concerned with vehicle holding durations. In Section 7.2, the above five alternative distribution functions are discussed. In particular, their theoretical bases as distribution functions for durations are discussed. Section 7.3 examines the statistical results for the five distributions and concludes that the Weibull distribution is the most suitable distribution. Section 7.4 discusses in detail the Weibull model of vehicle holding durations and Section 7.5 presents models by household vehicle ownership level. Section 7.6 presents ordered-response logit models of intended transaction timing. The ordered response logit model is first described briefly, then estimation results are discussed in Sections 7.6.2 through 7.6.5. Models for replacement timing by vehicle ownership level are discussed in Section 7.6.5. Predictive models of vehicle transaction timing that are based on Weibull distributions are briefly presented in Section 7.7. Section 7.8 is a brief summary.

7.2 Vehicle Holding Duration Models

Models of household vehicle holding durations are developed with the aim of better understanding household vehicle holding decision which underlies vehicle transaction behavior. In this modeling effort, vehicle holding durations are assumed to be a function of the attributes of:

the vehicle,
the primary user, and
the household.

The vintage (or, model year) of the vehicle is not used in the analysis as this variable is collinear with the holding duration when the vehicle was acquired new. The attributes of the other vehicles in the household fleet are not used as explanatory variables. This reflects the consideration that, given the utilization of a household vehicle (included in the duration models of this section), the holding duration of that vehicle is conditionally independent of the attributes of the other vehicles in the household fleet.

In this section, alternative distribution forms are first examined. Based on the results, the Weibull distribution is selected and duration models are developed by household vehicle ownership level.

7.2.1 Comparison of Alternative Distribution Functions

With the objective of determining the most suitable distribution function for household vehicle durations, alternative distribution functions are evaluated in this section. The analysis here assumes that the expected holding duration varies from household to household depending on their attributes as measured during the survey.

Therefore the best distribution identified here may not necessarily fit well the overall distribution of durations for all households.

The distributions examined here are: negative exponential distribution, Weibull distribution, log-normal distribution, log-logistic distribution, and generalized gamma distribution.

7.2.1.1 Negative Exponential Distribution

Let the probability density function (pdf) of T be

$$f_T(t) = \alpha \exp(-\alpha t), \quad t > 0$$

where

$$\alpha = e^{-\mu}.$$

This distribution function has a mean of $1/\alpha = e^{\mu}$.

Let W = *ln*T. Then the pdf of ξ is

$$g_W(w) = f_T(e^w)\,|dT/dW| = f_T(e^w)e^w = e^{(w-\mu)}\exp(-e^{(w-\mu)}).$$

This distribution is used to estimate the parameters of the distribution function with $\mu = \beta'X$, where β is the vector of coefficients and X is the vector of explanatory variables.

The negative exponential distribution is associated with the stochastic process of purely random events, i.e., events that occur with a probability that is invariant over time and whose occurrence neither depends on the past event history nor affects the occurrence of future events. The distribution represents the elapsed time between two successive events that are purely random, while the frequency of such events counted over intervals of a fixed length will have a Poisson distribution.

In survival analysis, the negative exponential distribution represents the basic distribution where the hazard function, h(t), is constant:

$$h(t) = f_T(t)/[1 - F_T(t)] = \alpha,$$

where $F_T(t)$ is the cumulative distribution function (CDF) of T. When durations have a negative exponential distribution, therefore a constant hazard function, they are duration independent, i.e., the fact that the duration in a state has reached a value, t_0, does not in any way influence the probability of staying in that state for another duration of time, say x. In this sense, these durations are "memoriless." More formally,

$$\Pr[T \leq t_0 + x \mid T > t_0] = \Pr[t_0 < T \leq t_0 + x]/\Pr[T > t_0]$$

$$= \{F_T(t_0+x) - F_T(t_0)\}/(1 - F_T(t_0)) = \{-\exp(-(t_0+x)) + \exp(-t_0)\}/\exp(-t_0)$$

$$= 1 - \exp(-x) = F(x), \quad x > 0, \ t_0 \geq 0$$

which is independent of t_0.

7.2.1.2 Weibull Distribution

T has a Weibull distribution if there exist γ (> 0), α (> 0) and ξ_0 such that

$$Y = [(T - \xi_0)/\alpha]^\gamma$$

has the standard negative exponential distribution, $f_Y(y) = e^{-y}$, $y > 0$. The pdf of T is

$$f_T(t) = (\gamma/\alpha)[(t - \xi_0)/\alpha]^{\gamma-1}\exp\{-[(t - \xi_0)/\alpha]^\gamma\}, \quad t > \xi_0.$$

For duration models we may assume $\xi_0 = 0$. Then

$$f_T(t) = (\gamma/\alpha)(t/\alpha)^{\gamma-1}\exp[-(t/\alpha)^\gamma], \quad t > 0.$$

$W = \ln T$ has

$$g_W(w) = (\gamma/\alpha)e^{\gamma w}\exp[-e^{\gamma w}/\alpha^\gamma], \quad -\infty < w < \infty.$$

Letting $\sigma = 1/\gamma$ and $\mu = \ln\alpha$, therefore $\alpha = e^\mu$ and $\alpha^\gamma = e^{\gamma\mu}$, we may rewrite this as

$$g_W(w) = \sigma^{-1}\exp[(w - \mu)/\sigma]\exp\{-\exp[(w - \mu)/\sigma]\}, \quad -\infty < w < \infty.$$

The mean and variance of T are, with $\xi_0 = 0$, given as

$$E[T] = \alpha\Gamma(1 + \gamma^{-1}) = e^\mu\Gamma(1 + \sigma), \text{ and}$$

$$\mathrm{Var}(T) = \alpha^2\{\Gamma(1 + 2\gamma^{-1}) - [\Gamma(1+\gamma^{-1})]^2\} = e^{2\mu}\{\Gamma(1 + 2\sigma) - [\Gamma(1+\sigma)]^2\}.$$

The distribution was used in 1939 by a Swedish physicist Waloddi Weibull to represent the distribution of the breaking strength of materials (Johnson & Kotz, 1970a). It can be argued that a distribution of this form arises when we consider the limiting distribution of the failure time of a system when it consists of n elements, each of which consists of γ redundant components. Each element fails when all γ

components fail, while the system fails if at least one element fails. Because of this linkage to failure times, the Weibull distribution is often used in survival analysis.

7.2.1.3 Log-normal Distribution

If there is a number, ξ_0, such that $Z = ln(T - \xi_0)$ has a normal distribution, then T is said to have a log-normal distribution. Let μ and σ be the expected value and standard deviation of Z, respectively. Let

$$U = \{ln(T - \xi_0) - \mu\}/\sigma$$

have the unit normal distribution. Then the pdf of T is

$$f_T(t) = [(t - \xi_0)(2\pi)^{-1/2}\sigma]^{-1}\exp\{-(ln(t - \xi_0) - \mu)^2/\sigma^2\}, \quad t > \xi_0.$$

With $\xi_0 = 0$, therefore $T > 0$, we have

$$f_T(t) = \exp\{-(ln t - \mu)^2/\sigma^2\}/\{(2\pi)^{1/2}\sigma t\}, \quad t > 0$$

The pdf of $W = lnT$ is

$$g_W(w) = \exp[-(w - \mu)^2/2\sigma^2]/\{(2\pi)^{1/2}\sigma\}, \quad -\infty < w < \infty.$$

And

$$E[W] = \mu \text{ and } Var(W) = \sigma^2.$$

Consider n independent random variables, $X_1, X_2, ..., X_n$, and let T_n be the product of the X's. Then lnT_n tends to be normal as n approaches infinity. The limiting distribution of T_n would then be log-normal. Thus a log-normal distribution is a theoretical distribution for durations if they are determined as a product of independent positive random variables.

7.2.1.4 Log-logistic Distribution

Let the CDF of T be

$$F_T(t) = 1 - 1/(1 + \alpha t^{\gamma})$$

and its pdf be

$$f_T(t) = \alpha\gamma t^{\gamma-1}/(1 + \alpha t^{\gamma})^2, \quad t > 0.$$

Then the pdf of $W = \mathit{ln}T$ is

$$g_W(w) = \alpha\gamma e^{\gamma w}/(1 + \alpha e^{\gamma w})^2, \quad -\infty < w < \infty.$$

Letting $\gamma = 1/\sigma$ and $\alpha = e^{-\mu/\sigma}$,

$$g_W(w) = \sigma^{-1}\exp[(w - \mu)/\sigma]/\{1 + \exp[(w - \mu)/\sigma]\}^2, \quad -\infty < w < \infty,$$

and

$$E[W] = \mu \text{ and } Var(W) = (\pi^2/3)\sigma^2.$$

The logistic function has often been used as a growth curve based on the differential equation,

$$dH/dx = c[H(x) - A][B - H(x)]$$

where c (> 0), A and B (> A) are constant parameters. Function H may be viewed to represent growth from a lower asymptote A to an upper asymptote B, and the rate of growth is proportional to the product of the distances from the two asymptotes. The above H is a CDF when A = 0 and B = 1. The logistic distribution has been shown to be the limiting distribution (as $n \rightarrow \infty$) of the standardized mid-range (average of largest and smallest sample values) of random samples of size n (Johnson and Kotz, 1970b).

7.2.1.5 Generalized Gamma Distribution

The standardized forms ($\mu = 0$, $\sigma = 1$) of the generalized gamma distributions used in this analysis can be written as:

$$f_T(t) = \delta(t^\delta/\kappa)^{1/\kappa}\exp(-t^\delta/\kappa)/t\Gamma(1/\kappa), \quad t > 0,$$

and

$$g_W(w) = \delta(e^{\delta w}/\kappa)^{1/\kappa}\exp(-e^{\delta w}/\kappa)/\Gamma(1/\kappa), \quad -\infty < w < \infty,$$

where $\kappa = \delta^2$. The mean and variance of T are:

$$E[T] = \delta^{2/\delta}\Gamma(1/\kappa + 1/\delta)/\Gamma(1/\kappa), \text{ and}$$

$$Var(T) = \delta^{4/\delta}\{\Gamma(1/\kappa + 2/\delta) - [\Gamma(1/\kappa + 1/\delta)]^2/\Gamma(1/\kappa)\}/\Gamma(1/\kappa).$$

The gamma distribution includes as its special case the chi-square distribution. The latter is the distribution of the sum of squares of independent unit random variables.

Namely, if U_1, U_2, ..., U_υ are independent unit normal random variables, then $X = U_1^2 + U_2^2 + \ldots + U_\upsilon^2$ has a χ^2 distribution with degrees of freedom υ. It also contains as its special case the Erlang distribution, which is the distribution of the sum of α (= 1, 2, ...) independent negative exponentially distributed random variables. Gamma distributions have been used as approximate distributions for chi-square statistics. Johnson and Kotz (1970a) note that "In applied work, gamma distributions give useful representations of many physical situations. They have been used to make realistic adjustments to exponential distributions in representing lifetimes in 'life-testing' situations. Of recent years, Weibull distributions have been more popular for this purpose, but this may not be permanent. The fact that sum of independent exponentially distributed random variables has a gamma distribution ... leads to the appearance of gamma distribution in the theory of random counters and other topics associated with random processes in time ...."

7.3 Statistical Results

Observed durations used in the analysis are: (i) the reported holding durations of the vehicles which the sample households had available at the time of the survey (up to six vehicles per household), and (ii) the reported holding durations of the up to six vehicles the sample households most recently had prior to the survey. The former set of durations are all censored. It is conceivable that the latter set of durations tended to involve inaccurate reporting and that those vehicles that were disposed in a distant past tended not to have been reported at all by the respondents. The latter may create the problem of sample selectivity. This issue is not addressed at this stage where the primary objective is to examine the suitability of the alternative distribution functions.

7.3.1 Vehicle Attributes

The variables used to describe the attributes of the vehicle include those concerned with vehicle ownership, form of acquisition, body type, and utilization (Table 7.1). As noted earlier, model year is not used as an explanatory variable as it is collinear with the holding duration when the vehicle was acquired brand new.

The most significant variable is the dummy variable indicating whether the vehicle was acquired as a used vehicle. The variable has a negative coefficient in all models, indicating that vehicles acquired as used vehicles tend to be kept for shorter durations. The variable has extremely high t-statistics whose absolute values exceed 10 in all models.

Table 7.1: Vehicle Holdings Duration Models with Alternative Distribution Functions

		EXPONENTIAL		LOG-NOMAL		LOG-LOGISTIC		GAMMA		WEIBULL	
Variable		Coef.	t	Coef.	t	Coef.	t	Coef.	t	Coef.	t
Intercept		5.833	22.52	5.480	27.90	5.095	29.02	5.231	31.65	5.240	30.91
Vehicle Attributes											
The vehicle is	Leased	-0.196	-0.77	-0.297	-1.57	-0.321	-1.88	-0.321	-2.01	-0.318	-1.93
	Company car	-0.295	-0.85	-0.475	-1.93	-0.471	-2.08	-0.403	-1.85	-0.422	-1.91
The vehicle was	Acquired as a used vehicle	-0.877	-11.01	-0.797	-13.11	-0.686	-12.70	-0.615	-11.43	-0.643	-12.24
The vehicle was	Addition to the household fle	0.090	0.93	0.101	1.36	0.092	1.39	0.074	1.23	0.079	1.27
Vehicle class	Mini	0.829	3.34	0.638	3.68	0.616	3.86	0.583	3.66	0.597	3.75
	Subcompact	0.343	2.59	0.261	2.61	0.282	3.20	0.314	3.80	0.304	3.58
	Midsize	0.086	0.61	0.026	0.25	0.058	0.61	0.118	1.34	0.099	1.10
	Large	0.167	0.77	0.102	0.64	0.146	1.02	0.165	1.22	0.162	1.17
	Luxury	0.260	1.62	0.210	1.76	0.209	1.96	0.234	2.33	0.227	2.21
	Sports car	0.390	2.61	0.313	2.77	0.349	3.48	0.392	4.20	0.380	3.96
	Small pick-up truck	0.260	1.50	0.214	1.67	0.194	1.70	0.187	1.72	0.192	1.73
	Standard pick-up truck	0.230	1.19	0.123	0.87	0.182	1.41	0.218	1.79	0.210	1.69
	Mini van	-0.541	-3.44	-0.387	-2.98	-0.335	-3.07	-0.363	-3.78	-0.360	-3.56
	Standard van	0.331	1.08	0.271	1.18	0.285	1.38	0.335	1.74	0.324	1.64
	Utility vehicle	-0.225	-1.46	-0.218	-1.81	-0.149	-1.42	-0.128	-1.35	-0.136	-1.38
Miles driven per year (in 10,000mi.)		-0.171	-0.89	-0.230	-1.66	-0.204	-1.66	-0.185	-1.62	-0.194	-1.64
Average miles driven per day		-0.0057	-2.30	-0.0029	-1.30	-0.0039	-2.17	-0.0044	-2.90	-0.0043	-2.68
Attributes of Primary User											
Age		0.012	3.36	0.012	4.53	0.012	5.01	0.011	4.95	0.011	4.95
Gender	Female	-0.026	-0.33	-0.039	-0.66	-0.066	-1.24	-0.086	-1.72	-0.077	-1.51
Personal annual incom	$100,000 or more	-0.682	-3.00	-0.593	-3.28	-0.488	-3.08	-0.417	-2.98	-0.443	-3.04
Have a second job		0.144	0.97	0.041	0.37	0.090	0.90	0.117	1.27	0.109	1.14
Work type (regularly)	Other	0.214	1.67	0.184	1.92	0.134	1.57	0.120	1.48	0.126	1.53
Work type (irregularly)	Outside sales	-0.392	-1.08	-0.345	-1.11	-0.346	-1.36	-0.369	-1.67	-0.357	-1.53
	Other	0.210	1.13	0.101	0.69	0.161	1.26	0.170	1.47	0.169	1.41
	Don't work	0.237	2.03	0.135	1.57	0.138	1.79	0.129	1.76	0.134	1.78
Attributes of Household											
Number of drivers		-0.169	-2.61	-0.115	-2.39	-0.130	-2.96	-0.127	-3.10	-0.130	-3.10
Number of vehicles		0.333	5.42	0.202	4.76	0.233	5.70	0.241	6.24	0.240	6.08
Years in area	< 1 year	-0.377	-2.88	-0.391	-3.92	-0.352	-3.97	-0.331	-4.04	-0.336	-3.99
	1~2 years	-0.283	-2.40	-0.263	-2.91	-0.250	-3.14	-0.271	-3.71	-0.263	-3.50
	3~4 years	-0.096	-0.81	-0.083	-0.93	-0.092	-1.17	-0.128	-1.72	-0.115	-1.51
	5~6 years	-0.189	-1.57	-0.182	-1.98	-0.128	-1.56	-0.110	-1.49	-0.117	-1.52
Rent Home		-0.160	-1.67	-0.184	-2.50	-0.149	-2.32	-0.134	-2.26	-0.138	-2.27
Household annual inco	Less than $15,000	0.497	2.46	0.392	2.70	0.341	2.58	0.312	2.43	0.324	2.50
	$61,000 to $75,000	-0.115	-0.94	-0.057	-0.60	-0.091	-1.11	-0.119	-1.55	-0.112	-1.41
	$76,000 to $100,000	-0.028	-0.21	-0.015	-0.15	-0.041	-0.46	-0.048	-0.58	-0.046	-0.54
	$101,000 to $125,000	0.320	1.41	0.178	1.13	0.152	1.04	0.127	0.87	0.139	0.96
	$125,000 and over	0.048	0.23	0.053	0.33	0.010	0.07	-0.048	-0.36	-0.027	-0.20
	Don't know	-0.188	-0.92	-0.301	-2.00	-0.203	-1.46	-0.118	-0.92	-0.142	-1.08
	Refused to answer	0.150	0.88	0.080	0.64	0.085	0.74	0.141	1.31	0.124	1.14
								1.266	6.90		
Shape Parameter (1/g)		1.000		1.275	8.350*	0.590	-24.83*	0.535	-7.22*	0.639	-20.94*
N		7136		7136		7136		7136		7136	
L(C)		-2580.8		-2571.3		-2537.4		-2533.1		-2533.4	
L(b)		-2403.1		-2345.7		-2301.1		-2291.8		-2293.3	
-2[L(C)-L(b)] (df)		355.5	(39)	451.1	(39)	472.6	(39)	482.5	(40)	480.2	(39)
Censored observations = 6399								*The t- Statistics is for H0 : b=1			
Noncensored observations = 737											

Body type is also significantly associated with holding duration. The estimated coefficients offer consistent indications that: mini, subcompact and sports cars tend to be held longer (the coefficient of the compact car is set to 0 as a reference). These variables have significant (at $\alpha = 1\%$) positive coefficients in all models. Luxury cars also tend to be held longer as indicated by the significant (at $\alpha = 5\%$) positive coefficients of the log-logistic, gamma and Weibull models. Minivans, on the other hand, tend to be held for shorter periods. This could be due to the fact that minivans are relatively new in the market. In fact standard-size vans have not necessarily significant but positive coefficients in all models. Luxury cars also have significant positive coefficients in some models.

Holding duration is negatively associated with vehicle utilization. Vehicles with greater annual mileage or those with greater miles driven per day tend to be held shorter.

7.3.2 Attributes of Primary User

The estimation results offer a strong indication that the age of the primary user of a vehicle is positively associate with the holding duration. It is consistent with our expectation that younger individuals tend to undergo frequent changes in life cycle, employment, and residential location, which tend to induce vehicle transaction; while older individuals tend to undergo fewer changes and less frequent vehicle transactions and therefore longer holding durations. The age variable is very significant (at $\alpha = 1\%$) with their t-statistics exceeding 4.5 in all models but the exponential distribution model.

Primary users whose personal annual incomes exceed $100,000 tend to keep their vehicles shorter. The coefficient estimates for the dummy indicator of high personal income are consistently significant at $\alpha = 1\%$. There are indications that women tend to have shorter durations, but the gender variable is not consistently significant.

7.3.3 Attributes of Household

Several household attributes show consistent and significant association with vehicle holding duration in all five models. Number of drivers is associated with vehicle holding duration negatively, number of vehicles positively, years of residence in the area negatively, renting home negatively, and low household income (less than $15,000) positively.

Number of vehicles has extremely significant positive coefficients. The positive coefficient estimates of number of vehicles combined with the negative coefficient estimates of number of drivers imply that a household with a vehicle fleet which is large relative to the number of drivers tends to hold each vehicle longer.

The negative association between years of residence and vehicle holding duration indicates the relation suggested earlier that those with frequent changes in their life tend to have shorter vehicle holding durations and frequent vehicle transactions.

The result that low income households tend to hold their vehicles longer implies the limited economical abilities these households have to replace their vehicles to best serve their needs. The coefficient estimates of income dummy variables for other income ranges do not show consistent tendencies.

7.3.4 Comparison of the Five Distribution Functions

The five distributions can be ranked as follows according to the log-likelihood values:

gamma	-2291.8
Weibull	-2293.3
log-logistic	-2301.1
log-normal	-2345.7
negative exponential	-2403.1

The gamma and Weibull distributions have the largest log-likelihood values that are practically identical; they fit the data almost equally well and outperform the other three distributions. Note the significant gaps in log-likelihood statistics between the log-logistic distribution and the log-normal distribution, and between the log-normal distribution and the exponential distribution.

The estimation results offer strong evidence that the distribution of vehicle holding durations is not negative exponential. Vehicle holding duration is not an outcome of a memoriless decision process, nor is vehicle transaction invoked by purely random events. In fact the estimate of the Weibull shape parameter (an estimate of $1/\gamma$ is shown in the Table 7.1) indicates that vehicle holding duration is duration dependent.

The best three distributions that have similar log-likelihood statistics -- gamma, Weibull and log-logistic -- each offer a certain theoretical basis as a distribution for durations. In particular, the gamma distribution represents a duration as a sum of several random variables. The Weibull distribution, on the other hand, has an explicit connection to the failure time of a system when it consists of several components, all of which must function for the system to function. In terms of the goodness of fit, there are practically no differences between these two distributions which offer the best fit. Considering the computational ease it offers as well as its theoretical foundation, the Weibull distribution is chosen here for further analysis.

7.4 Weibull Model of Vehicle Holding Durations

Based on the comparative analysis of the five alternative distribution functions, the Weibull distribution is selected as the most suitable distribution function. Using this distribution, household vehicle holding duration models are estimated by the level of vehicle ownership. As noted earlier, the models are formulated using the attributes of the vehicle, its primary user, and the household, but the age of the vehicle is not used. Considering the endogeneity, the attributes of the other vehicles in the household are not included as explanatory variables. It is noted that the attributes of the vehicle user and household are as measured at the time of the survey, and do not necessarily correspond to user and household attributes when acquisition or disposal decisions were made. It is anticipated that better explanatory variables will be available when the second-wave survey results are available. Weibull models of household vehicle holding durations are shown in Table 7.2 by vehicle ownership level. Salient results are summarized below, first for the model for all vehicle ownership levels, then by ownership level.

7.4.1 Vehicle Attributes

The Weibull model presented in Table 7.1 along with the other four models is discussed in more detail here. As noted earlier, the ownership of a vehicle available to the household is associated with its holding duration. As indicated by their negative coefficient estimates, leased vehicles and company cars both tend to be held shorter compared with vehicles owned by the household (suppressed in the model for normalization). The coefficients are both statistically significantly different from zero at α = 10%.

The coefficient of the dummy variable indicating that the vehicle was acquired as a used vehicle is negative and, with a t-statistic value of -12.24, is highly significant. Clearly households tend to hold their vehicles shorter if they acquire them as used vehicles compared with vehicles they acquire brand new. With exp(-0.643) = 0.526, the coefficient estimates implies that, *ceteris paribus*, a vehicle acquired used tends to be held approximately half as long as one acquired brand new.

Smaller cars (minis and subcompacts) tend to be held longer. The coefficient estimates indicated that, *ceteris paribus*, the mean holding duration of mini cars is 82% longer, and that of subcompact cars 36% longer, than that of compact cars, which are suppressed in the model as a reference vehicle body type. Likewise, luxury vehicles and sports cars tend to be kept longer, with their mean holding durations longer than that of compact cars by 25% and 46%, respectively. The estimation results also indicate that mini-vans tend to have shorter holding durations. This, however, may be due to the fact that mini-vans were introduced into the market relatively recently.

Table 7.2: Weibull Models of Vehicle Holdings Durations by Vehicle Ownership Level

		1 VEHICLE		2 VEHICLES		3+ VEHICLES	
Variable		Coef.	t	Coef.	t	Coef.	t
Intercept		5.108	16.06	5.474	23.89	6.554	13.10
Vehicle Attributes							
The vehicle is	Leased	-0.388	-1.36	-0.457	-2.39	1.021	1.25
	Company car	-0.952	-1.48	-0.405	-1.50	0	
The vehicle was	Acquired as a used vehic	-0.738	-7.10	-0.497	-7.74	-0.919	-6.21
The vehicle was	Addition to household fle	-0.179	-1.34			0.427	2.61
	First vehicle for househol	-0.139	-0.97			0	
Vehicle class	Mini	0.379	1.46	0.569	2.88	0.820	1.44
	Subcompact	0.159	1.04	0.368	3.51	0.244	0.90
	Midsize	-0.181	-1.08	0.202	1.83	0.144	0.52
	Large	0.141	0.52	0.052	0.31	0.498	1.26
	Luxury	0.312	1.32	0.224	1.79	0.133	0.49
	Sports car	0.352	2.00	0.383	3.18	0.406	1.45
	Small pick-up truck	0.037	0.16	0.183	1.38	0.350	1.05
	Standard pick-up truck	0.248	0.74	0.199	1.33	0.473	1.53
	Minivan	-0.200	-0.86	-0.290	-2.44	-0.671	-2.34
	Standard van	0.139	0.36	0.585	1.91	0.227	0.52
	Compact utility vehicle	-0.247	-0.97	-0.250	-1.79	-0.199	-0.62
	Standard utility vehicle	-0.337	-1.12	0.191	1.01	-0.271	-0.79
	Mini utility vehicle	-0.458	-0.72	0.168	0.29	-1.557	-1.80
Miles driven per year (in10,000 mi.)				-0.190	-1.39	-0.455	-1.35
Most miles driven per day				-0.0014	-1.90		
Average miles driven per day						-0.0079	-3.16
Attributes of Primary User							
Age		0.016	3.51	0.014	4.56	0.0080	1.15
Gender	Female			-0.147	-2.21		
Employment	Self-employed, full-time			-0.164	-1.32		
	Self-employed, part-time			-0.279	-1.39		
	Retired, not employed	-0.228	-1.67			0.328	1.36
	Student	0.624	1.57	-0.178	-1.88	1.204	1.88
Education	Some highschool or less	0.580	1.52			-0.041	-0.12
	Highschool graduate					0.271	1.54
	Four-year college degree	0.151	1.32			0.198	0.99
	Graduate work or degree	0.206	1.41			0.019	0.09
Number of days per week in school		-0.161	-1.77			-0.174	-1.15
Personal annual income	No answer					0.723	1.50
	$50,000 to $100,000			-0.152	-1.64	0.301	1.30
	$100,000 or more			-0.254	-1.32	-0.937	-3.03
Need a car for job						0.391	2.28
Have a second job		0.260	1.46	0.180	1.39		

Table 7.2 continued

Work type (regularly)	School			-0.030	-0.30	-0.333	-1.52
	Retail			0.078	0.56		
	Manufacturing			0.151	0.89		
	Construction			0.078	0.36		
	Outside sales			0.173	0.57		
	Other			0.154	1.62	-0.418	-2.47
Work type(irregularly)	Outside sales	-1.173	-3.42	-0.066	-0.21		
	Free-lance work	-0.906	-2.85	-0.337	-0.96		
	Self-employed	0.328	0.74	0.092	0.50		
	Other	0.298	1.00	0.393	1.77		
	Don't know			0.153	1.54		
Attributes of Household							
Household size						-0.139	-2.39
Number of members 16 or older				-0.137	-2.79		
Number of drivers		-0.165	-1.93				
Years in area	< 1 year	-0.163	-1.17	-0.456	-4.42	-0.307	-0.97
	1 to 2 years	-0.410	-3.31	-0.153	-1.55	-0.392	-1.77
	3 to 4 years			-0.194	-2.12	0.021	0.10
	5 to 6 years			-0.134	-1.44	-0.569	-2.58
Rent Home		0.164	1.46	-0.221	-2.87	-0.488	-2.33
Dwelling type	Apartmemt					0.850	1.74
	Condominium	0.228	1.45				
	Mobile home			-0.473	-2.15	1.122	1.84
	Group quarters	-0.826	-1.31				
Parking at home	Carport			0.139	1.39	-0.733	-3.23
	None	0.208	1.59	0.104	0.88	-0.412	-1.74
Household annual income	Less than $15,000	0.530	2.66	0.328	1.38	0.303	0.53
	$15,000 to $30,000	-0.005	-0.03	0.124	1.14	-0.118	-0.45
	$31,000 to $45,000	0.066	0.45	0.023	0.28	0.444	1.84
	$61,000 to $75,000	-0.086	-0.38	0.001	0.01	0.052	0.25
	$76,000 to $100,000	0.220	0.85	-0.073	-0.73	0.146	0.62
	$101,000 to $125,000	-0.613	-1.54	0.263	1.48	0.090	0.28
	$125,000 and over	-0.396	-0.87	-0.031	-0.19	0.401	1.36
	Don't know	0.622	1.27				
Shape Parameter (1/g)		0.607	-11.95*	0.577	-20.40*	0.749	-5.27*
N		1453		3744		1956	
L(C)		-611.4		-1330.2		-565.2	
L(b)		-525.9		-1181.5		-485.9	
-2[L(C)-L(b)] (df)		171.1 (45)		297.2 (53)		158.7 (49)	
Censored observations		1263		3337		1816	
Noncensored observations		190		407		140	
				*The t- Statistics is for H0 : b=1			

As anticipated, vehicles with higher vehicle-miles driven tend to be held for shorter durations. The coefficient estimate, however, is only marginally significant at α = 10%. More significant is the variable representing the reported average miles driven per day (significant at 1%). Theoretically these two variables are linearly related to each other. Perceived values, however, evidently do not have this collinearity and produced results with different magnitudes.

7.4.2 Attributes of the Primary User

The age of the primary user of the vehicle is positively associated with the holding duration. The estimated coefficient of 0.011 implies that, *ceteris paribus*, a vehicle will be held on average 11.6% longer when its primary user is 10 years older. The coefficient estimate of the gender variable indicates that female primary users tend to hold their vehicles shorter, on average by 7.4%. But the coefficient is not significant at α = 10%.

Primary users with high personal incomes ($100,000 annually or more) tend to have shorter holding durations. The coefficient estimate of -0.443 is significant at α = 1% and indicates that vehicle holding durations of these higher income individuals are on average 36% shorter than those of individuals in other income classes.

Holding durations vary depending on the type of work of the primary user. For example, individuals in occasional (once a week) outside sales tend to have shorter holding durations while those in manufacturing, (more frequent) outside sales, free-lance work, self-employed or in other categories tend to have longer holding durations. The coefficient estimates of these variables are, however, in general not significant.

7.4.3 Attributes of the Household

Households with more drivers tend to hold their vehicles shorter, while those with more vehicles tend to hold them longer. The coefficient estimates of these variables are both significant at α = 1%. The coefficient estimates imply that a household with one more driver tends to have a mean vehicle holding duration that is 12.2% shorter, while a household with one more vehicle tends to have a mean holding duration longer by 27%. If a household holds as many vehicles as its drivers, then having one more driver and one more vehicle would imply a mean vehicle holding duration that is 11.6% longer.

The set of dummy variables indicating the number of years spent in the area shows a clear tendency that the duration of residence is positively associated with vehicle holding durations (the category for "seven years or more" is dropped in the model to have a reference coefficient value of 0). The result suggests that residentially mobile households, which probably tend to be going through periods of demographic and socio-economic changes, tend to make frequent vehicle transactions. The negative coefficient of the variable for "rent home" (significant at 5%) is consistent with this result.

Households in the lowest income category (less than $15,000 annually) tend to have mean vehicle holding durations that are 38% longer than their higher income counterparts. The coefficient estimate is significant at 2%. Other than this, however, the coefficients of the income category variables do not show a clear and significant tendency.

7.4.4 Shape Parameter

The shape parameter, $1/\gamma$, is estimated as 0.639, which is significantly different from 1.0. The Weibull distribution reduces to a negative exponential distribution when γ = 1, and represents "memoriless" vehicle holding behavior. Namely, the decision to dispose of or replace a vehicle is independent of the duration that the vehicle has been held so far. The result that γ is significantly (at α = 1%, one-tail test) greater than 1 offers empirical evidence that vehicle holding behavior is *positively* duration dependent. Vehicle holding behavior does have strong memory; the longer one has held a vehicle, the more likely will the vehicle be replaced or disposed of in the near future.

7.5 Weibull Vehicle Holding Durations Models by Vehicle Ownership Level

Weibull vehicle holding duration models are estimated by household vehicle ownership level (one vehicle, two vehicle, three or more vehicles) and summarized in Table 7.2. Consistent with the overall model, the models for one-vehicle and two-vehicle households have negative coefficient estimates for the indicators of a leased vehicle and a company car, although they are not all significant. This is not the case for the model for households with three or more vehicles.

Vehicles acquired as used vehicles tend to be held for shorter durations at all levels of vehicle ownership. Consistent with the overall model, the coefficient estimates of the indicator of a used car are negative and significant at 1% in all models. The tendency found in the overall model that a vehicle acquired as an additional vehicle tends to be held longer, however, is found for households with three or more vehicles only.

The association between annual vehicle-miles and holding duration is weak in these models by vehicle ownership level. Average mileage per day is significant only in the model for households with three or more vehicles. Models show associations between vehicle body type and holding duration that are similar to those found in the overall model.

The age of the primary user again can be found to be positively associated with the holding duration. It is notable that the age variable is not significant in the model for households with three or more vehicles. The tendency that a vehicle whose primary driver is a female tends to be held shorter can be found for two-vehicle households only. The result suggests that gender differences in vehicle holding are most pronounced in two-vehicle households which most likely comprise younger male-female couples without driving-age children. The negative association between personal income and vehicle holding duration can be clearly observed for households with three or more vehicles.

Of the household attributes, the set of dummy variables indicating years of residence in the area no longer exhibits clear tendencies in these models by vehicle ownership. Renters with two vehicles or three or more vehicles have shorter holding durations, but not renters with one vehicles. As before, household income does not show any meaningful tendencies.

Quite notably the estimate of the shape parameter, $1/\gamma$, is closer to 1 in the model for households with three or more vehicles. The result suggests that vehicle transaction behavior becomes more random and the distribution of holding durations approaches a negative exponential distribution as the number of vehicles in the household fleet increases.

7.5.1 Summary

The comparative analysis of alternative distribution functions indicated that the Weibull distribution best fits the vehicle holding durations data obtained from the first-wave of the SCE/PG&E survey. Salient findings include that household vehicles acquired as used vehicles tend to be held for shorter durations. The same tendency can be found for leased vehicles. As one might expect, vehicles that are driven more tend to be held for shorter durations. The results also indicated that mini and subcompact, luxury and sports cars tend to be kept longer, while mini-vans tend to have shorter holding durations. Of the attributes of the primary user of a vehicle, age is most important and is positively associated with holding duration; older individuals tend to keep their cars longer. The results also show that overall the duration of residence in an area is positively associated with vehicle holding duration, and those who rent their homes tend to have shorter vehicle holding durations.

7.6 Ordered-Response Logit Models of Stated Transaction Timing

The analysis of this section examines the response to the stated-preference question on the timing of the next intended vehicle transaction. The analysis is motivated by the desire to explain a household's intended timing for the next vehicle transaction in terms of the characteristics of the vehicle fleet, vehicle utilization, and the attributes of the household. Ordered-response logit models are applied to response categories used in the survey questionnaire.

7.6.1 Models of Stated Transaction Timing by Transaction Type

The dependent variable of this analysis comes from the section of the first-wave survey which starts with: "The next time your household sells or buys a vehicle, which of the following is most likely?" This is followed by a set of questions, which varies depending on the transaction type that the respondent gives, and includes "When are you most likely to replace this vehicle?" or "sell or scrap," or "buy or lease". The response categories used in these questions are (1) less than a year, (2) 1 - 2 years, (3) 3 - 4 years, (4) More than 4 years, (5) Don't know, and (6) Refused. In this analysis the sub-sample of those who indicated intended transaction timings are included, and ordered-response logit models are applied to the first four response categories.

The ordered-response logit model can be described as follows. Let

$$F(t) = 1/(1 + \exp(-t)),$$

which is the cumulative standard logistic distribution function which has a mean of 0 and a variance of $\pi^2/3$. Let k be the number of categories that the ordered-response dependent variable, Y_i, will assume; let x_i be the vector of explanatory variables; β be the vector of coefficients; and α_j, $j = 1, \ldots, k - 1$, be threshold parameters. Then the probability that the response will be j is given as:

$$\Pr[Y_i = j | x_i] = \begin{cases} F(\alpha_1 + \beta' x_i) & j = 1 \\ F(\alpha_j + \beta' x_i) - F(\alpha_{j-1} + \beta' x_i) & j = 1, \ldots, k - 1 \\ 1 - F(\alpha_{k-1} + \beta' x_i) & j = k \end{cases}$$

The models include as their explanatory variables the attributes of the household and those of its vehicle fleet. The variables that describe the household fleet are:

the average age of the vehicles,
the age of the oldest vehicle in the fleet,
the fraction of vehicles acquired brand new,
the total annual VMT by the fleet, and
the average holding duration of the vehicles.

Estimation results are summarized in Table 7.3.

Table 7.3: Ordered-response Logit Models of Stated Transactions Timing by Transactions Type for All Vehicle Ownership Levels

		REPLACEMENT		ACQUISITION		DISPOSE	
Variable		Coef.	t	Coef.	t	Coef.	t
Intercept	α1	-1.993	-6.05	-0.294	-0.32	-8.835	-4.64
	α2	-0.565	-1.72	2.522	2.74	-7.214	-3.91
	α3	0.882	2.69	4.418	4.39	-6.146	-3.37
Vehicle Attributes							
Average age of the vehicles		0.149	6.80				
Age of the newest vehicle in fleet						0.092	1.91
Age of the oldest vehicle in fleet		-0.061	-4.62				
Number of leased vehicle in current household fleet		0.247	1.98				
Number of company vehicle in current household fleet		0.518	2.86				
Number of new vehicle in current household fleet						0.693	1.41
Fraction of new vehicles in current household fleet		-0.296	-2.64			-1.768	-1.57
Total annual VMT (in 1,000 miles)		0.124	5.12	0.339	2.38	0.419	3.02
Average annual VMT (in 1,000 miles)				-0.451	-2.02		
Average number of months since acquisitions (in 100 m		0.252	1.90				
Attributes of Household							
Household Type	Couple			-0.602	-2.02	2.098	2.98
	Couple & children 6 to 15 yrs. old			-0.746	-2.08	4.664	4.14
	Couple & children 16 to 20 yrs. old					1.508	1.76
	Single parent			-0.927	-1.52	3.218	2.69
	Couple & other adults	0.504	2.85	0.969	1.33		
	Single person	-0.542	-5.32	-1.368	-3.17	1.831	1.75
Presence of member leaving household		-0.324	-2.70	0.596	1.53		
Number of vehicles				-0.535	-1.77	-1.285	-2.72
Number of vehicles per driver		0.682	3.51	0.642	1.65	2.689	3.27
More vehicles than drivers in household		-0.450	-2.35	-0.900	-1.77	1.435	1.95
Number of workers in househhold						0.583	2.22
Number of fulltime workers in household		0.120	2.26				
Number of children15 yrs. old or younger						-0.629	-1.88
Number of members 16 yrs. old or older						1.853	3.36
Number of members 16-20 yrs. old				-0.303	-1.38		
Years in area	4 years or less	0.107	1.37				
Rent home		0.171	1.91				
Dwelling type	Duplex	-0.334	-1.77				
Parking at home	Carport			-0.443	-1.51		
Household annual inco	Less than $15,000					-2.300	-3.47
	Less than $30,000	-0.284	-3.01	-0.672	-2.59		
	$46,000 and over	-0.250	-2.98				
L(0)		-3679.2		-528.2		-202.4	
L(C)		-3627.9		-361.8		-187.6	
L(b)		-3478.7		-341.0		-136.1	
-2[L(0)-L(b)] (df)		401.0(21)		374.4(17)		132.6(19)	
-2[L(C)-L(b)] (df)		298.4(18)		41.6(14)		103.0(16)	
N		2654		381		146	
1. < 1 Year		656		176		64	
2. 1~2 Years		810		175		33	
3. 3~4 Years		724		25		18	
4. > 4 Years		464		5		31	

7.6.2 Replacement Timing Model

Most of the descriptors of household vehicle fleets are significant in the replacement timing model. The most significant are two variables describing the age of the vehicles. The coefficient of the average age of the vehicles is positive, while that of the age of the oldest vehicle is negative. Thus those owners of older fleets tend not to plan a transaction in the near future, given that they plan to replace a vehicle. The age of the oldest vehicle in the fleet, however, tends to motivate a transaction planned in the nearer future.

Both total annual VMT and average holding duration have positive coefficients, implying that heavier users of their vehicle fleets, and those who have kept their vehicles longer, tend to plan a transaction at a later time. The estimation results altogether suggest the bipolar tendency that households can be classified into two groups, one group comprising households that replace their vehicles frequently, and the other group comprising households that plan to hold their vehicles longer and make extensive use of them.

Households which rent their homes, those living in mobile homes, and those with members who have left, tend to plan to replace their vehicles in the farther future. Single persons, on the other hand, tend to plan to replace in the nearer future. The coefficients of household type variables also suggest that households in earlier stages of life cycle tend to plan to replace in the nearer future. The variables for household income categories have significant coefficients, but do not exhibit clear tendencies.

7.6.3 Acquisition Timing Model

Total annual VMT and average annual VMT per vehicle are both significant ($\alpha = 5\%$) with opposite signs. The two coefficients combined indicate that a household tends to plan to acquire a vehicle sooner if the vehicles in the household fleet are used more heavily, but less so when there are more vehicles in the fleet. The coefficient of the number of vehicles, however, is negative and significant at 10%.

Significant coefficients of the household type variables offer strong indications that younger households (couples, couples with children between 6 and 15 years old, and single persons) tend to plan to acquire a vehicle sooner.

7.6.4 Disposal Timing Model

Consistent with the replacement timing model, the coefficient of total annual VMT is positive and significant ($\alpha = 1\%$). Age of the newest vehicle in the fleet also has a positive coefficient (significant at $\alpha = 10\%$). The more the vehicles in the household fleet are utilized, and the older the vehicles are, the later will the disposal of a vehicle tend to be. Number of vehicles has a significant negative coefficient (α = 1%); *ceteris paribus* households with more vehicles tend to intend sooner disposal of vehicles.

The estimation results offer strong indications that couples, couples with children between 6 and 15 years old and single parents tend to plan later disposal. Likewise, the number of workers and the number of household members of at least 16 years old both have positive coefficients (at 5% and 1%, respectively). Notably, having more vehicles than drivers does not appear to motivate the household to plan to dispose of its vehicles quicker. Also notable is the negative coefficient for the lowest income category (significant at α = 1%). Lower income households intend to dispose of their vehicles sooner, which is an economically rational behavior.

7.6.5 Models of Vehicle Replacement Timing by Vehicle Ownership Level

Ordered-response logit models of vehicle replacement timing are developed by vehicle ownership level. The results are summarized in Table 7.4. (Unfortunately the available sample size is not sufficient to develop models for acquisition and disposal.)

Vehicle Attributes: One- and Two-Vehicle Households
Average age of the vehicles in the household fleet, which is a highly significant variable in the replacement timing model for all vehicle ownership levels, is also significant (at 1%) in the models for one- and two-vehicle households. Total annual VMT also has significant positive coefficients in these models (at 5% in the model for one-vehicle households, at 1% in the model for two-vehicle households). *Ceteris paribus*, households with older fleets and mobile households tend to plan vehicle replacement later in the future.

The model for one-vehicle households indicates that a one-vehicle household tends to plan to replace the vehicle later if it is a company car, and sooner if it is new. The model for two-vehicle households indicates that a two-vehicle household tends to plan a sooner replacement when its older vehicle is more aged, and when its fleet has more replacement vehicles.

Overall the models for one- and two-vehicle households again indicate the tendency that households tend to split into two extreme groups, those who tend to replace vehicles at shorter intervals, and those who tend to hold their vehicles for long durations.

Table 7.4: Ordered-response Logit Models of Stated Replacement Timing by Vehicle Ownership Level

		1VEHICLE		2VEHICLES		3+ VEHICLES	
Variable		Coef.	t	Coef.	t	Coef.	t
Intercept	α1	-1.449	-2.85	3.625	1.50	-2.096	-2.50
	α2	-0.067	-0.13	5.110	2.12	-0.607	-0.73
	α3	1.373	2.70	6.599	2.73	0.970	1.16
Vehicle Attributes							
Average age of the vehicles		0.101	5.00	0.175	4.65		
Age of the oldest vehicle in fleet				-0.082	-3.57		
Age of the newest vehicle in current household fleet						0.059	2.02
Fraction of new vehicles in current household fleet				-4.460	-1.88	-0.514	-1.67
Number of company vehicles in current household fleet		2.454	2.76	0.413	1.71	0.439	1.49
Number of replaced vehicles in current household fleet				-0.164	-1.94		
Number of new vehicles in current household fleet		-0.412	-2.33				
Number of used vehicles in current household fleet				-2.185	-1.84		
Total annual VMT (in 1,000 miles)		0.151	2.30	0.131	3.56		
Average number of months since acquisitions (in 100 mo.)		0.333	1.47	0.338	1.76	-0.454	-1.68
Attributes of Household							
Household type	Couple			-0.487	-2.68		
	Couple & children 1 to 5 yrs. old	0.948	2.15	-0.575	-2.62		
	Couple & children 6 to 15 yrs. old			-0.441	-2.21		
	Couple & other adults			0.563	1.69	0.343	1.37
	Single person			-1.181	-3.42		
	Multi adult					-0.957	-2.21
Presence of member leaving household		-0.510	-2.28				
Number of drivers in household		-0.290	-1.63			0.220	1.24
More drivers than vehicles in household				-0.562	-1.86		
More vehicles than drivers in household						-0.375	-1.33
Number of vehicles per driver						0.707	2.32
Number of fulltime workers in household				0.150	2.04	0.148	1.35
Number of members 16 yrs. old or older		0.308	1.75	-0.327	-1.97		
Years in area	1 year or less	0.228	1.39				
	1~2 years					0.289	1.21
Rent home				0.303	2.47		
Dwelling type	Apartment	0.383	2.46				
	Condominium or townhome	0.440	2.24	-0.293	-1.60		
	Duplex					-1.198	-1.71
	Mobile home			0.936	2.59	-1.035	-1.25
Parking at home	Carport	-0.229	-1.49				
	Neither garage not carport			-0.380	-2.09		
Household annual inco	Less than $30,000	-0.429	-2.49	-0.283	-2.15		
	$30,000 to $46,000					-0.174	-0.79
	$46,000 and over	-0.324	-2.11	-0.256	-2.22		
L(0)		-1127.1		-1931.1		-621.1	
L(C)		-1116.6		-1884.5		-584.1	
L(b)		-1050.3		-1818.0		-568.2	
-2[L(0)-L(b)] (df)		153.6(18)		226.2(25)		105.8(17)	
-2[L(C)-L(b)] (df)		132.6(15)		133.0(22)		31.8(14)	
N		813		1393		448	
1. < 1 Year		153		358		145	
2. 1~2 Years		209		449		152	
3. 3~4 Years		242		376		106	
4. > 4 Years		209		210		45	

Vehicle Attributes: Households with Three or More Vehicles
None of the variables describing the characteristics of the household fleet is significant (at 5%) in the replacement timing model for households with three or more vehicles. The only exception is the age of the newest vehicle in the current household fleet, which has a positive coefficient estimate. It appears that intended timing for vehicle replacement is less dependent on household fleet attributes when the number of vehicles in the fleet exceeds two.

Household attributes: One- and Two-Vehicle Households
The dwelling type is a significant contributing factor in the model for one-vehicle households while the household type variables are highly significant in the model for two-vehicle households. One-vehicle households in apartments or condominiums, and two-vehicle households who rent their homes or are in mobile homes, tend to plan the next replacement farther in the future. Two-vehicle households which comprise a single person, a couple, and a couple and children, on the other hand, tend to plan a replacement in the nearer future. Notable is the significant (at 5%) negative coefficient of the dummy variable indicating the presence of a household member who left the household in the model for one-vehicle households. The resulting change in household composition appears to prompt sooner vehicle replacement. Household income variables are included in both models and are significant. With their non-linear relationships, however, income effects are difficult to interpret.

Household Attributes: Households with Three or More Vehicles
Partly due to the small sample size, only a few variables have significant coefficients in the model for households with three or more vehicles. They indicate: households comprising non-couple, multiple adults tend to plan vehicle replacement sooner, while households with more drivers tend to plan vehicle replacement at later times.

7.6.6 Summary

The analyses of this section have focused on the intended -- or stated -- timing of the next vehicle transaction. The application of ordered-response logit models to stated transaction timing indicated that the average age of the vehicles in the household fleet, the age of the oldest vehicle, average holding duration, and total annual VMT are all associated with the intended timing of the next transaction. Associations have also been found between certain household attributes and the intended transaction timing. The results offer strong indications for the presence of bipolar tendency that those who have held their vehicles longer tend to intend to make the next transaction in the farther future, while those who have had their vehicles for shorter durations tend to plan a transaction in the near future. Clearly the intended timing of the next transaction is duration dependent.

7.7 Predictive Models

The discussions of this section so far have focused on the exploration of the characteristics of vehicle holding durations and intended transaction timing. Consequently the models developed in the analysis have many variables whose future values are difficult to forecast, and which therefore cannot be included in models for predicting transaction timing. In addition, the ordered-response logit models of intended transaction timing models presented above have dependent variables that describe transaction timing in terms of four timing categories. This does not offer adequate levels of temporal resolution desired for predictive purposes. Considering these, a new set of models are estimated using a parsimonious set of explanatory variables, and assuming Weibull distributions for the elapsed time between successive transactions. Results are presented in Tables 7.5 and 7.6.

Table 7.5: Weibull Model of Stated Transactions Timing for All Vehicle Ownership Levels

Variable		Coef.	t
Intercept		5.025	102.02
Number of vehicles		-0.149	-8.17
Age of the oldest vehicle in fleet		-0.003	-1.68
Age of the newest vehicle in fleet		-0.032	-11.94
Household Type	Single person	0.116	3.08
	Couple with child(1,2)	-0.053	-1.95
	Couple with child(3)	-0.183	-4.20
	Single with child(1,2,3)	-0.104	-2.24
Household Income	$46,000 to $60,000	-0.064	-2.18
	$61,000 to $75,000	-0.069	-1.88
	$76,000 to $125,000	-0.084	-2.42
	Over $125,000	-0.185	-3.43
Household Head's Employment Status		-0.167	-5.86
Household Wife's Employment Status		-0.060	-2.30
Household Head's Education L	Graduate work or degree	0.084	3.14
Number of members 16 yrs. old or older		-0.051	-2.74
Shape Parameter (1/g)		0.552	66.65
N		3459	
L(C)		-6629.2	
L(b)		-6226.8	
-2[L(C)-L(b)] (df)		804.9(15)	
Right censored observations		553	
Interval censored observations		2906	

Table 7.6: Weibull Model of Stated Transactions Timing by Vehicle Ownership Level

		1 VEHICLE		2 VEHICLES		3+VEHICLES	
Variable		Coef.	t	Coef.	t	Coef.	t
Intercept		4.754	103.46	4.658	105.03	4.675	32.78
Number of vehicles						-0.139	-3.49
Age of the oldest vehicle in fleet		-0.046	-13.41	-0.010	-3.91	0.007	2.35
Age of the newest vehicle in fleet				-0.023	-5.04	-0.031	-5.38
Household Type	Single person	0.231	5.20			-0.244	-1.69
	Couple with child(1,2)			-0.060	-1.78		
	Couple with child(2)					-0.123	-1.95
	Couple with child(2,3)	0.309	2.37				
	Couple with child(3)			-0.219	-4.04	-0.217	-2.69
	Single with child	0.324	3.84	-0.348	-3.91		
	Single with child(2)	0.372	2.71	-0.446	-3.07		
	Single with child(3)	0.613	3.33	-0.467	-2.66		
	Other			-0.174	-3.01		
Household Income	Less than $15,000	0.108	2.00			-0.664	-3.40
	Less than $30,000			0.078	1.78		
	$46,000 to $60,000					-0.157	-2.37
	Over $60,000			-0.062	-1.84		
	Over $125,000					-0.324	-3.65
Number of Children		-0.178	-4.13				
Household Head's Employment Status		-0.171	-3.86	-0.131	-3.14	-0.209	-2.93
Household Wife's Employment Status				-0.133	-4.12		
Household Head's Education L	Graduate work or degree			0.071	1.96	0.181	2.73
Shape Parameter (1/g)		0.515	37.34	0.533	46.94	0.605	29.13
N		1105		1697		657	
L(C)		-2168.2		-3100.4		-1137.0	
L(b)		-2038.3		-3022.9		-1100.6	
-2[L(C)-L(b)] (df)		259.9(9)		155.1(13)		72.6(11)	
Right censored observations		245		245		63	
Interval censored observations		860		1452		594	

7.8 Conclusions

Characteristics of vehicle holding duration and intended timing for the next vehicle transaction are examined in this report using the data obtained from the first-wave of the panel survey of the project. The analysis is limited in the sense that no measurements are available for past attributes of the household and its members except a few selected variables such as household size, and that a household's past history of vehicle transaction that can be reconstructed from the survey results may not be accurate due to reporting errors. Nonetheless the survey results offer a rich data base that can be explored to reveal characteristics of vehicle holding and transaction behavior in California. Salient results of this study are summarized in this section.

Five alternative statistical distributions are examined to describe the distribution of vehicle holding durations reported by the respondents. They include: negative exponential distribution, Weibull distribution, log-normal distribution, log-logistic distribution, and generalized gamma distribution. Of these, the gamma and Weibull distributions exhibit virtually the same goodness-of-fit to the data, and considering primarily computational convenience, the Weibull distribution is used in the subsequent analysis.

Estimates of the shape parameter of the Weibull duration models developed in this study offer consistent indications that vehicle holding is positively duration dependent. Namely, the probability of replacing or disposing of a vehicle increases with the holding duration. This positive duration dependence is weakest among households with three or more vehicles; vehicle holding behavior of households with large fleets appears to exhibit more random, "memoriless" patterns.

The Weibull vehicle holding duration models have offered the following findings. As one might anticipate, used vehicles and vehicles heavily utilized tend to be kept shorter. Estimation results also indicate that leased vehicles and company vehicles tend to be held shorter.

Holding durations significantly vary by body type. Mini, subcompact, luxury and sports cars tend to be held longer. Minivans, on the other hand, tend to have shorter holding durations, but this may be the consequence of their relative newness in the market.

The age of the primary user of a vehicle is strongly correlated with how long is it held; older users tend to keep their vehicles longer. Primary users with personal incomes exceeding $100,000, on the other hand, tend to keep their vehicles for shorter durations.

A household with a vehicle fleet which is large relative to the number of drivers tends to hold its vehicles longer. A household with a shorter duration of residence at the current location, on the other hand, tends to have shorter vehicle holding durations. Low income households, presumably due to their limited economical capabilities, tend to hold their vehicles longer.

Durations models estimated by vehicle ownership level indicate that households with three or more vehicles tend to have different vehicle holding characteristics. Used vehicles tend to be held shorter at all levels of ownership. On the other hand, vehicle utilization is found to be associated with holding durations only for households with three or more vehicles. Leased vehicles and company vehicles tend to be held shorter by households with one or two vehicles, but not by those with three or more vehicles.

The primary use's age is not significant for households with three or more vehicles. Female primary users tend to keep their vehicles shorter only among two-vehicle households, which probably comprise a man-woman couple without driving-age children.

The intended timing for the next vehicle transaction as reported by respondents is analyzed using ordered-response logit models. The results indicate that households with older vehicle fleets and those with greater vehicle-miles per vehicle, tend to intend the next transaction at later times. As the age of the oldest vehicle in the fleet increases, however, the intended timing tends to be the nearer future. There are indications that those who rent their homes, reside in mobile homes, or whose members have left the household, tend to intend to have the next transaction later. Single persons and those households in earlier stages of life-cycle stage tend to do so sooner.

Consistent with the results of the analysis of vehicle holding durations, intended timing for vehicle replacement is less dependent on household fleet attributes when the number of vehicles exceeds two.

The results obtained in this study consistently indicate that a bipolar tendency exists and splits households into two groups; one comprising those which hold their vehicles for long durations and the other comprising those which frequently replace their vehicles. For example, households which have held their vehicles longer tend to intend to make the next transaction in the farther future. Households which have had their vehicles for shorter durations, on the other hand, tend to plan a transaction in the near future. Differences in expected holding durations among households may not fully account for this tendency. The results so far suggest that the tendency coexists with, and partly cancels, the positive duration dependence identified by the Weibull duration models. Further investigation of the issue is believed to lead to more accurate predictive models of vehicle transaction behavior.

8. HOUSEHOLD TRANSACTIONS TYPE CHOICE

8.1 Background

8.1.1 Research Objectives

Manufacturers and government agencies interested in promoting alternative-fuel vehicles, and public utilities who must provide adequate refueling infrastructure, need to know how demand is affected by attributes that distinguish these vehicles from conventional gasoline or diesel vehicles. Such attributes include: range between refueling, overnight recharging requirements (electric and compressed natural gas), the potential availability of at-home refueling (compressed natural gas), the availability of refueling and opportunity recharging stations, vehicle performance levels, cargo carrying capacity, and capital and operating cost differences compared to conventional-fuel vehicles. It is also important to establish the extent to which consumers are attracted to vehicles that have reduced tailpipe emissions, as well as the effectiveness of various proposed incentives designed to promote sales and use of alternative-fuel vehicles. This is especially important in states like California, where stringent vehicle emission standards have been adopted or proposed. All new cars sold in California will be required to emit 80 percent less hydrocarbons by the year 2000, and 50 to 75 percent less carbon monoxide and nitrogen oxide. The California Air Resources Board (CARB) has also mandated the production and sale of zero-emission (electric) vehicles, beginning with 2 percent of annual sales in 1998 and increasing to 10 percent in 2003.

In this paper we describe a model that has been developed to provide the personal vehicle choice component of an integrated microsimulation forecasting system. A major goal is to improve the quality of forecasts by focusing on vehicle transactions rather than vehicle holdings.

8.1.2 Overview of the Forecasting Model System

The system is being designed to forecast demand for all types of vehicles subject to clean air mandates for each of 79 geographic areas, called districts, within the urbanized regions in California, excluding San Diego County. The districts are defined to be consistent with utility company service planning areas. The types of vehicles include all cars and light-duty trucks (pickup trucks, vans, and sport utility vehicles), as well as medium duty trucks up to 14,000 pounds gross vehicle weight. There will be two separate components for personal vehicles and commercial fleet vehicles, respectively, that will be linked through a third component that takes into account price effects associated with the used vehicle market.

The system will provide forecasts for aggregated "vehicle classes," and is being designed to support the definition of a variety of conventional-fuel and alternative-fuel classes that might become available during future forecast periods. Thus, vehicle classes for today's existing vehicles are formed by clustering together all makes and models with similar attributes (e.g., body type and size) into relatively homogeneous groups. We are currently using 14 classes (7 car classes and 7 light truck classes) for conventional-fuel personal vehicles, with each class further subdivided into 10 model-year vintage subclasses.

The system will also forecast fuel usage for each type of vehicle in each district. To determine the impact of electric vehicle recharging on the electricity transmission and distribution system, it will also forecast recharge demand for electric vehicles by time of day. Currently, peak electricity demand in California occurs during summer afternoons, and minimum demands occur between midnight and 6:00 A.M. Therefore, electric vehicle recharging will be much cheaper and less polluting if it takes place during late night hours when electricity is generated by relatively clean baseline plants.

The model system must be able to simulate the dynamics of the new vehicle adoption process. Thus, the system is designed to produce a separate forecast for each period, and the next period's forecast must depend on all the previous forecasts. For this reason, it is desirable to focus on vehicle transactions, and to calibrate dynamic behavioral models that use panel data.

The system uses microsimulation: it starts with a database of representative households and commercial fleets for the base year, and then simulates vehicle transactions at the level of the individual decision-making unit. Forecasts are reported for the current period by aggregating the results to the district level. However, dynamic effects are preserved by maintaining individual disaggregated histories as required by the behavioral models. This structure is similar to the system of Hensher, et al. (1992), where the population is represented by a relatively small number of "synthetic" households. We use a large sample of actual households and fleets obtained from our surveys. Such a microsimulation approach requires more computation, but it should be more accurate.

The inputs to our transaction models are the current characteristics of the household (or fleet) and the current vehicle inventory and utilization. Since vehicle type decisions are discrete, the models can only provide probabilities that a particular household or firm will purchase a particular type of vehicle. Forecasting a particular choice from these models requires simulating an actual choice, which introduces some random noise into the forecasting process. Fortunately, the effect of this randomness disappears when forecasts for individual households or fleets are aggregated to predict market demand. The predicted changes in vehicle holdings and utilization are then combined with initial holdings to forecast vehicle stocks for the next period.

The effects of estimation errors on the resulting forecasts will be measured by a "bootstrapping" process (Efron and Tibshirani, 1993). A number of different forecasts will be generated using different parameter values chosen to represent the parameter estimation uncertainty. The resulting spread of forecasts will generate confidence regions for our forecasts.

The model system will internally set used car prices so that the demands for used cars forecast by the submodels equals the predicted number of used cars sold by the submodels. This price equilibration will be performed separately for small groups of vehicle type-vintage classes. Therefore, one important feature of our model system is that it will provide estimates of used prices for alternative-fuel vehicles. Our approach requires that the used vehicle market in California is closed, or that used-vehicle price differences do not cause people to move vehicles in or out of the state. This assumption is reasonable given California's geography: the main urban areas are far away from urban areas in neighboring states.

Although our personal vehicle and fleet demand submodels exclude rental and state and federal fleets, these fleets are an important source of vehicles entering the used market. At this time, it appears that rental fleets will be excluded from all alternative-fuel vehicle mandates, so we will model their behavior as fixed throughout the forecast period. Specifically we will assume that rental fleets purchase and sell the same type and number of vehicles as they did in 1993-1994.

For political reasons, state and federal fleets will need to meet alternative-fuel vehicle mandates. We will therefore assume that they purchase enough vehicles to meet these mandates in the lowest-cost fashion. We will also assume that these fleets continue to follow the same vehicle sales and scrappage policies as in 1993-1994. Clearly our rental and government fleets "models" could be considerably improved. Unfortunately, the required data collection is beyond the scope of the current project.

The key inputs to our forecasting system are vehicle technology, fuel prices, fuel infrastructure, and incentives. Vehicle technology includes all attributes of vehicles which will become available in the future, including fuel type, refueling or recharging range, price, operating costs, vehicle tailpipe emissions, payload, and performance. Although it is relatively easy to forecast these attributes two to three years ahead, it is very difficult to predict the state of new technology ten or more years ahead. Forecasts from the model system crucially depend on future vehicle technology, and users of the model system will need to continually update this information as time progresses. Since the model produces forecasts for each year, it is also important to forecast when new technology vehicles will be introduced. Finally, the model system assumes that manufacturers are willing to provide as many vehicles as demanded at the forecast vehicle price.

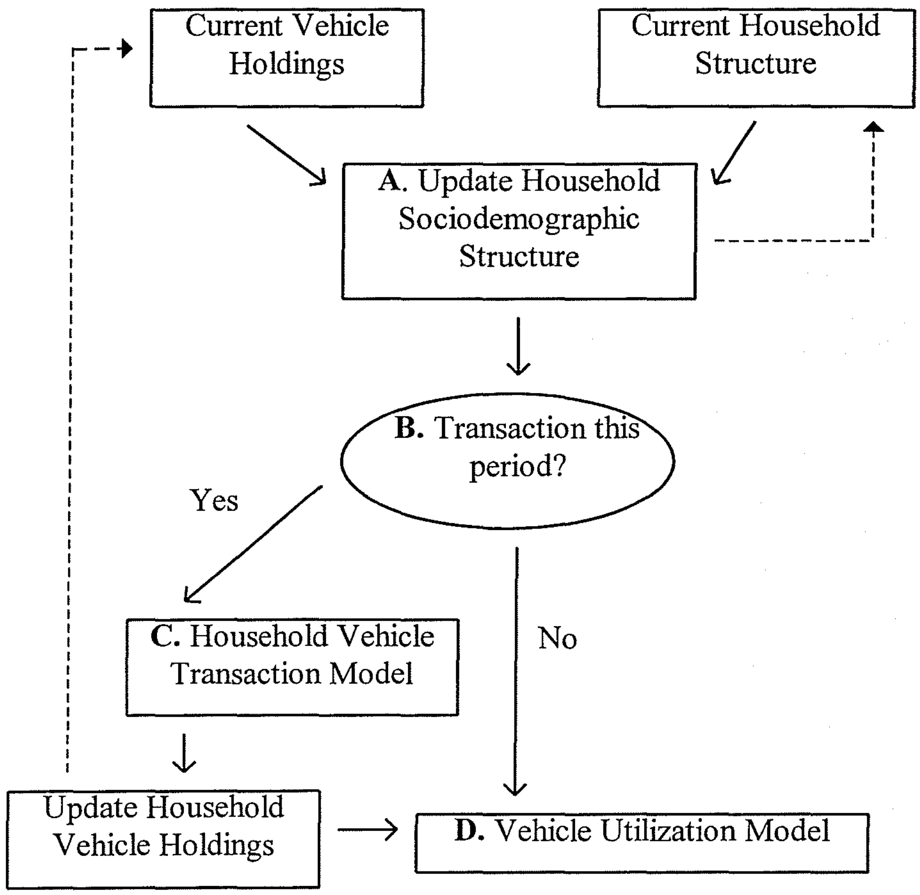
Fuel prices and availability are other exogenous inputs to the model system. Although prices are typically very difficult to forecast, we only need accurate forecasts of relative fuel prices. Fuel availability for compressed natural gas and methanol is described in comparison to the availability of gasoline (e.g., "one methanol station for every 10 gasoline stations"), while the electricity fuel infrastructure also includes the types of places (e.g. shopping centers, airports, etc.) where "opportunity charging" is available.

The final set of exogenous inputs are incentives for purchasing alternative-fuel vehicles. Many proposed incentives (such as, sales tax and vehicle registration fee subsidies) simply lower the capital and/or operating costs of these vehicles, so the effects of these incentives can be modeled by changing the appropriate cost variables in the vehicle technology section.

8.1.3 The Personal Vehicle Demand Model

Our framework for forecasting personal vehicle demand is summarized by the model in Figure 1, which consists of a number of linked sub-models. The initial current vehicle holdings and household structure are taken from the personal vehicle survey described below. Box A in Figure 1 represents a series of models which age each household, and simulate births, deaths, divorces, children leaving home, etc. Once the new household structure is determined, other models in Box A determine the household's income and employment status. The dotted line leaving Box A shows that this updated household is used as the starting point for aging the household in the next period. The models in Box A are mostly calibrated from the Panel Study of Income Dynamics (Hill, 1992).

Figure 1: Personal Vehicle Submodel



Ellipse B in Figure 1 takes the updated household and current (aged) vehicle holdings as inputs. It then decides whether or not a vehicle transaction takes place during this period. The period length is set at six months, in order to limit the number of transactions per period to one, but model system outputs are given annually. A vehicle transaction is defined to include: disposing of an existing vehicle, replacing an existing vehicle with another one, or adding a new vehicle to the household's fleet.

If the simulation from the transactions model in Ellipse B predicts that a vehicle transaction has taken place, the transaction type model in Box C determines exactly what type of transaction takes place. The household's vehicle holdings are updated accordingly, and these are used as inputs to the vehicle utilization model in Box D as well as starting values for the next period's forecast. The model outputs for each year accumulate the probabilities of all actions to the total numbers of vehicles owned or leased by type and vintage. For new vehicles, this represents market penetration. The focus in this paper is on the model represented by Box C in Figure 1.

Another important component is utilization (sub-model D.). It takes the updated vehicle holdings and household structure as inputs. It then predicts the annual vehicle miles traveled for each household vehicle. The usage forecasts are then converted to fuel demand by using average miles per gallon for liquid fuels and miles per equivalent gallons for non-liquid fuels. For electric vehicles, the utilization model also predicts the frequency of recharging at different times of day.

8.2 Literature Review

8.2.1 Alternative-fuel Vehicle Demand

Most of the earlier studies on alternative fuel vehicle demand focused on demand for electric vehicles (EV's). The SRI (1978) study used the model of Crow and Ratchford (1977) to forecast total sales of electric vehicles in the United States. Mathtech (Karfisi, Upton, and Agnew, 1978) forecasted electric vehicle demand by adapting a model in Wharton (1977). Beggs, Cardell and Hausman (1980) study the potential demand for EVs by applying an ordered logit model to stated preference data in which individuals provide rank orderings for hypothetical vehicle descriptions. Train (1980) uses a vehicle-type choice model (multinomial logit model), which was developed by Lave and Train (1978), to forecast the market share for several specific non-gasoline-powered automobiles. Hensher (1982) focuses on the demand elasticities for electric cars in Sydney, Australia. Calfee (1985) studies only the potential private demand for electric autos (i.e., no trucks or vans), using discrete-choice SP data and a fully disaggregated logit model. The work described here was preceded by a study described in Bunch et al. (1992). Bunch et al. (1992) employs nested multinomial logit models and multinomial probit models for vehicle choice and the binary logit model for fuel choice.

Probably the most comprehensive forecasting work performed to date is due to Train, which we describe here and in the next section. Train (1980) uses a vehicle-type choice model (multinomial logit model), which was developed by Lave and Train (1978), to forecast the market share for several specific non-gasoline-powered automobiles: three types of battery-powered vehicles (nickel-zinc, high-temperature #1, and high-temperature #2), a hybrid gas and battery vehicle, a hydrogen vehicle, and a vehicle run by the reaction of aluminum into energy and oxidation products. Train develops a "most likely case" scenario, and concludes that, for this scenario, 2.3% of passenger autos will be battery-powered by the year 2000. These results are similar to Dickson and Walton's (1977): they estimated that 3.4 million electric vehicles would be sold from 1990 to 2000, or about 2.4 percent of all vehicles sales during that period.

8.2.2 Vehicle Holdings and Transaction Models

There are many studies on vehicle holding and transactions: Farrell (1954), Janosi (1959), Kreini (1959), Huang (1964, 1966), Golob and Burns (1976), Johnson (1975, 1978), Lave and Train (1979), Lave and Bradley (1980), Train (1980a), Hocherman, Prashker, and Ben-Akiva (1982), Booz, Allen, and Hamilton, Inc. (1983), Hensher and Le Plastirer (1983), Mannering and Winston (1983), Winston and Mannering (1985), Berkovec and Rust (1985), Train (1986), Hensher, Barnard, Simith, and Milthorpe (1990), and Smith, Hensher, and Wrigley (1991). We summarize Train's model (1986) below.

Train (1986) develops a hierarchical structure to model auto ownership and use. This model has several submodels: vehicle quantity submodel, class/vintage submodel for one-vehicle households, class/vintage submodel for two-vehicle households, annual VMT submodel for one-vehicle households, annual VMT submodel for each vehicle for two-vehicle households, submodel for the proportion of VMT in each of two categories (work and shopping) for one-vehicle households, and a corresponding submodel for each vehicle in two-vehicle households.

Train's model has much in common with previous models: (1) it is a behavioral model that is estimated using choices from a household survey; (2) each household's choices depend on both the vehicle characteristics of each class/vintage (such as vehicle purchase price) and the household characteristics (such as household annual income); and (3) the model can be incorporated into a simulation framework to forecast the demand for and use of vehicles.

Compared to previous household vehicle demand models, Train's model has some advantages: (1) the model can forecast the number of vehicles owned and the annual VMT for each vehicle class/vintage; (2) it explicitly shows the interdependence of a household's choice of how many vehicles to own and of which vehicle class/vintage to own; (3) it explicitly indicates that a household's choice of how many and what vehicle(s) to own closely relates to how much the household drives, and vice versa; and

(4) it shows that each household chooses a particular make/model from within its chosen vehicle class without asking for a specification of the demand for each make/model.

Although there is a transaction dummy in Train's vehicle type submodel, the model only explains which class/vintage vehicle a household owns at some point in time, without considering the transaction(s) leading to this portfolio. Calibration of Train's model only requires information on households' vehicle holdings.

8.2.3 Combined Revealed Preference and Stated Preference

Several studies have been conducted on the issue of stated preference (SP) and revealed preference (RP): Kroes and Sheldon (1988), Fowkes and Wardman (1988), Hensher, Barnard, and Truong (1988), Wardmand (1988), Louviere (1988), Ben-Akiva and Morikawa (1990), Hensher (1992), Bradley and Daly (1993), and Morikawa (1993).

The major contribution of Morikawa (1993) is correcting the state dependence and correlation in the RP/SP by linking error components of separate RP and SP equations. Fully jointly estimated RP/SP models are not generally available. The present model represents an attempt to link SP and RP choice information within a single choice model structure.

Although we will use both RP and SP information, we will not estimate RP and SP choices jointly, but estimate SP vehicle choices conditioned on current RP holdings. Since the model we build will be used for one-step dynamic forecasting, using a conditional model is appropriate. Also forecasting SP vehicle choices by conditioning on RP vehicle holdings can capture some heterogeneity between households and therefore avoids some possible bias problems.

8.3 The Survey Data

The first wave of our personal vehicle panel survey was carried out in June and July, 1993. The sample was identified using pure random digit dialing and was geographically stratified into 79 areas covering most of the urbanized area of California. 7,387 households completed the initial computer-aided telephone interview (CATI). This initial CATI interview collected information on: household structure, vehicle inventory, housing characteristics, basic employment and commuting for all adults, and the next vehicle transaction.

The data from the initial CATI interview were used to produce a customized mail-out questionnaire for each sampled household. This questionnaire asked more detailed questions about each household member's commuting and vehicle usage, including information about sharing vehicles in multiple-vehicle and multiple-driver households. The mail-out questionnaire also contained two stated preference discrete-choice experiments for each household. Each of these experiments described three 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 be similar (but not identical) to the household's description of their next intended vehicle purchase.

After the households received the mail-out questionnaires, they were again contacted for a final CATI interview. This interview collected all the responses to the mail-out questions. Additional questions about the household's attitudes towards alternative-fuel vehicles were also included in this interview.

The 4747 households that successfully completed the mail-out portion of wave one of the personal vehicle survey in 1993 represent a 66% response rate among the households that completed the initial CATI survey. A comparison with Census data reveals that the sample is slightly biased toward home-owning larger households with higher incomes, and weights are being developed to balance the sample to the known population. Eighty percent of the households in the sample had exactly one driver per vehicle, proving that, in California, the number of drivers is the most important determinant of the vehicle ownership level. For two vehicle households, a little over one-third of the vehicles are driven 10,000 miles per year or less, a third are driven 10,000 - 15,000 miles per year, and almost a third are driven more than 15,000 miles per year.

An example SP task from the questionnaire is given in the Appendix. There are four fuel-types for vehicles: gasoline, compressed natural gas (CNG), methanol, and electric (EV). Three of the four fuel-types appear in each SP question. SP purchase prices and SP vehicle types are designed based on the household's intended spending and vehicle types. The framework of three vehicles per choice set allows estimating models which do not necessarily rely on the assumption of independence from irrelevant alternatives (IIA). Levels need to be chosen for six or seven attributes per vehicle per choice set, and four levels were used to cover the range of most attributes, allowing for estimation of nonlinear effects. The basic design producing the variation in attribute levels was an orthogonal main effects plan for a 4^{21} factorial in 64 runs (Golob et al., 1993).

8.4 Model Specification: SP Vehicle Choices Conditioned on RP Holdings

8.4.1 Dependent Variables

We are modeling the future demand for vehicles of four fuel-types: gasoline, EV, CNG, and methanol. Each household may have three actions: adding, replacing, or disposing. For adding or replacing, a household must decide which vehicle to add; for replacing or disposing, a household must decide which vehicle to dispose of. In our survey design each household faces six vehicle choices with different fuel types, vehicle types, vehicle sizes, and other attributes. A household could have 13, 20, or 27 alternatives depending on whether its current number of vehicles is 1, 2, or 3, respectively. For the present, zero-vehicle households are excluded, since there are only 53 households in the sample which own no vehicles.

Figures 2 and 3 depict the choice specification in the model. These tree structures suggest that the transactions should be modeled with a nested logit specification. Nested logit specifications were estimated for one-vehicle households and we found that the coefficient for each inclusive value is not significantly different from one. Thus, the conditional multinomial logit model (McFadden, 1974) was used for both one- and two-vehicle households. We applied a Hausman test to verify the IIA property for both one- and two-vehicle households.

Figure 2. One-Vehicle Household Transaction Tree

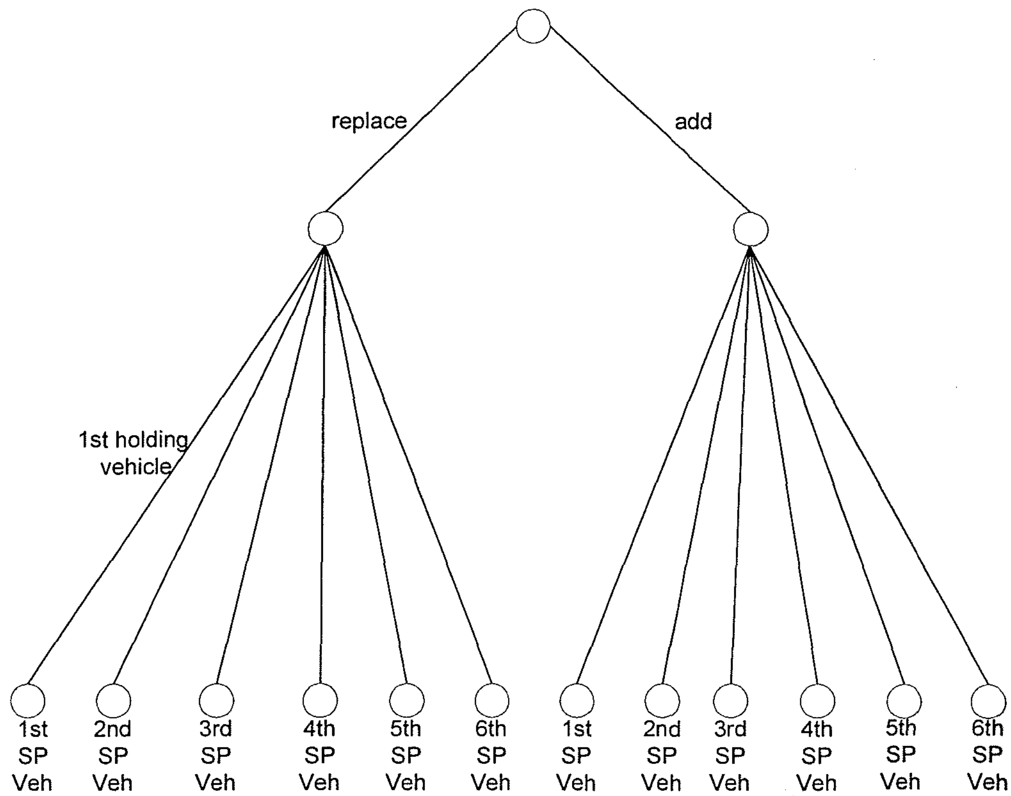
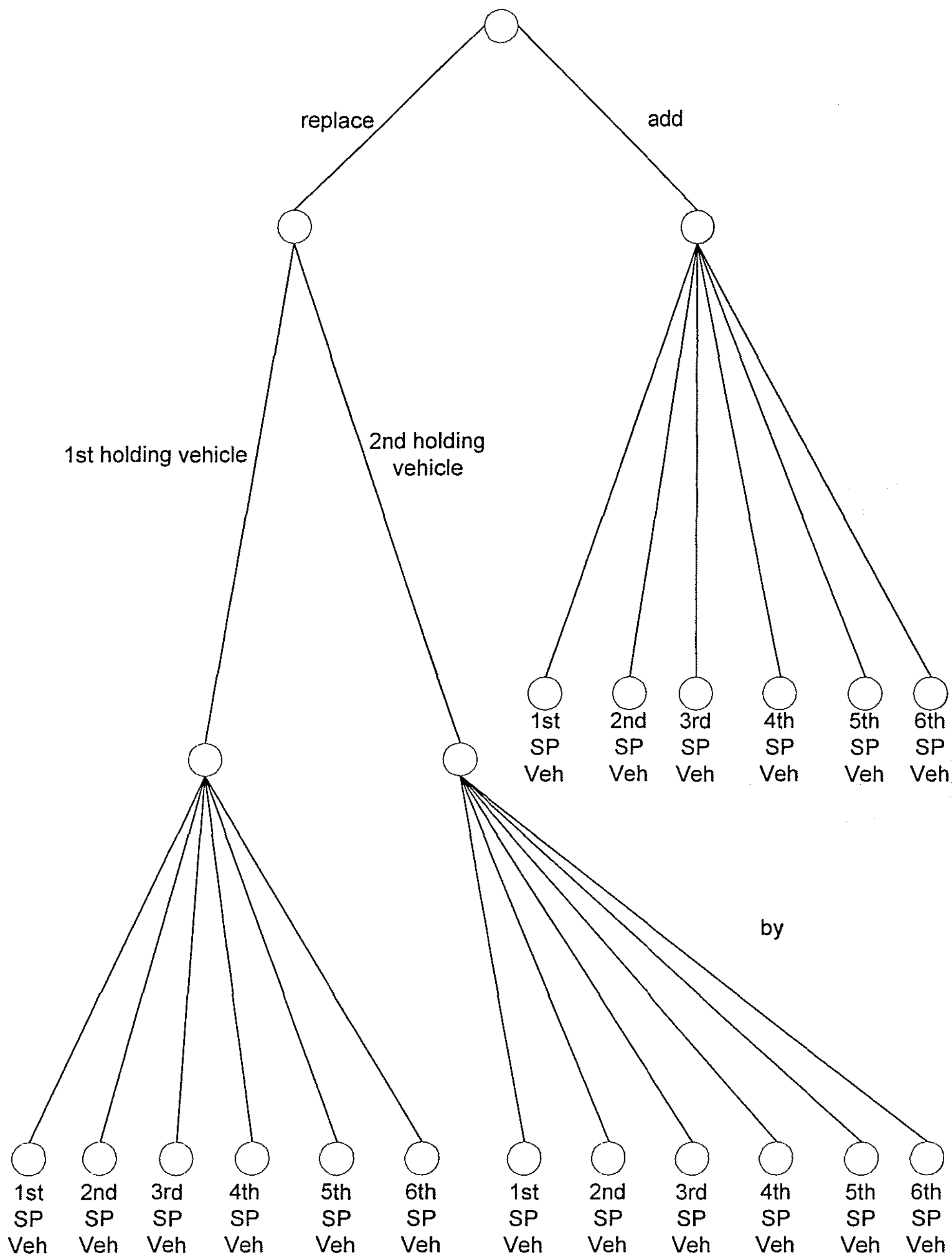


Figure 3. Two-Vehicle Household Transaction Tree



The specifications of the dependent variables for the one- and two-vehicle households are provided in Tables 1 and 2, respectively. The order of the 1st and 2nd vehicles corresponds to the order in which respondents entered their vehicles. The order of the 1st to 6th SP vehicles corresponds to the order on the survey form.

Table 8.1: The Dependent Variable for One-Vehicle Households

Value	Description
1	choose 1st SP vehicle to replace the holding vehicle
2	choose 2nd SP vehicle to replace the holding vehicle
3	choose 3rd SP vehicle to replace the holding vehicle
4	choose 4th SP vehicle to replace the holding vehicle
5	choose 5th SP vehicle to replace the holding vehicle
6	choose 6th SP vehicle to replace the holding vehicle
7	add 1st SP vehicle
8	add 2nd SP vehicle
9	add 3rd SP vehicle
10	add 4th SP vehicle
11	add 5th SP vehicle
12	add 6th SP vehicle

Table 8.2: Dependent Variable for Two-Vehicle Households

Value	Description
1	choose 1st SP vehicle to replace the 1st holding vehicle
2	choose 2nd SP vehicle to replace the 1st holding vehicle
3	choose 3rd SP vehicle to replace the 1st holding vehicle
4	choose 4th SP vehicle to replace the 1st holding vehicle
5	choose 5th SP vehicle to replace the 1st holding vehicle
6	choose 6th SP vehicle to replace the 1st holding vehicle
7	choose 1st SP vehicle to replace the 2nd holding vehicle
8	choose 2nd SP vehicle to replace the 2nd holding vehicle
9	choose 3rd SP vehicle to replace the 2nd holding vehicle
10	choose 4th SP vehicle to replace the 2nd holding vehicle
11	choose 5th SP vehicle to replace the 2nd holding vehicle
12	choose 6th SP vehicle to replace the 2nd holding vehicle
13	add 1st SP vehicle
14	add 2nd SP vehicle
15	add 3rd SP vehicle
16	add 4th SP vehicle
17	add 5th SP vehicle
18	add 6th SP vehicle

The estimates and forecasts described here do not distinguish between new and used SP vehicles. In the initial CATI interview we asked respondents whether they intended to purchase a new or used vehicle at their next transaction, and we also asked the price range for the vehicle purchased as part of the next transaction. Future work will use these data to model the choice of new/used vehicles as well as the vintage of the used vehicles. However, preliminary tests did not find any significant differences in preferences between new and used vehicle purchasers.

8.4.2 The Independent Variables

Since we are modeling the SP vehicle choices conditioned on current vehicle holdings, we decompose explanatory variables X_{ij} into five parts

$$X_{ij} = X^{SP\text{-}RP}_{ij} + X^{RRP}_{ij} + X^{SP}_{ij} + X^{SPRP}_{ij} + X^{OT}_{ij}, \qquad (1)$$

where $X^{SP\text{-}RP}_{ij}$ are the variables indicating the differences between the SP data and RP data; X^{RRP}_{ij} are the variables representing the attributes of the remaining holding vehicle; X^{SP}_{ij} are the attributes of SP vehicles; X^{SPRP}_{ij} are the variables interacting with SP data and RP data; and X^{OT}_{ij} are other attributes. The above specification shows the idea of estimating SP vehicle choices conditioned on vehicle holdings. The computation of the independent variables is demonstrated by example.

Example

a. Net capital cost ($X^{SP\text{-}RP}_{ij}$)

i. For replacing: alternatives 1 - 6;

= SP vehicle price - current market value of the holding vehicle

ii. For adding: alternatives 7 - 12;

= SP vehicle price

iii. For disposing: alternative 13;

= - current market value of the holding vehicle

b. Value of the remaining vehicle (X^{RRP}_{ij})

i. For replacing: alternatives 1 - 6

= 0

ii. For adding: alternatives 7 - 12

= current market value of the holding vehicle

iii. For disposing: alternative 13

= 0

The same procedure applies to the calculation of operating costs. The only difference, say for a one-vehicle household, is that for alternative 13 the difference is set to zero since after disposing of the holding vehicle, a household bears zero operating costs. The same procedure can also apply to top-speed and acceleration time.

The rationale for using these net benefit/cost variables is that a household not only compares the net gain or loss of a transaction, but also takes the benefit/cost left over from former holdings into account since this value does contribute to their utility. In other words, different remaining vehicles have different values to a household, so the utility function has to include this factor.

8.4.3 Testing the Independence of Irrelevant Alternatives

Using the MNL specification to model the transaction above, the basic assumption is that the disturbances are independent. To test the validity of the assumption versus the nested logit specification, Hausman and McFadden (1984) show that if a subset of the choices is irrelevant, eliminating it from the model will not systematically affect the estimates. However, excluding these choices will be inefficient. The irrelevancy is the basis for Hausman's specification test:

$$\chi^2 = (\hat{\beta}_r - \hat{\beta}_u)'[\hat{V}_r - \hat{V}_u]^{-1}(\hat{\beta}_r - \hat{\beta}_u), \qquad (2)$$

where r represents the estimators for the restricted subset, u represents the estimators for the full set of choices, β is the coefficient estimate, and V is the estimate of the asymptotic covariance matrix. The statistic is asymptotically distributed as chi-squared with K degrees of freedom, where K is the rank of the weight matrix. Note that after the choices are eliminated, some explanatory variables which are interacted with the eliminated alternatives are always zero. So, we can only use the remaining variables to do the test.

8.5 Estimation Results and Analysis

8.5.1 Estimation Results

Of 1607 one-vehicle households and 2220 two-vehicle households, 1153 and 1156 valid observations remained after excluding those with missing or incorrect data, primarily household income and vehicle year/make/model. Due to the small valid sample size for three-or more-vehicle households, in this paper we estimate and forecast for only one- and two-vehicle households. Due to lack of data on vehicle attributes, we excluded all vehicles with model years before 1979. Estimation results are obtained by using the first set of SP data.

For easy comparison, we will first list the results of one- and two-vehicle households, and then analyze and compare the results. Since this model is used for forecasting, more explanatory variables than usual are included.

In the tables below, HH stands for household; K stands for $1,000; # stands for number; and a dummy takes the value 1 when the condition is met, otherwise it is zero.

The estimation results for the sample of one-vehicle households are listed in Table 3. The Hausman test was conducted for one-vehicle households by excluding the replacement alternatives. At 95% significance level, we cannot reject the hypothesis that the MNL specification is correct.

The two-vehicle household estimation results are listed in Table 4. The Hausman test was also conducted for two-vehicle households by excluding the replacement alternatives. At 95% significance level, we cannot reject the hypothesis that the MNL specification is correct.

Table 8.3: Estimation Results for One-Vehicle Households

Explantory Variables	**Coefficient**	**t-stat**
Net capital cost (HH income ≤ 30K, HH has no child of age<21) *	-.0000809	-3.690
Value of the remaining vehicle (HH income ≤ 30K) *	.0000874	3.222
Net capital cost (31K ≤ HH income ≤ 75K, HH has no children<21) *	-.0000761	-3.848
Value of the remaining vehicle (31K ≤ HH income ≤ 75K) *	.0000605	2.613
Net capital cost (HH income > 75K,HH has a child of age<21) *	-.0000418	-0.919
Net capital cost (HH income > 75K, HH has no child of age<21) *	.000012	0.570
Net operating cost (HH income<=30K, HH has a child of age<21) **	-.0301917	-1.116
Net operating cost (HH income ≤ 30K, HH has no child of age<21) **	-.0750135	-4.021
Net operating cost(31K≤HH income≤75K, HH has a child of age<21) **	-.0938415	-2.916
Net operating cost(31K≤HH income≤75K,HH has no child of age<21) **	-.0413121	-2.148
Net operating cost (HH income > 75K, HH has no child of age<21) **	-.0825472	-1.879
Top-speed difference between the choice-set vehicle and the holding vehicle	.0013544	0.982
Acceleration time difference between the choice-set vehicle & the holding vehicle ***	-.0430502	-2.103
Refueling time	-.0007526	-1.287
Range of choice-set vehicle	.0045544	2.145
Range^2 of choice-set vehicle	-2.34e-06	-0.479
Service station availability for EV †	.1883788	0.440
Service station availability for dedicated CNG vehicle †	.6931062	1.743
Service station availability for methanol veh. and dual fuel CNG veh.	.4061086	1.895
Dual fuel (dummy)	.2558952	1.291
Pollution level of choice-set vehicle, for HH *with* child of age<21 †††	-.7822153	-2.716
Pollution level of choice-set vehicle, for HH *without* child of age<21 †††	-.7112071	-3.527
Luggage space of choice-set vehicle ††	.3559792	0.949
Van (HH size ≤ 3) (dummy)	-.5577494	-2.508
Van (HH size ≥ 4) (dummy)	.8886816	2.720
EV (Northern Calif. w/o SF, Oakland, San Jose) (dummy)	-.1458946	-0.595
EV*Subcompact (dummy)	.2597384	0.969
EV*Compact car (dummy)	.2044488	0.971
EV*Large (dummy)	.4978991	2.238
EV*Van (dummy)	-.1748806	-0.555
EV*Truck (dummy)	-.9020771	-3.080
EV*Utility vehicle (dummy)	-.4165532	-0.942
EV*Sport car (dummy)	-.4261871	-1.005
CNG*Mid-size car (dummy)	-.0670159	-0.403
CNG*Large car (dummy)	-.0864181	-0.448

CNG*Van (dummy)	.5585584	2.070
CNG*Utility (dummy)	.7891047	2.238
CNG*Sport car (dummy)	.0668197	0.191
Methanol*Mid-size car (dummy)	.1536637	0.932
Gasoline (dummy)	.6633813	2.403
Gasoline*Subcompact (dummy)	-.3741808	-1.407
Gasoline*Mini (dummy)	-1.428037	-2.399
.Gasoline*Compact (dummy)	-.5038076	-2.116
Gasoline*Mid-size car (dummy)	-.2806644	-1.221
Gasoline*Van (dummy)	-.0296736	-0.105
Gasoline*Truck (dummy)	-.6699467	-3.229
Gasoline*Utility (dummy)	-.3283719	-0.944
Gasoline*Sport (dummy)	.6058397	1.744
New holding--two vans (dummy)	-.2500269	-0.426
New holding--two utility vehicles (dummy)	-.7976523	-0.716
New holding--two cars (dummy)	.5810163	2.561
Alternative-add constant for HH, with children 15 or 16 years old (dummy)	.6565692	1.729
Alternative-add constant for HH with holding vehicle's type different from the choice vehicles'	.412067	1.935
Alternative-replace constant for HHs with # cars ≥ # adults (dummy)	1.021258	7.502
Alternative-replace constant for HHs with holding vehicle's type the same as choice set vehicle's (dummy)	1.201564	11.948
Choice-set-vehicles'-type=next-intended-vehicle-type ‡(dummy)	1.548205	14.905
Number of observations Initial Likelihood Final Likelihood "Rho-Squared" w.r.t. Zero	1334 -3314.8655 -2556.2844 0.2288	

* 1993 U.S. dollar.
** For EV, using home-refueling cost and home-refueling time. The unit for cost is cent/mile and the unit for refueling time is minute. The gasoline price is assumed 120 cents/gallon.
*** The time from 0 to 30 mph.
† It is the proportion of service stations which carry the fuel.
†† It takes the value of 1 (same size as RP vehicle) or .7 (30% smaller than RP vehicle).
††† It takes the value of 1 (1993 gasoline vehicle), or 0.4, 0.25, or 0 (for other alternative-fuel vehicles).
‡ Vehicle types are car, sport utility, sport car, truck, and van.

Table 8.4: Estimation Results for Two-Vehicle Households

Explanatory Variables	Coefficient	t-stat
Net capital cost (HH income ≤ 30K, HH has a child of age<21) *	-.0000831	-1.751
Net capital cost*(HH income ≤ 30K, HH has no child of age<21) *	-.0000269	-0.609
Value of the remaining vehicle (HH income ≤ 30K) *	.000055	1.095
Net capital cost (HH income > 30K,HH has a luxury vehicle and a child of age<21) *	.0000131	0.839
Net capital cost (HH income > 30K, HH has a luxury vehicle and no child of age<21) *	2.12e-06	0.214
Net capital cost (HH income > 30K, HH has no luxury vehicle, but a child of age<21) *	-.0000527	-2.753
Net capital cost (HH income > 30K, HH has no luxury veh. & no child of age<21) *	-.000073	-4.121
Value of the remaining vehicle (HH income > 30K, HH has no luxury vehicle) *	.0001469	6.230
Net operating cost (HH income ≤ 30K, HH has no child of age<21) **	-.0397447	-1.297
Net operating cost (HH income > 30K, HH has a luxury vehicle and a child of age<21) **	-.0514634	-1.108
Net operating cost (HH income > 30K, HH has a luxury vehicle and no child of age<21) **	-.0710147	-1.809
Operating cost of the remaining vehicle (HH income > 30K, HH has a luxury vehicle) **	.1004534	2.109
Net operating cost (HH income > 30K, HH has no luxury vehicle, but a child of age<21) **	-.0607736	-3.247
Net operating cost (HH income > 30K, HH has no luxury vehicle and no child of age<21) **	-.0757479	-4.359
Top-speed difference between the choice-set vehicle and the holding vehicle	.0032654	2.304
Acceleration time difference between the choice-set vehicle and the holding vehicle (HH income ≤ 30K) ***	.0426008	1.006
Acceleration time of the remaining vehicle (HH income ≤ 30K) ***	-.2176092	-0.735
Acceleration time difference between the choice-set vehicle and the holding vehicle (HH income > 30K) ***	-.0690339	-3.249
Refueling time of the choice-set vehicle	-.0007574	-1.361
Range of the choice-set vehicle	.0062765	3.088
Range^2 of the choice-set vehicle	-5.98e-06	-1.304
Service station availability for EV †	.2511702	0.597
Service station availability for edicated CNG vehicle w/o home-refueling †	.7791451	1.812
Service station availablity for edicated CNG vehicle w/ home-refueling†	.5187576	1.136
Dual fuel (dummy)	.2300668	1.889
Pollution level of choice-set vehicle for HH *with* child of age<21 †††	-.372541	-1.774
Pollution level of choice-set vehicle for HH *without* child of age<21 †††	-.0892202	-0.463
Luggage space of choice-set vehicle ††	.6608735	1.964
Van (HH size ≤ 3) (dummy)	.1859008	1.058
Van (HH size ≥ 4) (dummy)	1.123961	5.882
EV*(LA & Orange Counties) (dummy)	-.2281306	-1.100
EV*(S.F., Oakland, San Jose) (dummy)	-.099904	-0.491

EV*(Northern Calif. w/o SF, Oakland, and San Jose (dummy)	-.0253294	-0.103
EV*(Subcompact, Mini, Compact Cars) (dummy)	.3790473	1.808
EV*Mid-size car (dummy)	.6872253	3.123
EV*Van (dummy)	-.2743813	-1.200
EV*Truck (dummy)	-.6247521	-2.767
EV*Utility vehicle (dummy)	-.1675988	-0.466
EV*Sport car (dummy)	.0749827	0.187
CNG*Van (dummy)	-.1950898	-0.933
CNG*Utility (dummy)	.1307011	0.457
CNG*Truck (dummy)	-.508754	-2.413
CNG*Sport car (dummy)	-.0419138	-0.111
Methenol*Subcompact car (dummy)	-.3113108	-1.456
Gasoline*Subcompact (dummy)	.026683	0.115
Gasoline*Mini (dummy)	1.288469	2.473
Gasoline*Compact (dummy)	-.0253625	-0.158
Gasoline*Large car (dummy)	.0316858	0.175
Gasoline*Van (dummy)	-.3982977	-1.902
Gasoline*Truck (dummy)	-.0742639	-0.458
Gasoline*Utility (dummy)	.1367836	0.531
Gasoline*Sport (dummy)	-.3164726	-0.880
New holding--two or more vans (dummy)	-.8378513	-2.892
New holding--two or more trucks (dummy)	-.1873752	-0.817
New holding--two or more utility vehicles (dummy)	-1.129437	-2.153
New holding--two or more cars (dummy)	.3235226	3.458
Alternative-add constant for HHs with # cars < # adults (dummy)	.6051908	1.962
Alternative-add constant for HHs with a child 15 or 16 years old (dummy)	.229257	0.743
Alternative-add constant for HHs with holding vehicle's type different from the choice-set vehicle's type	-.2384718	-1.559
Alternative-replace constant for HHs with # cars ≥ # adults (dummy)	.5701477	2.118
Alternative-replace constant *(Lower value vehicle) (dummy)	.4718061	5.298
Alternative-replace constant for HHs with holding vehicle's type the same as choice-set vehicle's (dummy)	1.072339	14.230
Choice-set-vehicles'-type=next-intended-vehicle-type ‡ (dummy)	1.174223	14.501
Number of observations	1423	
Initial Likelihood	-4112.999	
Final Likelihood	-3400.8	
"Rho-Squared" w.r.t. Zero	0.1732	

* 1993 U.S. dollar.
** For EV, using home-refueling cost and home-refueling time. The unit for cost is cent/mile and the unit for refueling time is minute. The gasoline price is assumed 120 cents/gallon.
*** The time from 0 to 30 mph.
† It is the proportion of service stations which carry the fuel.
†† It takes the value of 1 (same size as RP vehicle) or .7 (30% smaller than RP vehicle).
††† It takes the value of 1 (1993 gasoline vehicle), or 0.4, 0.25, or 0 (for other alternative-fuel vehicles).
‡ Vehicle types are car, sport utility, sport car, truck, and van.

8.5.2 Interpretation of Results

8.5.2.1 Net Capital Cost

The net capital cost is the difference between the price of the SP vehicle and the current market value of the holding vehicle. Table 5 shows that the net capital cost for one-vehicle households with annual income less than $75,000 has, as we expect, a negative sign. For households with annual income greater than $76,000 the coefficient for net capital cost is insignificant.

For the two-vehicle households with annual income less than $30,000, the results are very similar to the one-vehicle results in that both have a negative sign. However, for the two-vehicle households with income greater than $31,000, the result varies significantly between households with and without luxury cars. The households without luxury cars behave more like "rational" people in that they demand less when the price is high. The households with luxury cars, however, prefer high-priced vehicles as reflected in the positive and significant coefficient. This result implies that there is a "name-plate" effect; that is, some people not only buy a vehicle but also buy status. This specification--with and/or without luxury vehicles--does capture some unobservable characteristics existing in the households.

Both results also show the big variation in coefficients for households with and without children under 21. This variation captures the difference for households with and without children under 21, although it is not clear in which direction the coefficient should vary.

8.5.3 Net Operating Cost

The net operating cost is the difference between the operating cost of the SP vehicle and the operating cost of the holding vehicles. As indicated by the formulae, which were discussed in the model specification, the net operating cost shows the net amount of money that people have to spend when they use the chosen vehicle.

Except the two-vehicle households with luxury cars, the coefficients of net operating costs for both one- and two-vehicle households have the expected negative sign. For two-vehicle households with luxury cars and with income greater than $31,000, the coefficient for net operating cost is positive and significant, as it was for net capital cost. Coefficients vary according to household income and with/without children under 21.

8.5.4 Value and operating cost of the vehicles in the resulting household fleet

The remaining vehicles are the remaining holding vehicles after a household's transaction. Since the value of the remaining vehicles is an asset to a household, the coefficient should have a positive sign. The estimation does support this expectation.

However, the operating cost of the remaining vehicle is still a cost or negative value to a household, so the sign of the coefficient should be negative. The estimation also supports this expectation. The coefficient of the value and the coefficient of operating cost of the remaining vehicle each varies with households' income and with/without children under 21. However, we can not foretell in which direction the coefficient should vary.

8.5.5 Top speed and acceleration time

The coefficients of the difference in top-speed have expected positive signs for both one- and two-vehicle households. However, the coefficient does not show significance for the one-vehicle household, and has a t-statistic of 1.6 for the two-vehicle household.

For the one-vehicle households, the coefficient of the difference in acceleration has a t-statistic of -1.6 and an expected sign, negative.

For the two-vehicle households, the coefficient for a household with income of $30,000 or less has a positive sign, and the coefficient for income of $31,000 or higher has an expected negative sign and is significant. Although it is not clear why the coefficient for a low-income household is positive, this does show that a low-income household, in contrast to a high-income household, does not care too much about acceleration time.

For a two-vehicle household, acceleration time of the remaining vehicle for low- and high-income households is specified. Acceleration time and operating cost of the remaining vehicle are similar in that they both have a negative value to a household. So, the negative and significant coefficients are expected.

8.5.6 Refueling time

Refueling time is service station refueling time for a non-EV and home-refueling time for an EV. For both one- and two-vehicle households the coefficients of refueling time have the expected signs, but are not significant. The reason is that people can recharge an EV at home, so the refueling time does not matter too much

8.5.7 Vehicle range

As expected, the coefficient of range for both one- and two-vehicle households has a positive sign and is significant. This implies that the range is a very important factor when households buy an alternative-fuel vehicle. The coefficient for $(\text{range})^2$ has a negative sign and is not significant. Although the coefficients of $(\text{range})^2$ are not significant for both one- and two-vehicle households, the implication is important: the increase in value from increasing vehicle range declines.

8.5.8 Service station availability

For both one-vehicle and two-vehicle households, the service station availability coefficients have the expected positive signs and their *t*-statistics range from 1.2 to 2.3. For two-vehicle households the coefficient for dedicated CNG vehicles without home-refueling is, as expected, the largest. Service station availability for dedicated CNG vehicles with and without home-refueling have the same value for one-vehicle households, so they are combined. For two-vehicle households, this coefficient is significant and relatively large in magnitude.

8.5.9 Emission level

For both one- and two-vehicle households, these two coefficients have expected negative signs and are significant. Also, as expected, the coefficient for households with children has a larger negative value than that for households without children. This is especially so for two-vehicle households, where the coefficient for households with children under 21 years of age is almost 10 times greater than that of households without children.

8.5.10 Vehicle and fuel-type interactions

There are many significant interactions between vehicle type and fuel type in both one- and two-vehicle models. To summarize, the results show that people are more likely to buy electric cars, as opposed to electric light-duty trucks and vans, and they are more likely to buy CNG utility and sport utility vehicles.

One-vehicle households generally prefer a gasoline vehicle to other alternative-fuel vehicles. For two-vehicle households this coefficient is zero; that is, for two-vehicle households a gasoline vehicle has no special advantage over alternative-fuel vehicles.

8.5.11 Vehicle type = vans

For both one- and two-vehicle households, the coefficients of van dummies for household size greater than 3 are significant and have expected positive signs. This result implies that households with 4 or more people will be more likely to buy a van.

For one-vehicle households with size less than 4, the coefficient has an expected negative sign and is significant. For two-vehicle households the coefficient has an expected negative sign, but is not significant. This difference between one- and two-vehicle households implies that for households with 3 or fewer people the value of a van is much less for a one-vehicle household than for a two-vehicle household.

8.5.12 Holdings of two or more vehicles of the same type

When a household decides to add a vehicle, a one-vehicle household will become a two-vehicle household and a two-vehicle household will become a three-vehicle household. We generally expect a household to have two or more cars, but not two or more special vehicles, such as two vans. For one-vehicle households these coefficients are not significant, but it is not immediately obvious to us why the coefficient for new-holding-two-trucks has a positive sign. For two-vehicle households, all the signs of the coefficients are as expected. The coefficients for new-holding-two-or-more-vans and for new-holding-two-or-more-cars are negative and significant.

8.5.13 Household adding vehicles

For both one- and two-vehicle households, coefficients associated with adding vehicles in households with fewer vehicles than drivers, and in households with children 15 or 16 years old, have the expected positive signs and have t-statistics ranging from 1.1 to 3.1. Obviously, when a household has more drivers than cars, or has a child 15 or 16 years old (close to or at legal driving age), the household will be more likely to add a car.

The coefficient associated with households in which the holding vehicle's type is different from the SP vehicle's type variable is designed to determine if a household would like to add a vehicle which is different in type from the holding vehicle. For one-vehicle households the coefficient is negative and not significant, which implies that one-vehicle households may or may not add a new vehicle that is different in type from the holding vehicle; that is, any combination of two types of vehicle is possible.

For two-vehicle households the coefficient is negative and significant, which implies that it is unlikely for a two-vehicle household to add a new vehicle that is different in type from both holding vehicles; that is, a three-vehicle household is unlikely to have, for example, a car, a truck, and a van.

8.5.13.1 Households replacing or disposing of vehicles

For both one- and two-vehicle households, the alternative-replace constant for the variable defining households with more vehicles than drivers has the expected positive sign. That is, if a household has more vehicles than drivers, it is more likely to replace than it is to add. This coefficient is significant for two-vehicle households.

For both one- and two-vehicle households, the alternative-dispose constant for households with a member over 60 years old is, as expected, positive and significant. This obvious result shows that older people are more likely to get rid of their vehicles.

8.5.13.2 Other vehicle type effects

The coefficient associated with replacing a station wagon by a van has an expected positive sign for both one- and two-vehicle households; that is, people are more likely to replace a holding station wagon by a van.

Also, for both one- and two-vehicle households, the alternative-replace constant for households in which the holding vehicle's type is the same as the SP vehicle's type, is positive and significant. This implies that most households just replace their old vehicle by a new vehicle with the same type.

8.5.13.3 Alternative-replace constant of replacing a cheaper vehicle

This variable is designed only for two-vehicle households. When a household decides to replace one of their holding vehicles, the one that is more likely to be replaced is not the older one but the one which has lower market value. The estimation supports this idea with a positive and significant coefficient.

8.5.13.4 Electric vehicle interactions with geographic variables

For two-vehicle households, the fuel type electric (EV) interacts with three geographic dummies: Los Angeles metropolitan area; San Francisco, Oakland, and San Jose; and Northern California excluding San Francisco, Oakland, and San Jose. All three coefficients are negative. The coefficient of EV fuel-type interacting with Los Angeles has the largest negative value, and is the only significant one. This implies that households in the Los Angeles Metropolitan Area are less inclined to purchase EV's than households in other urban areas in California, *ceteris paribus*. This is consistent with the hypothesis that those choosing to live in the Los Angeles area have demonstrated a higher tolerance for air pollution.

8.6 Used and New Vehicle Model

8.6.1 The Used and New Vehicle Issue

In the SP questions, we do not constrain the SP vehicles to new vehicles: these SP vehicles could be used vehicles. In order to model the used vehicles, in addition to the vehicle class we have to model the vehicle vintage. However, in the questionnaire we only ask what type of vehicle a household will buy and if the vehicle is new or used. If a household states that it wants to buy a used vehicle, we do not know the intended model year. Since we ask how much a household expects to spend on the next vehicle and we also have the data to show the market prices for vehicles of all fourteen classes from 1973 to 1993, I can "estimate" the intended model year of the next purchase by choosing the minimum difference between the market price and the expected spending. If a household states that it will buy a new vehicle, but the money the household expects to spend is not enough for the new vehicle, I will treat the household as a used-vehicle buyer. Similarly, if a household states that it will buy a used vehicle, but the money the household expects to spend is enough for the new vehicle, I will treat the household as a new-vehicle buyer.

8.6.2 The Grouping of the Vintages

There are fourteen vehicle classes. In this study, the model year goes from 1993 back to 1973. The total possible vehicle class-vintage combination is 14 (class)* 21 (year) = 294 (class-vintage). This fine classification is not necessary and also not easy to model in most econometric packages. In order to reduce the number of vehicle class-vintage combinations, I group the model years as shown in Table 5.

Table 8.5: The Grouped Model Year

Model Year	Mode Year Group
1993	1
1992	2
1991	3
1990	4
1989	5
1988	6
1987	7
1986	7
1985	8
1984	8
1983	8
1982	9
1981	9
1980	9
1979	9
1978	10
1977	10
1976	10
1975	10
1974	10
1973	10

Vehicles with different class, vintage, or fuel-type have different attributes, such as purchase prices, operating cost, acceleration time, and the number of makes. Using the number of makes as weight, the grouped vehicle attributes are the weighted average of those within the model year group. For example, the market price for vehicle class-1 and model year group 7 is calculated as follows:

$$\text{Price}_{1-7} = \frac{\sum_{i=1}^{4}(\text{Price}^{i}_{1-1987} * \#\text{Makes}^{i}_{1-1987} + \text{Price}^{i}_{1-1986} * \#\text{Makes}^{i}_{1-1986})}{\sum_{i=1}^{4}(\#\text{Makes}^{i}_{1-1987} + \#\text{Makes}^{i}_{1-1986})}, (3.8)$$

where, Price $_{1\text{-}7}$ is the grouped market price for vehicle-class 1 in model year group 7; Price $_{1\text{-}1987}$ is the market price for vehicle-class 1 in 1987; Price$_{1\text{-}1986}$ is the market price for vehicle-class 1 in 1986; #Makes$_{1\text{-}1987}$ is the number of makes for vehicle-class 1 in 1987; #Makes$_{1\text{-}1986}$ is the number of makes for vehicle-class 1 in 1986; and i stands for fuel type: gasoline, compressed natural gas, methanol, and electric.

The formula above implies that we average the attributes across both the model year and fuel type. The basic assumption here is that consumers, based on their budget, will first choose the vehicle type and model year without considering the fuel type.

8.6.3 Defining Dependent Variables

The dependent variable for the new/used vehicle model is defined in Table 6. Note that the vehicles before 1979 have no class 14.

Table 8.6: Dependent Variables for Used/New Vehicle Model for One-Vehicle Households

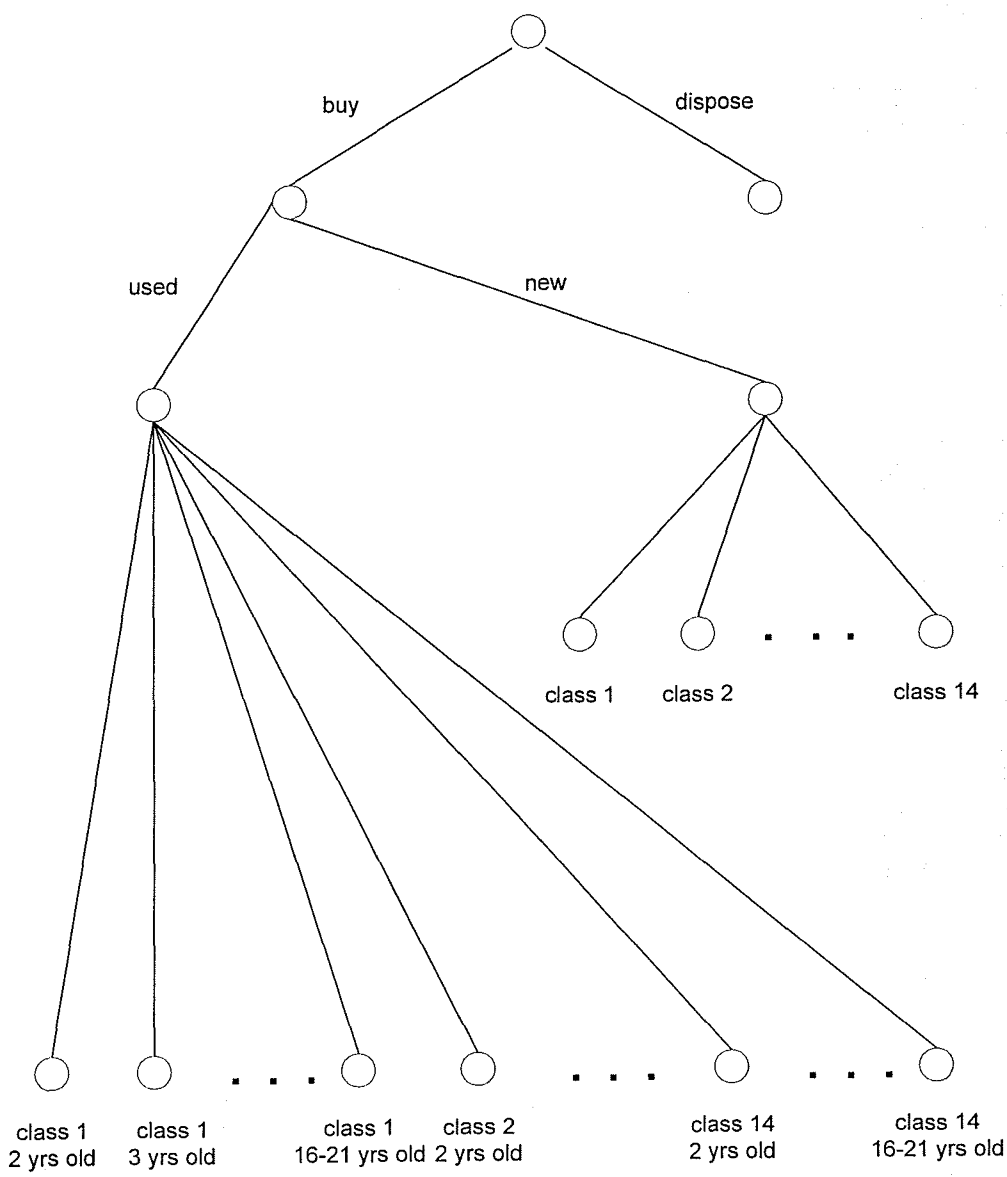
Value	Vehicle Class of Next Purchase	Grouped Model Year	Value	Vehicle Class of Next Purchase	Grouped Model Year
1	1	2	36	4	10
2	1	3	37	5	2
3	1	4	38	5	3
4	1	5	39	5	4
5	1	6	40	5	5
6	1	7	41	5	6
7	1	8	42	5	7
8	1	9	43	5	8
9	1	10	44	5	9
10	2	2	45	5	10
11	2	3	46	6	2
12	2	4	47	6	3
13	2	5	48	6	4
14	2	6	49	6	5
15	2	7	50	6	6
16	2	8	51	6	7
17	2	9	52	6	8
18	2	10	53	6	9
19	3	2	54	6	10
20	3	3	55	7	2
21	3	4	56	7	3
22	3	5	57	7	4
23	3	6	58	7	5
24	3	7	59	7	6
25	3	8	60	7	7
26	3	9	61	7	8
27	3	10	62	7	9
28	4	2	63	7	10
29	4	3	64	8	2
30	4	4	65	8	3
31	4	5	66	8	4
32	4	6	67	8	5
33	4	7	68	8	6
34	4	8	69	8	7
35	4	9	70	8	8

Table 8.6 continued

Value	Vehicle Class of Next Purchase	Grouped Model Year	Value	Vehicle Class of Next Purchase	Grouped Model Year
71	8	9	106	12	8
72	8	10	107	12	9
73	9	2	108	12	10
74	9	3	109	13	2
75	9	4	110	13	3
76	9	5	111	13	4
77	9	6	112	13	5
78	9	7	113	13	6
79	9	8	114	13	7
80	9	9	115	13	8
81	9	10	116	13	9
82	10	2	117	13	10
83	10	3	118	14	2
84	10	4	119	14	3
85	10	5	120	14	4
86	10	6	121	14	5
87	10	7	122	14	6
88	10	8	123	14	7
89	10	9	124	14	8
90	10	10	125	14	9
91	11	2	126	1	1
92	11	3	127	2	1
93	11	4	128	3	1
94	11	5	129	4	1
95	11	6	130	5	1
96	11	7	131	6	1
97	11	8	132	7	1
98	11	9	133	8	1
99	11	10	134	9	1
100	12	2	135	10	1
101	12	3	136	11	1
102	12	4	137	12	1
103	12	5	138	13	1
104	12	6	139	14	1
105	12	7	140	dispose	dispose

We can draw a tree to show the relation in Table 1. See Figure 1.

Figure 4. New/Used Vehicle Model



The dependent variable for two-vehicle households is very similar to that for one-vehicle households. The first 139 values of the dependent variable are exactly the same as for one-vehicle households. The values 140 and 141 are defined as the disposal of the first and second household holding vehicles, respectively. Therefore, compared to the tree for one vehicle households, there will be only one more node under the dispose branch.

8.6.4 Estimation Results

Of 1676 one-vehicle households and 2170 two-vehicle households, 1098 and 1230 valid observations remain. The reduction in size of the sample is due to missing or incorrect data, primarily household income, vehicle year/make/model, and the statement of next purchase.

After grouping the vintages, we still have 14 (class) * 10 (vintage) = 140 (class-vintage) combinations. Then, the data matrix for one-vehicle households has 1098 (households) * 140 (class-vintage) = 153,720 rows and 22 (number of variables) columns for one-vehicle household; and the data matrix for two-vehicle households has 1230 (household) * 141 (class-vintage) = 173,430 rows and 22 (number of variables) columns. Due to the limitation of the 32 megabyte memory of my computer, estimation of this model on the full set of alternatives is not feasible. As explained by Train (p. 47, 1986), estimation on a subset of alternatives will not influence the consistency. Furthermore, Train shows that beyond a minimal number of alternatives, the estimates are not sensitive to the number of alternatives included in the estimation. In his example, he uses 10 percent of the total alternatives.

I use 21 alternatives, about 15 percent of the total alternatives, as the subset of alternatives. These 21 alternatives are the household's chosen alternative plus 20 alternatives randomly selected from the remaining 139 alternatives for one-vehicle households and 140 alternatives for two-vehicle households.

For easy comparison, I will first list the results for one- and two-vehicle households, and then analyze and compare the results. In the tables below, HH stands for household; K stands for $1,000; # stands for number; and a dummy takes the value 1 when the condition is met, otherwise it is zero.

Table 8.7: New/Used: One-Vehicle Household

Explanatory Variables	Coefficients	t-stat
Purchase price (HH income>=31K)	-.0000347	-3.738
Purchase price (HH income<=30K)	-.0000808	-6.659
Operating Cost(HH income<=30K)	-.0131432	-0.578
Acceleration Time (HH income>= 61K)	-.2070171	-1.357
New-Car dummy (HH income>=46K)	1.620144	10.630
Subcompact car dummy	0.8942339	3.784
Compact car dummy	2.417079	11.842
Mid-size car dummy	2.620264	12.707
Large-size car dummy	1.300304	5.310
Luxurious car dummy (HH income>=31K)	2.499933	6.113
Luxurious car dummy (HH income<=30K)	2.03783	5.992
Sport car dummy	1.209337	5.259
Pickup dummy	0.8655282	3.895
Van dummy (HH size>=4)	2.76883	8.941
Van dummy (HH size<=3)	0.4173439	1.665
Utility vehicle dummy	0.1605688	0.670
Alternative-dispose constant for HHs with at least one member's age>60	2.644101	7.420
3-to-4-yr-old vehicle	-1.425501	-12.687
5-yr-old vehicle	-1.652665	-10.537
6-to-8-yr-old vehicle	-1.733317	-13.194
9-to-21-yr-old vehicle	-1.804498	-11.354
Number of observations Initial Likelihood Final Likelihood "Rho-Square" w.r.t. Zero	1098 -3394.6425 -2531.2291 0.2543	

Table 8.8: New/Used: Two-Vehicle Households

Explanatory Variables	Coefficients	t-stat
Purchase price(HH income>=31K)	-.0000579	-7.090
Purchase price(HH income<=30K)	-.0000841	-4.831
Operating cost (HH income<=75K)	-.0138858	-0.587
Acceleration Time(HH income>=61K)	-.2308725	-1.818
Top Speed	.009253	1.869
New-Car dummy (HH income>=31K)	2.703064	18.564
New-Car dummy(HH income<=30K)	1.771132	7.580
Subcompact car dummy	1.384278	4.088
Compact car dummy	2.770002	9.047
Mid-size car dummy	3.390846	11.195
Large-size car dummy	2.396474	7.452
Luxurious car dummy (HH income>=31K)	3.925489	10.333
Luxurious car dummy (HH income<=30K)	3.384701	8.135
Sport car dummy	1.554859	4.605
Pickup dummy	2.18398	7.104
Van dummy (HH size>=3)	3.29098	10.127
Van dummy (HH size<=3)	1.9758	6.121
Utility vehicle dummy	1.75092	5.600
Alternative-dispose constant for HHs with at least one member's age>=60	1.520025	1.562
3-to-5-yr-old vehicle	-.1851662	-1.192
6-yr-old vehicle	-.2557022	-1.349
7-yr-old vehicle	-.4356248	-2.187
9-to-21-yr-old vehicle	-.6930717	-3.184
Number of observations	1287	
Initial Likelihood	-3918.3004	
Final Likelihood	-2723.6236	
"Rho-Square" w.r.t. Zero	0.3049	

8.6.5 Analysis

Purchase Price

Households with income greater than $30,000 are less sensitive to purchase price than households with income less than or equal to $30,000.

Operating Cost

Households with income greater than $30,000 are less sensitive to operating cost than households with income less than or equal to $30,000.

Acceleration Time

Households with income greater than $30,000 are more sensitive to acceleration time than households with income less than or equal to $30,000.

Top Speed

This coefficient has an expected sign, but is not significant.

New Car Dummy

Households with income greater $30,000 are more likely to buy a new vehicle than households with income less than or equal to $30,000.

Vehicle Type

For one-vehicle households, the preference of body type in decreasing order is mid-size, compacts, large vehicles, sport, subcompact, and utility vehicles.

For two-vehicle households, the order is, from large to small, mid-size, compact, large, utility, sport, and subcompacts.

Luxury Vehicle

Households with income greater than $30,000 are more likely to buy a luxury car than households with income less than or equal to $30,000.

Van Dummy

Households with 4 or more people are more likely to buy a van than those with 3 or less people.

Dispose

Households with at least one driver's age is over 60 are more likely to dispose of a vehicle.

Vehicle Model Year

The coefficients have expected signs and, as expected, become smaller when vehicles get older.

APPENDIX 8-1
Vehicle Choice Survey Question

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

	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 on CNG	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 Fuel Cost	2 cents per mile (50 MPG gasoline equiv.) for recharging between 6 pm and 10 am 10 cents per mile (10 MPG gasoline equiv.) for recharging between 10 am and 6 pm	4 cents per mile (25 MPG gasoline equiv.)	
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 equiv.)	4 cents per mile (25 MPG gasoline equiv.)	4 cents per mile (25 MPG gasoline equiv.)
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 new1993 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

1. 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)

2. Would this vehicle most likely be purchased as a replacement vehicle for your household, or as an additional vehicle?

    1) Replacement   2) Additional

3. If you choose "Replacement" in Question 2, please cross off the household vehicle that would be replaced from the following list:

    1) 1990 Ford Bronco   2) 1989 Toyota Camry
    3) ...

9. HOUSEHOLD VEHICLE USAGE ADJUSTMENT

9.1 Objectives

Household vehicle usage behavior *by type of vehicle* is modeled in the research reported here. Forecasts of future vehicle emissions, including potential gains that might be attributed to introductions of alternative-fuel (clean-fuel) vehicles, critically depend upon the ability to forecast vehicle miles of travel by the fuel type, body style and size, and vintage of the vehicle.

Households acquire different vehicles to satisfy both the transportation needs and the preferences of the household members. Consequently, vehicle usage by type of vehicle can be considered to be a function of three categories of variables: (1) household characteristics, (2) principal driver characteristics, and (3) characteristics of the vehicle itself. Examples of household variables are income, residential location, number of vehicles, number of drivers, number of workers, and number of household members by age group.

Important driver characteristics include age, gender, and employment status. Usage of a specific vehicle depends heavily on which household member typically drives the vehicle. Workers, younger persons, and males are likely to drive more, as demonstrated in several usage models (Hensher, 1985; Hensher, *et al.*, 1992; Mannering, 1983; Mannering and Winston, 1985; and Train, 1986).

Usage is also affected by vehicle characteristics, such as vehicle age (vintage), operating cost, passenger and cargo capacity, body style, and value. Moreover, alternative-fuel vehicles are distinguished by vehicle attributes that are potentially even more important influences on usage patterns, such as limited range between refueling, coupled with limited fuel availability or the necessity to refuel or recharge the vehicle at home overnight. Differences between conventional-fuel and alternative-fuel vehicles in terms of fuel costs, cargo capacity, performance, and image is also expected to influence vehicle usage (van Wissen and Golob, 1992).

Applying a vehicle-type usage model in travel demand forecasts requires obtaining or developing forecasts of all of the model's exogenous variables. The first category of variables, household characteristics, can be readily forecast using Census data or household sociodemographic models used in regional planning. For example, the usage model developed here is driven as part of a microsimulation system (Brownstone, *et al.*, 1994) by a competing-risk hazard model of changing household demographics (Kazimi, 1995; Kazimi and Brownstone, 1995).

Forecasts of the second category of explanatory variables, principal driver characteristics, are problematic for multi-vehicle households, and for single-vehicle households with more than one driver. For such households, vehicle usage behavior involves allocating vehicles to drivers in satisfying activity needs (Golob, *et al.,* 1995). And for multi-vehicle households, usage behavior involves distributing total travel among the vehicles and drivers. While, in principle, forecasts of household and vehicle characteristics are all that are needed to forecast vehicle usage for single-vehicle households with only one driver, exogenous forecasts of principal driver characteristics for individual vehicles in multi-vehicle and multi-driver households are not feasible. The present usage models incorporates driver allocation models so that usage can be forecast by specific vehicle type while taking into account important principal driver characteristics.

Finally, exogenous forecasts of household vehicle holdings by type of vehicle are obtainable using vehicle type choice models, such as those developed by Lave and Train (1979), Manski and Sherman (1980), Hensher and Manefield (1982), Hocherman, et al. (1983), Berkovec (1985), Hensher and Le Plastrier (1985), Mannering and Winston (1985), Train (1986), McCarthy and Tay (1989), Hensher, *et al.* (1992). Such vehicle type-choice models are based on vehicle holdings and transactions data (so called "revealed-preference" or RP models). Because consumers do not have actual experience with alternative-fuel vehicles of the type that are likely to be available in 1998 and beyond, a vehicle type choice model based on stated preference (SP) data is required to forecast demand for these new vehicle types. One such model (Ren, *et al.*, 1995) is being coupled with the present usage models to forecast alternative-fuel vehicle usage for the State of California.

The current model is similar to previous models of vehicle allocation and use in multi-vehicle households (Mannering, 1983; Hensher, 1985; Train, 1986; and Hensher, *et al.*, 1992) in that separate equations with correlated error terms are developed for each vehicle in the household. However, this research differs from previous efforts because there are additional equations for principal-driver characteristics that cannot be readily forecast and need to be "solved out" of the problem; reduced-form equations needed for forecasting purposes are developed through a structural specification of vehicle allocation to drivers. This research is also unique in that the models use both RP and SP data simultaneously; the models are estimated with a mix of RP and SP observations.

The household's choices of the number of vehicles to own and the types of these vehicles are taken as given in these models. This is theoretically incorrect (as described in Golob, *et al.*, 1995), because a household's travel requirements will influence its choice of vehicle type. If the error terms of the vehicle type choice model and the vehicle usage model are correlated, the parameter estimates will be biased. A linear correction term involving a transformation of predicted vehicle type choice probabilities can be applied to the usage model to account for self selectivity bias (McFadden et al., 1985; Mannering and Winston, 1985; Train, 1986; Hensher, *et al.*, 1992). Empirically, however, the selectivity corrections applied in utilization models to account for endogeneity bias have not had substantial effects on estimation results

(Train, 1986; Hensher, 1992). The structural equation system developed here can be estimated jointly with vehicle type choice, or it can accept correction terms that are linear-in-parameters; this is relegated to future research.

9.2 Data

The data are from a 1993 survey conducted using geographically stratified pure random digit dialing. The survey, covering most of urbanized California excluding San Diego County, was composed of three distinct components, as described in Brownstone, *et al.* (1994) and Golob, *et al.* (1995). An initial computer-aided telephone interview (CATI) collected information on household structure, vehicle inventory, housing characteristics, employment data, commuting for all workers and students, and information about the intended next vehicle transaction. These CATI data were then used to produce a customized mail-out questionnaire which asked detailed questions each household member's commuting and vehicle usage.

The mail-out questionnaire also contained two SP (stated preference) vehicle type choice experiments for each household. Each of these experiments described three hypothetical vehicles, from which households were asked to choose their preferred vehicle and indicate whether the chosen vehicle would replace an existing household vehicle or be added to the household fleet. These hypothetical vehicles included both alternative-fuel and gasoline vehicles, and the body types and prices were customized to be similar (but not identical) to the household's description of their next intended vehicle purchase. Vehicle usage SP questions followed the type experiment. Responses to the customized mail-out questionnaire were collected in the third part of the survey, a follow-on CATI survey.

Of the 7,387 households that completed the initial CATI survey, 66%, or 4747 households, successfully completed the mail-out portion of the survey. A comparison with Census data reveals that the sample is slightly biased toward home-owning larger households with higher incomes, and weights are being developed to balance the sample to the known population (Brownstone et al., 1994). An unweighted sample is used here.

The breakdown by vehicle ownership level was: 1% zero vehicles, 34% one vehicle, 47% two vehicles, 13% three vehicles, and 5% four or more vehicles. For one-vehicle households, 75% had exactly one driver, while 25% had two or more drivers. Thus, approximately 73% of the households in the sample were either multi-vehicle or single-vehicle/multiple-driver, where driver allocation behavior is relevant.

The model variables are divided into three groups: (1) *behavioral* vehicle usage characteristics, capturing the ways in which households use their vehicles, (2) *physical* vehicle characteristics and (3) household structural characteristics. Vehicle usage for each household's present vehicles (RP usage data) is self-reported in terms of "How many miles per year is this vehicle driven?" It would be more accurate to calculate annual usage from vehicle odometer readings one year apart, but such data are not available in a cross-sectional survey. Vehicle usage for the hypothetical future vehicles (SP usage data) was collected through a series of questions asking how many miles the vehicle chosen in a choice experiment would e driver each week, and who in the household would typically use the chosen vehicle to commute to work or school.

Exogenous forecasts must be available for all household variables selected. The variable "Mean age of household heads," was computed as the mean of the ages of mates in spousal-like households, or the age of the single parent or person who can be identified as the major income-earner. The dummy variable "Household heads are retired" is set to one if one or both household heads are retired and neither household head is employed or temporarily unemployed; it is possible that another person, perhaps a grown child, is employed in such a household.

Separate models are developed for single-vehicle households and multi-vehicle households. The sample sizes are: 2,260 single-vehicle observations, comprised of households currently holding one vehicle (RP data) and households ending up with one vehicle after the SP choice task; 3,150 multi-vehicle observations, comprised of households currently holding two or more vehicles (RP data) and households ending up with two or more vehicles after the SP choice task. (In addition, a model of third-vehicle use was developed using a sample of 445 households with 3 or more vehicles, but this model is not reported here.) Each of the samples consists of households with known type and vintage of their current vehicle (single-vehicle sample), or no missing data on the newest two vehicles in their fleet (multi-vehicle sample). It is also required that there be no missing data on the age, sex and employment status of the principal drivers of each of these vehicles.

9.3 Specification

9.3.1 Partition of the Variables into Endogenous and Exogenous Sets

A distinguishing feature of this research is the endogenous treatment of driver allocation behavior. In order to avoid omitted-variables bias, vehicle miles of travel (VMT) is specified as a function of principal driver characteristics in addition to exogenous household and vehicle type characteristics. However principal driver characteristics are also specified as a function of the exogenous variables. This allows the principal driver characteristics, for which no exogenous forecasts are available, to be replaced by their predictors in the final forecasting equations.

There are four endogenous variables for each vehicle. These are listed in Table 1. In the multi-vehicle case, household vehicles are arranged such that the newest of the vehicles is defined as "vehicle 1," described by the first four endogenous variables and the first group of vehicle-type exogenous variables. The second-newest vehicle is defined to be the "vehicle 2" and is described by the last four endogenous variables and the last group of vehicle-type variables. If two vehicles are of the same vintage, the order of listing by the respondent is preserved.

The exogenous variables in each model are divided into two blocks: physical vehicle characteristics and household characteristics. The first block, listed in Table 2, is made up of 16 physical vehicle characteristics for each vehicle.

Table 9.1: Endogenous Variables for Each Vehicle

Variable	Acronym
Natural log of vehicle miles traveled per year	Ln(VMT)
Age of principal driver in years	Driver Age
Gender of principal driver (0 = male; 1 = female)	Driver Gender
Employment status of principal driver (1 = working)	Driver Empl St.

Table 9.2: Exogenous Variables for Each Vehicle

Variable	Acronym
Vehicle Age (in years from 1993)	Vehicle Age
Mini class (dummy)	Type: Mini
Subcompact car class (dummy)	Type: Subcompact
Compact car class (dummy)	Type: Compact
Mid-size or full-size car class (dummy)	Type: Mid-size
Full-size (standard) car class (dummy)	Type: Full-size
Sports car (dummy)	Type: Sports Car
Compact pickup truck (dummy)	Type: Small Truck
Full-size (standard) pickup truck (dummy)	Type: Std. Truck
Minivan (compact van) (dummy)	Type: Minivan
Full-size (standard) van (dummy)	Type: Van
Compact sport utility vehicle (dummy)	Type: Compact SUV.
Full-size (standard) sport utility vehicle (dummy)	Type: Full-size SUV.
Operating cost per mile (in cents)	Operating Cost
Electric vehicle (dummy)	Electric Vehicle
Range between refueling in miles	Range

The second block of exogenous variables is comprised of up to eleven household characteristics. This list is reproduced with associated acronyms for further reference in Table 3. The dummy variable for three or more vehicles is used only in the two-vehicle model. These variables, together with the driver characteristic variables listed in Table 1, were selected on the basis of published vehicle usage model results (Mannering, 1983; Hensher, 1985; Mannering and Winston, 1985; Hensher and Smith, 1986; Train, 1986; Golob, 1990; Hensher, at al., 1992).

Table 9.3: Exogenous Variables - Household Characteristics

Variable	Acronym
Number of household members aged 16-20	No. 16-20 Yr. Olds
Total number of household members > 15 years	No. 16+ Yr. Olds
Number of children in household aged 0 to 5	No. less than 5 Yrs. Old
Total number of children in household	Total no. of Kids
Household income less than $31,000 (dummy)	Income < $31k
Household income more than $60,000 (dummy)	Income > $60k
Household is a couple	Couple HH
Household head(s) are retired (dummy)	Retired HH
Mean age of household heads	Ave. Age of Heads
Total Number of workers in household	No. Heads Working
Household has three or more vehicles (dummy)	3+ Vehicle HH

9.3.2 The Structural Equation Model Form

The standard structural equations model (without latent variables) is given by

$$\mathbf{y} = \mathrm{B}\mathbf{y} + \Gamma\mathbf{x} + \zeta \qquad (1)$$

where $\mathbf{y}$ is an mx1 column vector of endogenous variables, and $\mathbf{x}$ is an nx1 column vector of exogenous variables. The structural parameters are the elements of the matrices:

$$\underset{(m\times m)}{\mathrm{B}} = \text{matrix of causal links between the endogenous variables,}$$

and

$$\underset{(m\times n)}{\Gamma} = \text{matrix of direct causal effects of the exogenous variables.}$$

And the error-term parameters are the elements of the variance-covariance matrix:

$$\underset{(m\times m)}{\Psi} = E(\zeta\zeta') \quad = \text{ symmetric variance-covariances matrix of unexplained, or unique, terms of the endogenous variables.}$$

For identification of system (1), it must be assumed that **B** is chosen such that **(I-B)** is non-singular, where **I** denotes the identity matrix of rank m.

The total effects of the endogenous variable on each other is given by

$$\mathrm{H} = (\mathbf{I} - \mathrm{B})^{-1} - \mathbf{I}\,. \tag{2}$$

And the total effects of the exogenous variables on the endogenous variables in a structural equations model of this type are given by

$$\mathrm{T} = (\mathbf{I} - \mathrm{B})^{-1}\,\Gamma. \tag{3}$$

These are the so-called reduced-form equations.

9.3.3 Division of the Problem into Separate Models

Comparisons of sample sizes to the number of variables and potential number of parameters revealed that separate models could be developed for single-vehicle households and for two-vehicle households. However, the number of households with three or more vehicles was insufficient for the development of a three-vehicle model. The alternative was to expand the two-vehicle model to cover households with two *or more* vehicles, and to add a third-vehicle model for households with three or more vehicles. It would be better to model utilization of three vehicles simultaneously, but the expansion of the present structural equation system to 12 endogenous variables and up to 58 exogenous variables (16 for each vehicle plus 10 household variables) is infeasible with the present data. The use of a Third-Vehicle Model (not reported here), with only four endogenous and 26 exogenous variables, is a pragmatic solution to the problem. The Multi-Vehicle Model is the most complex, and its specification is described here in detail. The Single-Vehicle Model is a simplification of the Two-Vehicle Model.

9.3.4 Specification of the Multi-Vehicle Model

Each model specification can be subdivided into endogenous effects, the **B** matrix in equation system (1), exogenous effects (the Γ matrix), and error-term variance-covariances (the Ψ matrix). This specification is based on structure of the RP (revealed-preference) utilization model developed in Golob, *et al.* (1995), but we introduce several new features relevant to joint SP-RP estimation. The postulated causal relationships between the endogenous variables are depicted in Table 4. There are two types of effects: within-vehicle effects and between-vehicle effects.

The within-vehicle effects are those in the upper left-hand (first vehicle) and lower right-hand (second vehicle) quadrants of the **B** matrix. Each of these effects is expected to be identical for the two vehicles, and equality restrictions are specified for corresponding pairs of **B**-matrix parameters. Use is postulated to be less for vehicles primarily driven by older persons ($\beta_{1,2} = \beta_{5,6}$), and women ($\beta_{1,3} = \beta_{5,7}$), and use is postulated to be greater for vehicles primarily driven by employed persons ($\beta_{1,3} = \beta_{5,8}$). Male principal drivers are more likely to be employed ($\beta_{4,3} = \beta_{8,7}$), as are younger principal drivers ($\beta_{4,2} = \beta_{8,6}$), and older drivers are expected to be male ($\beta_{3,2} = \beta_{7,6}$). An important feature of this specification is that, for each of the two household vehicles, VMT is postulated to be a function of all three of the principal driver variables. Thus, while driver allocation is endogenous, VMT is a function of driver characteristics.

Table 9.4: Multi-Vehicle Model
Postulated Direct Effects Between Endogenous Variables

	Influencing Variable							
Influenced variable	Ln (VMT_1)	Driver Age_1	Driver $Gender_1$	Driver Empl St_1	Ln (VMT_2)	Driver Age_2	Driver $Gender_2$	Driver Empl St_2
Ln (VMT_1)		$\beta_{1,2}$ (-)	$\beta_{1,3}$ (-)	$\beta_{1,4}$ (+)				
Driver Age_1								$\beta_{2,8}$ (-)
Driver $Gender_1$		$\beta_{3,2}$ (-)					$\beta_{3,7}$ (-)	
Driver Empl St_1		$\beta_{4,2}$ (-)	$\beta_{4,3}$ (-)					$\beta_{4,8}$ (-)
Ln (VMT_2)						$\beta_{5,6}=\beta_{1,2}$	$\beta_{5,7}=\beta_{1,3}$	$\beta_{5,8}=\beta_{1,4}$
Driver Age_2				$\beta_{6,4}=\beta_{2,8}$				
Driver $Gender_2$			$\beta_{7,3}=\beta_{3,7}$			$\beta_{7,6}=\beta_{3,2}$		
Driver Empl St_2				$\beta_{8,4}=\beta_{4,8}$		$\beta_{8,6}=\beta_{4,2}$	$\beta_{8,7}=\beta_{4,3}$	

Regarding the between-vehicle effects, we expect strong negative relationships between principal-driver genders and employment status, and this is operationalized by specifying equated pairs of reciprocal effects ($\beta_{3,7} = \beta_{7,3}$) and ($\beta_{4,8} = \beta_{8,4}$). And the ages and employment stutuses of the two drivers are expected to be negatively related ($\beta_{2,8} = \beta_{6,4}$). The postulated model is parsimonious in that it has only nine free parameters in the **B** matrix, representing nine pairs of equated direct effects.

The postulated structure of the vehicle-characteristic exogenous effects is depicted in Table 5. This vehicle-type effects specified in the exogenous variable structure were developed by considering vehicle usage stereotypes. For example, there are typically more male principal drivers of compact and full-size pickup trucks, subcompact cars might have younger principal drivers, and minivans are likely to be driven by females. Logically, older vehicles and higher operating cost vehicles should be driven less.

The major restrictions applied in specifying these exogenous vehicle type influences are that the effects be the same for the two vehicles. It is a straightforward procedure to subsequently test whether the model can be significantly improved by releasing these cross-vehicle parameter equality restrictions. It is also quite possible that the characteristics of the first vehicle can affect the VMT and principal driver characteristics of the second vehicle, and conversely. The model was initially specified by setting all such cross-vehicle effects to zero. Tests were then conducted to ascertain whether cross-vehicle effects significantly improved model fit.

Examples of direct household effects to be tested include: Principal drivers in households with more workers and in high-income households are more likely to be employed. Usage is higher in households with more children and in high-income households. Principal drivers are younger in households with young children. Drivers in retired households are older and are less likely to be employed (although some drivers in retired households, such as adult children living with their parents, could be employed). Finally, households with three or more vehicles have lower levels of usage on their first and second vehicles, all else held constant. The default restriction on all of these postulated household influences involves equating the corresponding effects on the two vehicles, and then testing whether the relaxation of each equality results in a significant model improvement.

The final specification step involves the error-term variance-covariance matrix Ψ. If the unique (error) component of any one of the four endogenous variables of the first vehicle is correlated with the unique component of the corresponding variable for the second vehicle, then we should find statistically significant coefficients for the Ψ matrix terms $\psi_{5,1}$, $\psi_{6,2}$, $\psi_{7,3}$, or $\psi_{8,7}$. That is, if what is not explained about a variable for one vehicle is correlated with what is not explained about the same variable for the other vehicle, these sub-diagonal parameters should be found to be significant. The freely estimated main-diagonal variances of the Ψ matrix produce R^2 values:

$$R^2 = (s_{i,i} - \psi_{i,i}) / s_{i,i} \qquad (4)$$

where $s_{i,i}$ is the sample variance of endogenous variable i.

Table 9.5: Multi-Vehicle Model
Postulated Direct Effects of the Exogenous Variables

<table>
<tr><th rowspan="2">Exogenous Variable</th><th colspan="8">Endogenous Variable</th></tr>
<tr><th>Ln (VMT_1)</th><th>Driver Age_1</th><th>Driver $Gender_1$</th><th>Driver Empl St_1</th><th>Ln (VMT_2)</th><th>Driver Age_2</th><th>Driver $Gender_2$</th><th>Driver Empl St_2</th></tr>
<tr><td>Vehicle Age_1</td><td colspan="4" rowspan="16">First vehicle:
vehicle characteristic effects on VMT and relationships between vehicle characteristics and principal driver characteristics

(relatively dense submatrix, equated with 2nd vehicle effects)</td><td colspan="4" rowspan="16">Effects of characteristics of the 1st vehicle on VMT and principal driver allocation of the 2nd vehicle

(relatively sparse submatrix, initially specified null)</td></tr>
<tr><td>$Type_1$: Mini car</td></tr>
<tr><td>$Type_1$: Subcompact</td></tr>
<tr><td>$Type_1$: Compact</td></tr>
<tr><td>$Type_1$: Mid-size car</td></tr>
<tr><td>$Type_1$: Full-size car</td></tr>
<tr><td>$Type_1$: Sports car</td></tr>
<tr><td>$Type_1$: Small Truck</td></tr>
<tr><td>$Type_1$: Std. Truck</td></tr>
<tr><td>$Type_1$: Minivan</td></tr>
<tr><td>$Type_1$: Std. Van</td></tr>
<tr><td>$Type_1$: Small SUV</td></tr>
<tr><td>$Type_1$: Std. SUV.</td></tr>
<tr><td>Operating $Cost_1$</td></tr>
<tr><td>Electric $Vehicle_1$</td></tr>
<tr><td>$Range_1$</td></tr>
<tr><td>Vehicle Age_2</td><td colspan="4" rowspan="16">Effects of characteristics of the 2nd vehicle on VMT and principal driver allocation of the 1st vehicle

(relatively sparse submatrix, initially specified null)</td><td colspan="4" rowspan="16">Second vehicle:
vehicle characteristic effects on VMT and relationships between vehicle characteristics and principal driver characteristics

(relatively dense submatrix, equated with 1st vehicle effects)</td></tr>
<tr><td>$Type_2$: Mini car</td></tr>
<tr><td>$Type_2$: Subcompact</td></tr>
<tr><td>$Type_2$: Compact</td></tr>
<tr><td>$Type_2$: Mid-size car</td></tr>
<tr><td>$Type_2$: Full-size car</td></tr>
<tr><td>$Type_2$: Sports car</td></tr>
<tr><td>$Type_2$: Small Truck</td></tr>
<tr><td>$Type_2$: Std. Truck</td></tr>
<tr><td>$Type_2$: Minivan</td></tr>
<tr><td>$Type_2$: Std. Van</td></tr>
<tr><td>$Type_2$: Small SUV.</td></tr>
<tr><td>$Type_2$: Std. SUV.</td></tr>
<tr><td>Operating $Cost_2$</td></tr>
<tr><td>Electric $Vehicle_2$</td></tr>
<tr><td>$Range_2$</td></tr>
<tr><td>No. 16-20 Yr. Olds</td><td colspan="4" rowspan="7">Effects of household characteristics on VMT and principal driver allocation of 1st vehicle
(equated across vehicles)</td><td colspan="4" rowspan="7">Effects of household characteristics on VMT and principal driver allocation of 2nd vehicle
(equated across vehicles)</td></tr>
<tr><td>No. of 16+ Yr. Olds</td></tr>
<tr><td>No. 1-5 Yr. Olds</td></tr>
<tr><td>Total No. of Kids</td></tr>
<tr><td>Income > $60k</td></tr>
<tr><td>Ave. Age of Heads</td></tr>
<tr><td>3+ Vehicle HH</td></tr>
</table>

9.4 Estimation Method

Structural equations systems of this type can be generally estimated using methods of moments (also known as, variance analysis methods). The method proceeds by defining the sample variance-covariance matrix of the combined set of endogenous and exogenous variables, partitioned with the endogenous variables first:

$$\mathbf{S} = \begin{bmatrix} S_{yy} & S_{yx} \\ S'_{yx} & S_{xx} \end{bmatrix}, \tag{5}$$

where $\mathbf{S_{yy}}$ denotes the variance-covariance matrix of the endogenous variables, $\mathbf{S_{yx}}$ denotes the covariance matrix between the endogenous and exogenous variables, and $\mathbf{S_{xx}}$ denotes the variance-covariance matrix of the exogenous variables. In the Multi-Vehicle Model, there are 8 endogenous variables and 38 exogenous variables, so $\mathbf{S}$ is a (46 by 46) symmetric matrix.

It can be easily shown using matrix algebra that the corresponding variance-covariance matrix replicated by model system (1), denoted by

$$\Sigma = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma'_{yx} & \Sigma_{xx} \end{bmatrix}, \tag{6}$$

is:

$$\Sigma_{yy} = (I - B)^{-1}(\Gamma S_{xx}\Gamma' + \Psi)\left((I - B)^{-1}\right)', \tag{7}$$

$$\Sigma_{yx} = (I - B)^{-1}\Gamma S_{xx}, \tag{8}$$

and $\Sigma_{xx} = S_{xx}$ is taken as given.

The structural equation system here is estimated using the normal-theory maximum likelihood method (Bollen, 1989). The fitting function for structural equations maximum likelihood (ML) estimation is

$$F_{ML} = \text{Log}|\Sigma(\theta)| - \text{Log}|S| + \text{tr}[S\Sigma(\theta)] - (p + q) , \tag{9}$$

where the components are defined in (1) through (8) above. This fitting function $\mathbf{F_{ML}}$ is two times the log-likelihood ratio divided by the sample size, n. Under assumptions of multivariate normality, $n\mathbf{F_{ML}}$ is X^2-distributed, providing a test of model rejection and criteria for testing hierarchical models. Function (9) is minimized in the LISREL8 program using a modified Fletcher-Powell algorithm (Jöreskog and Sörbom, 1993a).

Because four of the eight endogenous variables are dichotomous, the coefficient estimates will be consistent, but the estimates of parameter standard errors for certain coefficients and the overall model Chi-square goodness-of-fit will be biased (Bentler and Bonett, 1980). Unbiased estimates can be generated using the asymptotically distribution-free weighted least squares method (Browne, 1982, 1984), but this requires a much larger sample size. (The rule-of-thumb is that the sample size must be at least three times greater than the number of free entries in the asymptotic variance-covariance matrix of the correlation matrix, the fourth order moments; with 36 variables, this requires approximately 3,250 observations.) However, ML estimates will be consistent, and they have been shown to be fairly robust (Boomsma, 1983). Furthermore, the two endogenous variables of most interest are continuous.

9.5 Results: Multi-Vehicle Model

9.5.1 Model Fit and Structure

The Two-Vehicle Model fits extremely well according to all goodness-of-fit criteria. The chi-square distributed, -2 log likelihood ratio is 210.5 with 237 degrees of freedom, corresponding to a probability value of 0.892. Thus, the model *cannot* be rejected at the p = .05 level. The estimated R^2 value for VMT of the first (newest) vehicle is 0.115, and that of the second (oldest) vehicle is 0.131. As expected, significant positive error-term covariances were found between the usages of the two vehicles (*t-statistic* = 11.3), between principal driver ages (*t-statistic* = 8.4), and between principal driver genders (*t-statistic* = 15.2).

The estimated direct effects between endogenous variables are listed with their *t*-statistics in Table 6. This endogenous variable structure model is basically in accordance with the hypotheses depicted in Table 4, with a few exceptions that can be identified by comparing Tables 4 and 6. All six of the within-vehicle endogenous-variable effects postulated for each vehicle were found to be statistically significant and five of the six effects are equal across the two vehicles. The three postulated cross-vehicle effects were also found to be significant and symmetric. However, three additional cross-vehicle effects were found to be necessary for good model fit: If the driver of the first vehicle is older, use of the second vehicle is less than otherwise expected (effect $\beta_{5,2}$), and if the driver of the second vehicle is female, use of the first vehicle is greater than expected (effect $\beta_{5,2}$). Also, the employment status of the first-vehicle driver predicts the gender of the second-vehicle driver (effect $\beta_{7,4}$).

9.5.2 Total Effects

The total effects of the endogenous variables on the two vehicle usage variables are listed in Table 7. For simplicity, only the total effects on the two VMT variables, elements $\eta_{1,j}$ and $\eta_{5,j}$ (j = 1 to 8) of matrix H defined in equation system (2), are shown. Results show that driver age is a significant effect on vehicle usage that is uniform for the two vehicles; if *either* driver is younger, *both* the first and second vehicles are likely to be used more. In contrast, the gender and employment status effects are consistent and reciprocal across the two vehicles. If the principal driver of either vehicle is female, that vehicle is driven less and the other vehicle is driven more, and if either driver is employed, that vehicle is driven more, and the other vehicle is driven less, *ceteris paribus*. These reciprocal pairs of effects are generally strongest on the driver's own vehicle.

Table 9.6: Multi-Vehicle Model
Estimated Direct Effects Between Endogenous Variables (*t*-statistics in parentheses)
(Coefficients that are restricted to be equal for the two vehicles are shown in **bold**)

	Influencing Variable							
Influenced variable	Ln (VMT_1)	Driver Age_1	Driver $Gender_1$	Driver Empl St_1	Ln (VMT_2)	Driver Age_2	Driver $Gender_2$	Driver Empl St_2
Ln (VMT_1)		**-0.0043 (-4.05)**	**-0.131 (-6.52)**	**0.179 (4.54)**			0.0797 (2.70)	
Driver Age_1								**-2.81 (-5.85)**
Driver $Gender_1$		-0.0051 (-7.78)					**-0.693 (-21.6)**	
Driver Empl St_1		**-0.0065 (-10.3)**	**-0.103 (-11.3)**					**-0.140 (-15.9)**
Ln (VMT_2)		-0.0028 (-2.12)				**0.179 (4.54)**	**-0.131 (-6.52)**	**0.179 (4.54)**
Driver Age_2				**-2.81 (-5.85)**				
Driver $Gender_2$			**-0.693 (-21.6)**	0.506 ((3.72)		-0.0036 (-6.46)		
Driver Empl St_2				**-0.140 (-15.9)**		**-0.0065 (-10.3)**	**-0.103 (-11.3)**	

The total effects of the exogenous variables on the usage endogenous variables are listed in Table 8. These are the coefficients of the reduced-form equations for two of the eight endogenous variables, which are given by matrix equation (2).

The total effects of vehicle age on VMT are strongest on the second vehicle, but the effects are consistent for both vehicles: The older a vehicle is, the less it is used, *ceteris paribus*. Also, the older the first vehicle is, the less the other vehicle is used as well. The forecasting implication of this is reduced usage of the household fleet over time if no vehicle transactions occur. If household structure, income and employment does not change, the reduction in fleet will be further accentuated through the negative total effect on usage of driver age. It is implied that a household will acquire a newer one to accommodate new travel demand.

The effects of operating cost are negative, as expected, but these effects are imprecisely estimated. Also, operating cost of the second vehicle has the anticipated positive cross-vehicle positive effect on usage of the first vehicle, but the coefficient has a relatively high standard error. This is probably due to the colinearity between operating cost, vehicle age, and the vehicle-type dummies.

Table 9.7: Multi-Vehicle Model
Total Effects of the Other Endogenous Variables on the Two Usage Variables

	Influenced Variable			
Endogenous Variable	Ln (VMT_1)		Ln (VMT_2)	
	Total effect	*t*-statistic	Total effect	*t*-statistic
Driver Age_1	-0.00358	-3.36	-0.00363	-2.79
Driver $Gender_1$	-0.40013	-7.13	0.19975	5.33
Driver Empl St_1	0.20385	5.03	-0.02565	-3.10
Driver Age_2	-0.00116	-4.72	-0.00362	-2.91
Driver $Gender_2$	0.35896	5.60	-0.28460	-6.24
Driver Empl St_2	-0.02098	-3.26	0.15278	3.84

The effects of the electric vehicle (EV) dummy variable on VMT are potentially important for pollution and energy policies. If either vehicle in multi-vehicle households is a future EV, the model results imply that the EV will be driven less, *ceteris paribus.* Moreover, if the EV is the newest (first) vehicle in the household, the second vehicle will be driven more than otherwise expected. Thus, this model captures a shift in usage from EV's to conventional fuel vehicles, somewhat mitigating the emissions gains of the electricity verses conventional fuels. The magnitude of this cross-vehicle substitution effect can be assessed by using this utilization model for forecasting combined with demographic, vehicle transaction, and vehicle type-choice models (Brownstone, *at al.*, 1994).

The range variable also captures a reduced VMT effect for all limited-range vehicles (potentially including dedicated compressed natural gas vehicles in addition to EV's). For limited-range second vehicles, there is also a shift in usage from the second vehicle to the first vehicle.

Table 9.8: Multi-Vehicle Model
Total Effects of the Exogenous Variables on the Two Vehicle Usage Variables

Exogenous Variable	Total Effects			
	Ln (VMT_1)		Ln (VMT_2)	
	Total effect	*t*-statistic	Total effect	*t*-statistic
Vehicle Age_1	-0.01301	-2.98	-0.00095	-2.48
$Type_1$: Mini car	-0.23091	-5.33	-0.00043	-0.10
$Type_1$: Subcompact	0.01675	3.33	0.01983	3.49
$Type_1$: Compact car	0.08289	2.41	0.00710	2.39
$Type_1$: Mid-size car	-0.01416	-2.69	-0.06500	-1.57
$Type_1$: Full-size car	-0.08733	-1.57	-0.00872	-2.32
$Type_1$: Sports car	0.03025	3.97	0.01494	2.44
$Type_1$: Small truck	0.07210	8.35	-0.03599	-6.52
$Type_1$: Std. Truck	0.08037	8.37	-0.04012	-6.53
$Type_1$: Minivan	0.12686	2.44	0.01668	4.12
$Type_1$: Std. Van	0.02095	2.05	-0.01046	-2.01
$Type_1$: Small SUV.	0.23267	4.33	-0.02145	-3.25
$Type_1$: Std. SUV.	0.07242	6.74	-0.02177	-2.80
Operating $Cost_1$	-0.00057	-1.36	-0.00058	-1.31
Electric $Vehicle_1$	-0.25025	-2.51	0.09420	1.25
$Range_1$	0.00153	3.30	0	- -
Vehicle Age_2	0.00443	1.14	-0.03372	-9.34
$Type_2$: Mini car	0.00323	3.19	-0.16784	-3.29
$Type_2$: Subcompact	-0.00290	-0.59	0.02421	3.99
$Type_2$: Compact car	0.00140	2.26	0.00436	1.93
$Type_2$: Mid-size car	0.01319	2.46	-0.01784	-3.21
$Type_2$: Full-size car	0	- -	0	- -
$Type_2$: Sports car	-0.09826	-2.12	0.02713	3.63
$Type_2$: Small truck	-0.06246	-6.43	0.05819	6.99
$Type_2$: Std. Truck	-0.08374	-6.86	0.05863	6.70
$Type_2$: Minivan	0	- -	0	- -
$Type_2$: Std. Van	-0.01879	-2.01	0.01490	2.02
$Type_2$: Small SUV.	-0.03829	-4.68	0.12405	2.17
$Type_2$: Std. SUV.	-0.05326	-5.30	0.05463	5.84
Operating $Cost_2$	0.00427	0.626	-0.00860	-1.19
Electric $Vehicle_2$	0	- -	-0.22579	-1.20
$Range_2$	-0.00096	-1.53	0.00072	0.81
No. of 16-20 Yr. Olds	0.00956	4.80	0.00822	1.55
No. less than 5 Yrs. Old	-0.00917	-3.41	0.04667	1.95
Total no. of Kids	0.03060	3.30	0.03662	3.88
Income>$60k	0.11339	4.13	0.08506	3.10
Retired HH	-0.05129	-4.84	-0.04452	-4.19
Ave. Age of Heads	-0.00350	-4.24	-0.00545	-5.88
No. Heads Working	0.11234	5.51	0.10618	5.20
3+ Vehicle HH	0	- -	-0.04580	-1.59

The number of household members between 16 and 20 years old has a positive influence on VMT of both the first and second vehicle. The number of children 1 to 5 years old positively influences VMT mostly of the second vehicle, while the total number of children positively influences VMT of *both* the first and second vehicles. The income effect has the expected sign, but, as in the case of average age of the heads, the effects are imprecisely estimated. Finally, as expected, the presence of three or more household vehicles reduces VMT of both the first and second vehicles.

9.5.3 Scenarios of Changes in VMT Implied by the Total Effects

The endogenous variables are in terms of the natural logarithms of VMT, so the natural exponent of each reduced-form equation coefficient represents a multiplicative factor applied to the endogenous VMT variable in question. That is, $\exp(\delta_j \tau_{1,j})$ and $\exp(\delta_j \tau_{5,j})$ express multipliers of VMT for vehicle 1 and VMT for vehicle 2, respectively, where the T matrix of total exogenous effects is defined in equation system (3) and δ_j is the level of change in the jth exogenous variable. Some selected VMT multiplier effects are listed in Table 9.

Each scenario listed in Table 9 assumes that all factors not defined in the scenario remain constant. In the case of vehicle replacements, this includes the vehicle type class and operating cost. However, to provide realism, when vehicles are assumed to be replaced with identical vehicles with different ranges or fuels, it is assumed that the replacement vehicle is one year newer.

Of all of the model predictions computed in Table 9, the most substantial effects are those attributable to vehicle range and the electric vehicle (EV) designator. In the case of the first (newer) vehicle, a reduction in range of 150 miles reduces VMT by a factor of 0.81, but there is no effect on VMT of the second vehicle. In the case of the second vehicle, a similar reduction in range of 150 miles reduces VMT by a factor of only 0.93, but first-vehicle VMT is predicted to increase by a factor of 1.15. The weaker second-vehicle range effect is partially due to an offsetting stronger second-vehicle age effect. Combining reduced range with the EV effect, the model predicts that a if the first vehicle is an EV with 100 miles range, VMT will reduce by a factor of 0.58, and second-vehicle VMT will increase by a factor of 1.10. If the second vehicle is an EV with 100 miles range, VMT on this vehicle will reduce by a factor of 0.70, but there will be more of a shift to usage of the first vehicle, with first-vehicle VMT increasing by a factor of 1.24.

It must be noted that the SP data was not sufficient to capture the possible effects on VMT of fuel availability away from home for EV's and other non-gasoline limited-range vehicles operating. This might result in an overestimation of range and EV effects. The modest effect of operating cost on VMT means that the range and EV scenario results would not be substantially changed by imposing accompanying realistic changes in operating costs.

Table 9.9: Multi-Vehicle Model
Exponentiated Total Effects on VMT of Selected Changes in Exogenous Variables

	Exogenous change	Multiplier Effect on Veh. 1 VMT	Multiplier Effect on Veh. 2 VMT
	Vehicle Age_1 (in years)		
V	vehicle ages one year	0.98	0.99
E	replace with same type veh. 1 year newer	1.01	1.00
H	replace with same type veh. 5 years newer	1.07	1.01
I	Electric $Vehicle_1$ (EV_1)		
L	$Range_1$ (in miles)		
E	replace 300 mi. veh. with 200 mi. non-EV, 1 yr. newer	0.87	1.00
	replace 300 mi. veh. with 150 mi. non-EV, 1 yr. newer	0.81	1.00
1	replace 300 mi. veh. with 100 mi. EV, 1 yr. newer	0.58	1.10
	replace 300 mi. veh. with 75 mi. EV. 1 yr. newer	0.56	1.10
	Vehicle Age_2 (in years)		
V	vehicle ages one year	1.00	0.97
E	replace with same type veh. 1 year newer	0.99	1.03
H	replace with same type veh. 5 years newer	0.98	1.18
I	Electric $Vehicle_2$ (EV_2)		
L	$Range_2$ (in miles)		
E	replace 300 mi. veh. with 200 mi. non-EV, 1 yr. newer	1.10	0.96
	replace 300 mi. veh. with 150 mi. non-EV, 1 yr. newer	1.15	0.93
2	replace 300 mi. veh. with 100 mi. EV, 1 yr. newer	1.21	0.72
	replace 300 mi. veh. with 75 mi. EV, 1 yr. newer	1.24	0.70
	No. 16-20 Yr. Olds		
	child passes 16th birthday, no other changes	1.01	1.02
	No. less than 5 Yrs. Old		
H	Total no. of Kids		
O	birth of child, no other changes	1.02	1.09
U	19 yr. old child moves out of home, no other changes	0.96	0.96
S	Income>$60k		
E	Retired HH		
H	No. Heads Working		
O	income rises above $60k, no other changes	1.12	1.09
L	+1 head working *and* income rises above $60k	1.25	1.21
D	1 head working: retires, income drops below $60k	0.85	0.88
	2 heads retire at same time, income stays above $60k	0.90	0.92
	3+ Vehicle HH		
	household adds third vehicle	1.00	0.96
	household disposes of third vehicle	1.00	1.05

In contrast to the range effects, the vehicle aging effects are weaker for the first (newer) vehicle than for the second (older) vehicle. If the newest vehicle in the household is replaced with a vehicle that is identical in type, operating cost, range, and fuel, but is five years newer, the model predicts that VMT for that vehicle will increase by approximately seven percent, with very little effect on VMT of the second vehicle. However, if the second vehicle is replaced with a vehicle that is identical in type, operating cost, range, and fuel, but is five years newer, the model predicts that VMT for that vehicle will increase by approximately eighteen percent, and VMT of the first vehicle will slightly decrease (by about two percent).

The predicted changes in VMT associated with the scenarios related to the number of children in the household are smaller in magnitude than expected, but the usage behavior appears consistent with conventional notions of first- and second-vehicles. For example, a new child places more pressure on the use of the second vehicle, the one that is less likely to be used for commuting.

In contrast, the predicted effects of income and the number of household heads working are relatively strong, especially in combination. The joint impacts of an additional worker and a higher household are predicted increases of twenty-five percent in usage of the first household vehicle, approximately half of which is attributable to an income effect; usage of the second vehicle increases by a slightly lower twenty-one percent. If one working head retires *and* income drops below the high-income cut-off, the model predicts that VMT of the first and second vehicles will be reduced by factors 0.85 and 0.88, respectively. If both household heads quit working due to retirement, the predicted change in VMT is only ten percent for the first vehicle, providing that household income remains above (or below) the high-income cut-off. Finally, the presence of a third household vehicle has a modest influence on VMT of the second vehicle.

9.6 Results: The Single-Vehicle Model

9.6.1 Model Fit and Final Structure

The structure of the final Single-vehicle Model is also basically in accordance with the structural hypotheses. This Model also fits extremely well according to all goodness-of-fit criteria, the chi-square statistic being 41.82 with 49 degrees of freedom, corresponding to a probability value of 0.757. The model *cannot* be rejected at the p = .05 level. The estimated R^2 value for VMT is 0.173. No significant error-term covariances were found between any pairs of the four endogenous variables.

The endogenous variable structure determined to be optimal in the Single-Vehicle Model is similar to the within-vehicle structure found for the Multi-Vehicle Models (the structure depicted in the upper-left-hand and lower-right-hand quadrants of the **B** matrix shown in Table 6. The only difference was that an additional direct effect was found between principal driver gender and age: if the principal driver of the vehicle in a single-vehicle household is female, she is younger than otherwise expected, *ceteris paribus*.

9.6.2 Total Effects

The total effects of the endogenous principal-driver endogenous variables on VMT for the Single-Vehicle Model are listed in Table 10. As in the multi-vehicle case, VMT is higher for younger, male, employed drivers, but the gender and employment status effects are relatively weaker for single-vehicle households.

Table 9.10: Single-Vehicle Model
Total Effects of the Other Endogenous Variables on Vehicle Usage

Endogenous Variable	Total Effect on Ln(VMT)	
	Total effect	*t*-statistic
Driver Age	-0.00396	-1.00
Driver Gender	-0.08037	-2.30
Driver Empl. Status	0.11671	2.46

Finally, the total exogenous effects on VMT for the Single-Vehicle Model are listed in Table 11. Once again, these effects are similar to those found for multi-vehicle households, with some exceptions. Usage patterns are consistent for eight types of vehicles, but sportscars, minivans, standard sport utility vehicles, and full-size cars exhibit different usage patterns in single-vehicle, versus multi-vehicle, households. Regarding alternative-fuel vehicles, the negative EV effect and the positive effect of range on VMT are consistent between single-vehicle and multi-vehicle households.

9.7 A Forecasting Method That Preserves Heterogeneity

This model is being applied in a dynamic microsimulation forecasting system (Brownstone, *et al.*, 1994), in which a sociodemographic transition model and vehicle transactions models are being used to forecast changes in households' sociodemographic structure and composition of the vehicle fleets. The usage model is then exercised to forecast VMT for both the before- and after-situations for the household. The calculated change in forecasts is then applied as a percentage change to the actual base level of usage for the household in the before-situation.

Even if the dynamic sociodemographic model predicts no change in household characteristics (household composition, employment status, or income), and the vehicle transactions model predicts no vehicle transactions for the household for the period in question, the present usage model will in general predict changes in VMT. This will be due to aging of the household heads, aging of the vehicles, and possible changes in the age categories of household members, particularly children.

Table 9.11: Single-Vehicle Model
Total Effects of the Exogenous Variables on Vehicle Usage

Exogenous Variable	Total effect	*t*-statistic
Vehicle Age	-0.01574	-4.14
Type: Mini car	-0.27808	-5.97
Type: Subcompact	0.09798	1.82
Type: Compact car	0.12140	2.26
Type: Mid-size car	-0.00259	-0.91
Type: Full-size car	0.00639	1.38
Type: Sports car	-0.00706	-1.46
Type: Small truck	0.26612	2.64
Type: Std. Truck	0.52883	3.26
Type: Minivan	0.45711	3.86
Type: Std. Van	0.34705	1.64
Type: Small SUV.	0.31306	2.98
Type: Std. SUV.	0.00000	0.00
Operating Cost	-0.01223	-1.31
Electric Vehicle	-0.15136	-1.58
Range	0.00138	3.62
No. 16-20 Yr. Olds	0.04246	2.58
No. 16+ Yr. Olds	0.03455	0.93
No. less than 5 Yrs. Old	0.12448	2.31
Total no. of Kids	-0.11225	-3.78
Income < $31k	-0.19112	-5.19
Income > $60k	0.10970	1.90
Couple HH	0.00833	1.69
Retired HH	-0.02178	-1.67
Ave. Age of Heads	-0.01071	-8.46
No. Heads Working	0.05588	2.21

The most effective application of the usage models in a micro-simulation forecasting system uses a "pivot" approach, rather than the traditional approach of using the expected value from a linear model. The pivot approach preserves heterogeneity across households. Heterogeneity due to spatial and lifestyle factors is to be expected; some households drive more miles per year than the model would predict while others drive fewer miles per year than the model would predict.

Consider the example of a household, i, i = 1 to m, with one vehicle. This household reports annual VMT for its vehicle in the base year (1993), denoted by $VMT_{i,0}^{Obs}$. Using the Single-Vehicle Model, the expected (predicted) $VMT_{i,0}^{Pred}$ can be computed. In the next time period, denoted time period 1, the vehicle and the household have aged by six months, some household characteristics may have changed, and the household might even be predicted to replace the vehicle. The new predicted VMT would be $VMT_{i,1}^{Pred}$ computed from the Model applied to updated exogenous data. However, in period 0 the household's actual VMT differed from the predicted VMT by the residual ($VMT_{i,0}^{Obs}$ - $VMT_{i,0}^{Pred}$). This is extremely useful information, as household behavior tends to be autocorrelated over time. A better estimate of VMT in period 1 is then:

$$VMT_{i,0}^{Est} = VMT_{i,0}^{Obs} \, (1 + \Delta_{Model})), \tag{10}$$

where

$$\Delta_{Model} = (VMT_{i,1}^{Pred} - VMT_{i,0}^{Pred}) \, / \, VMT_{i,0}^{Pred} \tag{11}$$

Using this approach is equivalent to:

$$VMT_{i,0}^{Est} = \delta_i \; VMT_{i,n}^{Pred} \tag{12}$$

where

$$\delta_i = VMT_{i,0}^{Obs} / VMT_{i,0}^{Pred}. \tag{13}$$

The household-specific multiplier δ_i can be computed for each sample and stored using the base-year household data, and then used repeatedly throughout the microsimulation. If a sample household is missing initial observed VMT, δ_i can be assumed to be unity. This multiplier approach can be extended to all vehicle ownership situations.

9.8 Conclusions and Directions For Further Research

The structural elegance of the models and their statistical fit to the sample data of the models is judged to be a confirmation of the modeling approach. Moreover, the correspondence between pure RP results (Golob, *et al.*, 1995) and the present SP-RP results is encouraging. The evidence is that the effects of the SP variables, such as those capturing limited range and electric vehicle recharge requirements in the present application, are more reliably captured in the context of a joint SP-RP estimation. A joint estimation approach is preferred to either a stand-alone SP model or a models in which SP results are scaled to RP results.

The SP usage questions in the 1993 household survey from which these data were extracted potentially limit the effectiveness of the model in capturing effects on usage attributable to fuel availability, peak and off-peak recharge costs for EV's, cargo capacity, performance, and other vehicle and fuel-system characteristics that might distinguish future vehicles. A second household survey, conducted in 1994, contained a different vehicle usage SP protocol. When the 1994 data are available, the robustness of present model results can be assessed, and hopefully the model can be extended.

Selectivity bias can be accounted for in this usage model by linking the model to a discrete type-choice model (e.g., Ren, et al., 1995), and adding into the structural equation system a correction term variable involving a transformation of the household's predicted type vehicle choice probabilities (McFadden et al., 1985; Mannering and Winston, 1985; Train, 1986; Hensher, et al., 1992). It is doubtful that such a correction term would have a pronounced effect on the results.

The known biases in the normal-theory maximum likelihood estimation method applied to dichotomous endogenous variables are concentrated on coefficient standard errors and overall goodness-of-fit criteria. The fit of the model is not in question, and hypothesis testing is subordinate to forecasting capability in this research. However, it would be possible to use unbiased generally weighted least squares estimation (Browne, 1982, 1984), as implemented in LISREL8 with PRE-LIS2 (Jöreskog and Sörbom, 1993b), with an increased sample size.

10. FLEET SAMPLE WEIGHTING

Unlike the personal vehicle sample, there are no reliable external sources for the population of business fleets in California. All data sources we could find suffered from obvious undercoverage errors, particularly for smaller fleets. We therefore had to generate our fleet survey sample, as well as our estimates of the fleet population, directly from a "snapshot" of the California Department of Motor Vehicles (DMV) Vehicle Registration file as of June 1, 1993.

Unfortunately, the DMV registration file is very large (approximately 6 Gigabytes) and contains numerous errors and omissions. This greatly complicated the task of identifying business fleets, plus the numerous errors in the name and address fields made it difficult to contact those fleets which we were able to identify. This chapter describes the procedures we used to identify a sample population for our fleet survey (discussed in Chapter 11), and it also describes the methods we used to estimate the population of business fleet vehicles. This estimated fleet population was then used to estimate forecasting (or expansion) weights for our fleet sample.

10.1 Fleet Sample Construction

The fleet sample was constructed from the 1993 DMV Vehicle Registration file by sorting the file to collect all vehicles registered at the same name and address, and then choosing all cases in the file with more than 10 vehicles that appeared to be legitimate business fleets. This task is almost impossible to complete exactly because there are no validity checks imposed on the DMV file itself. For example, SCE fleet vehicles are registered to "SCE Corp.", "Southern California Edison", "So. Cal. Edison", etc. There is also no uniform way of entering or abbreviating addresses, and frequently address fields reserved for leasing companies are incorrectly used to record the registered owner's address.

We began by first deleting all records registered to addresses in San Diego County. Although it is possible that some of these vehicles are operated within our sample area, we decided after pretesting from San Diego County records that it was too costly to identify these vehicles. After deleting the San Diego records, we then cleaned the file to standardize common abbreviations. For example, all abbreviations for "Street" were changed to "ST," all occurrences of post office boxes were changed to "BOX," all "First St." were changed to"1ST", and all occurrences of "&" were changed to "AND." In order to minimize matching errors due to misspellings, we then generated a sort key by removing all embedded punctuation and vowels. For example, "AT&T", "A.T.&T.", and "A T T" all are recoded to "ATT," while "American Products and Services" is recoded to "AMRCNPRDCTSSRVCS."

After matching as many records as possible, we then need to identify fleets in our target population of fleets with at least 10 light-duty vehicles operated by businesses and local governments. One unforeseen problem was that there were a large number of fleets with less than 30 vehicles owned by individuals with Hispanic surnames. Since many of these had addresses that seemed to be residences, we decided to try calling a sample to see if they were business fleets. We tried calling 500 such fleets, and only two respondents were willing to respond to our questions. It appears that these "fleets" are comprised of personal vehicles belonging to illegal aliens but registered to legal family members of friends. Based on this experience, we deleted these "Hispanic" fleets from our sample.

The final fleet list was constructed by removing all vehicles registered by car dealers, leasing companies, and car rental agencies. These fleets were identified by searching for keywords such as "FORD", "HERTZ RENT", and "LEASING" in the name fields. The resulting list of candidate fleets, comprising approximately four percent of the vehicles in the original DMV file excluding San Diego, was then used to draw the sample for the fleet behavior survey. To insure that enough large fleets were contacted, the survey sample was drawn proportional to the fleet size.

10.2 Fleet Population Estimation

Due to the problems with the DMV data described in the previous section, we were concerned that our sample might not accurately reflect the actual population of business fleets in California. Since the size of the DMV file precluded more careful analysis, we took a stratified random sample of 1,863,000 vehicles for more careful analysis. This sample was drawn by first sampling zip codes proportional to the number of vehicles registered in the zip code, and then taking all vehicles registered in those zip codes. This procedure insures that fleets (at least those registered to the same zipcode) are not fragmented by the sampling procedure, and it also yields a self-weighting sample since each vehicle in the DMV sample has an equal probability of being chosen.

We then followed the same cleaning and classification procedures described in the previous section. Since we were working with a smaller sample, we were able to manually check fleets that were not clearly classified by the matching programs. In particular, we looked at the distribution of class and vintage for all such fleets with 10 or more vehicles. If these fleets contained a large number of pickups or another distinctive vehicle type, they were classified as business fleets. If they had a range of types and classes, which was typical for the Hispanic fleets described in the previous section, they were classified as residential vehicles. There were also a number of fleets consisting solely of very old vehicles; we assume that these belong to collectors and were therefore deleted from the sample.

Due to the extensive analysis of this sample, we are confident that this is the best classification possible from the DMV records. Based on this sample, we are able to make inferences about the population of business and rental fleets in California. This population is described in the next section, and this population is also used to reweight our fleet survey sample to make it more representative. This weighting procedure is described in the last section of this chapter.

10.3 Descriptive Statistics for Fleet Population

In our sample of 1,863,000 vehicles described in the previous section, we found 1,460 commercial fleets (with at least 10 eligible vehicles) containing 84,839 vehicles. This implies that 4.5% of the vehicles in California are in commercial fleets. Since there are about 12,140,000 vehicles in the Southern California area, the total number of commercial vehicles in Southern California is about 551,453. The following charts show the distributions of vehicle type and vintage for these commercial fleet vehicles.

As we can see from Chart.10.3.1, the most common vehicles in commercial fleets are Full Size Pickup Trucks and Mid-size Cars.

CHART.10.3.1 Distribution of Known Types of Commercial Fleet Vehicles at Sites with 10+ Vehicles in 1993

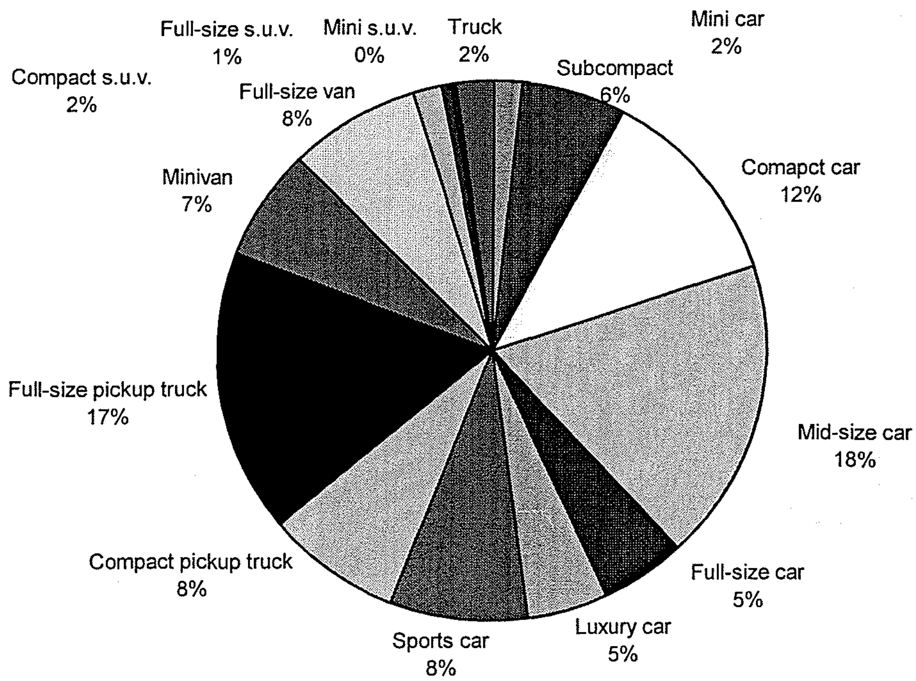


CHART.10.3.2 Model Year Distribution of Fleet Business Cars (93 DMV Data)

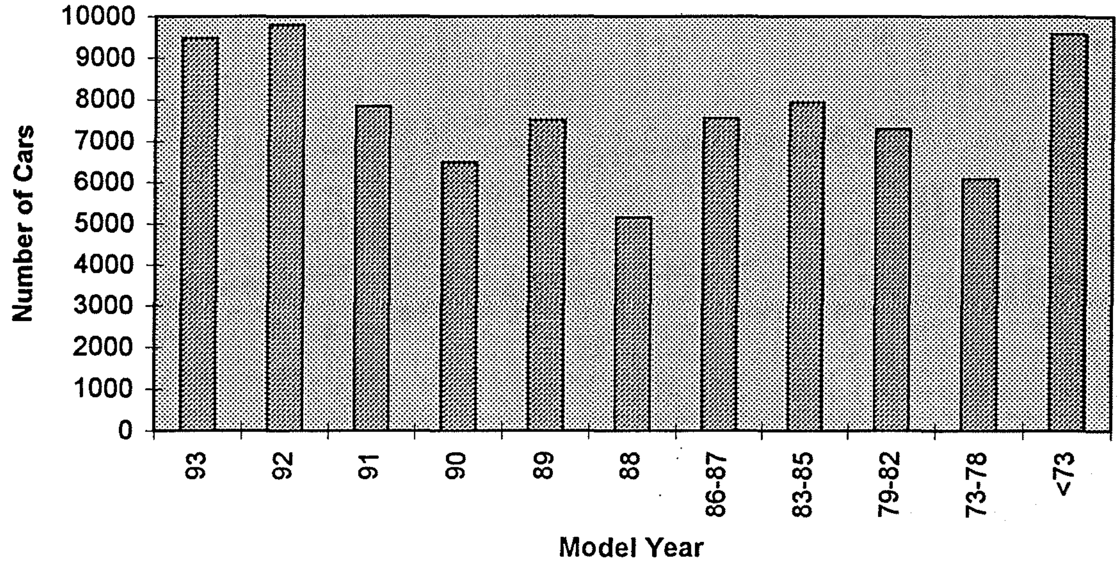


CHART.10.3.3 Joint Distribution of Type and Vintage for Fleet Vehicles at Sites with 10+ Vehicles

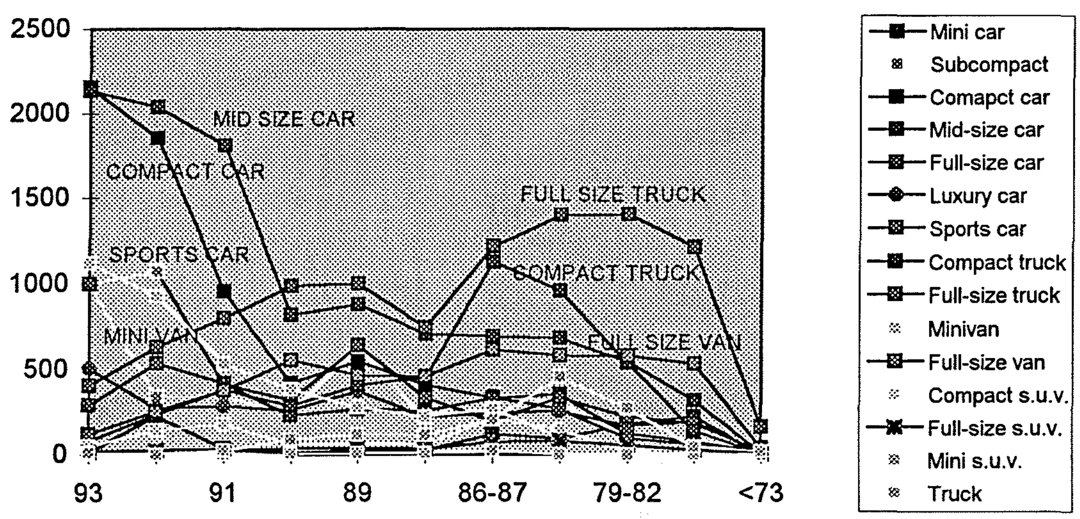
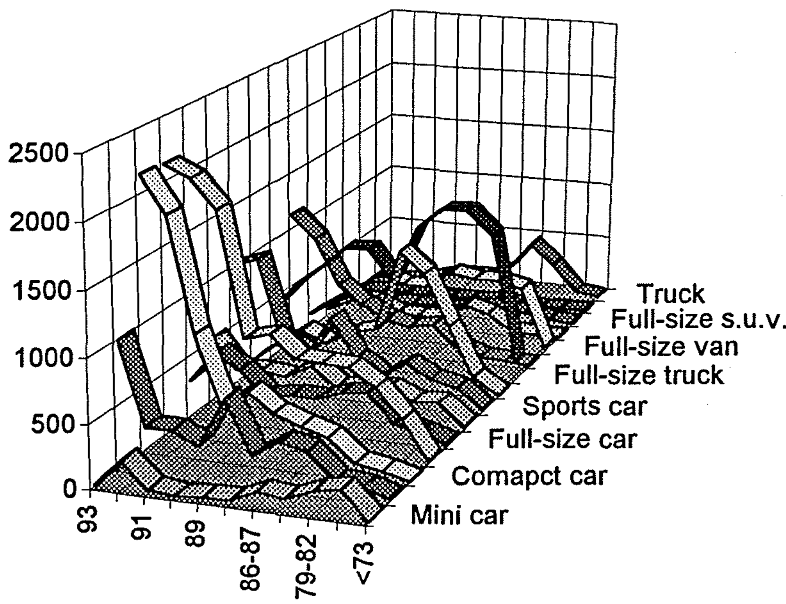


CHART.10.3.4 Joint Distribution Type and Vintage for Fleet Vehicles at Sites with 10+ Vehicles



In our sample of 1,863,000 vehicles, we also found 106,786 (5.7%) vehicles in rental fleets at sites with 10+ vehicles. This implies that there are about 694,109 fleet rental vehicles in Southern California. Chart.10.3.5 shows that the most common vehicles in rental fleets are Mid-size and Compact cars.

CHART.10.3.5 DISTRIBUTION of KNOWN TYPES of RENTAL FLEET VEHICLES at SITES with 10+ VEHICLES in 1993

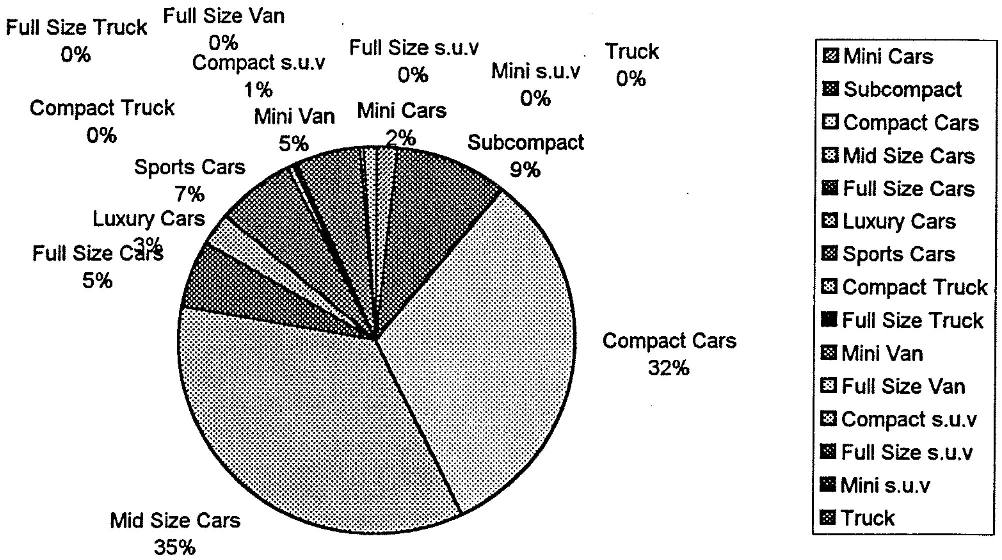


CHART.10.3.6 Model Year Distribution of Fleet Rental Vehicles at SITES with 10+ VEHICLES (93 DMV Data)

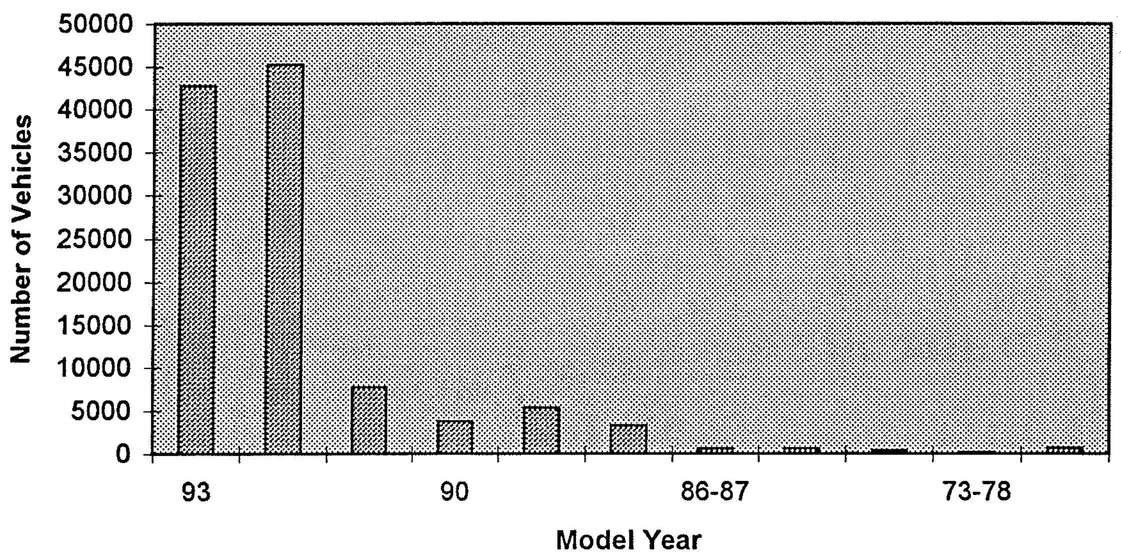


CHART.10.3.7 Joint Distribution of Type and Vintage for Rental Fleet Vehicles at Sites with 10+ Vehicles

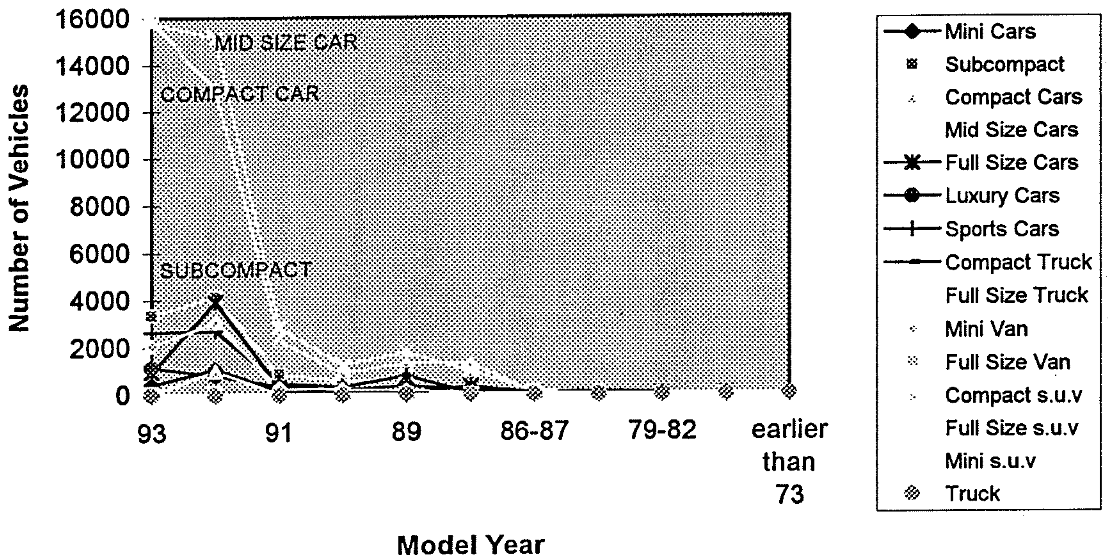
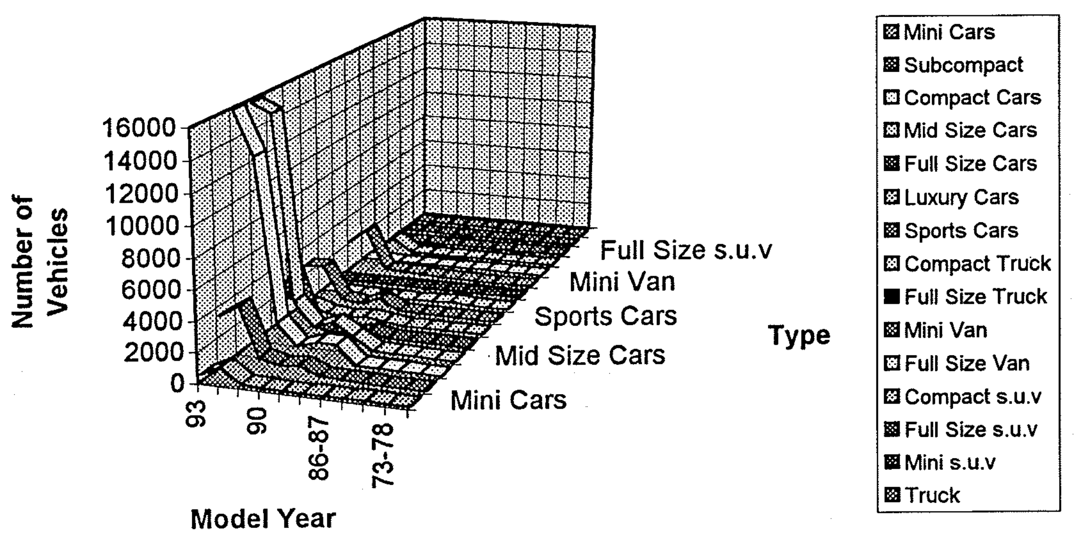


CHART.10.3.8 Joint Distribution of Type and Vintage for Fleet Rental Vehicles at Sites with 10+ Vehicles



10.4 Weight Construction for Fleet Sample

There are a number of reasons why our fleet sample might not be representative of the population described in the previous section. We initially drew our sample proportional to fleet size, so that we expected to oversample large fleets. A larger problem is that the DMV records only contain company names and registration addresses. In many cases it was impossible to find telephone numbers corresponding to the DMV records, so that we were unable to contact many firms in our initial sample. To make our final fleet sample more representative, we constructed sample weights (or expansion factors) as described in this section.

Ideally, we would like to proceed as in Chapter 5 and construct weights to match the joint distribution of fleet size, vehicle type, and vintage constructed as described in Section 10.2. This requires matching each fleet in our sample back to the DMV records to determine the vintage and vehicle type of each vehicle in our sample. For the reasons described in Section 10.1, this has turned out to be a very difficult task which we do not expect to complete before the end of 1995. The weights used in this version of the fleet model are constructed to match the marginal distributions of fleet size, vehicle type, and vintage.

Table 10.4.1 shows the expansion factors for each fleet size class. The larger factor for the smallest fleets reflects the deliberate undersampling of smaller fleets and the larger problems with finding valid telephone numbers for these fleets. The last columns show that the average fleet sizes in each group are similar for the sample and the SCE data. The total unexpanded sample average is higher (65 vs. 46 vehicles) because of the lower proportion of small fleets.

Table 10.4.1: Expansion by Fleet Size Class

Fleet Size	Sample Fleets	Target Fleets	Expans. Factor	Average Sample	Fleet Size vs SCE
10-19	744	5759	7.74	14	15
20-59	777	2497	3.21	31	38
60-119	238	561	2.36	78	85
120-499	192	408	2.13	212	239
>500	36	118	3.26	995	1150
Total	1987	9343	4.70	65	46

After the fleets are expanded by the appropriate size class factor, an adjustment is made to get the correct number of vehicles in each body class. For the six body classes in the fleet survey (shuttle buses are grouped with full size vans), the target number was determined from the distribution in the DMV data for years 85-92 applied to the total number of vehicles. We only used the years 85-92 to avoid problems with vehicle reclassifications. Since the fleet survey did not ask detailed vehicle type, we need to group the classes to correspond to the survey data. Therefore classes 1-7 were all grouped as class 3 - car and station wagon, and classes 12-14 (utility vehicles) were split across classes 8-12 and 15 in proportion to the number of vehicles already in those classes.

The results in Table 10.4.2 below show a further adjustment factor in the range of 1.5 to 2.2 for all classes except medium duty trucks, which has a much lower factor of .76. Respondents probably included a wider range of trucks in their estimates than what we had intended. In general, the factors are greater than 1 because the sample only includes a maximum of 2 body types per fleet, while many fleets have more than that.

Table 10.4.2: Adjustment by Vehicle Class

Vehicle Class	Sample Records	Sample Vehicles	Target SCE	Adjust. Factor	Adj. veh/Record
3-car/stat wagon	799	84159	173755	2.06	217
8-car/wag pickup	542	20622	36689	1.78	68
9-full size pickup	992	57320	120970	2.11	122
10-minivan	283	9925	15022	1.51	53
11-full size van	559	29143	64245	2.20	115
15-mod. duty truck	568	28846	21923	.76	39
Total	3743	230015	432605	1.55	116

In total there are 3743 fleet/body type records in the sample, with each record representing an average of 116 vehicles after expansion and adjustment. Before these records can be used in forecasting, some more information is necessary.

(1) The average VMT per year: This is taken directly from the record, and estimated using a regression equation if no response was given.

(2) The current distribution across fuel types: This was also taken directly from the survey record, though the split of alternative fuel and electric vehicles across body types had to be assumed in a small number of cases. There were very few non-gasoline vehicles in the data.

(3) The current distribution across vintages: We used the data directly from the survey records as to when vehicles of each type are replaced. Of the 3743 records, 2957 (80%) had a valid response, either directly in terms of the age at replacement, or indirectly in terms of the VMT at replacement divided by a valid response for average VMT (the regression -based VMT estimates were not allowed in this case.) Valid responses were assumed to be those between 0 and 20 years. In cases where both types of responses were given, an average of the two was used. For missing cases, a regression-based estimate of replacement age did not seem appropriate, since we are trying to estimate a distribution and the regression-to-the-mean problem would change the shape of the vintage distribution. As an alternative, the distribution across replacement ages was tabulated for each body type/fleet type combination, and a random draw from the appropriate distribution was made to fill in each case of missing data. The resulting numbers of vehicles in each cell and the average age of replacement within that cell are shown in Table 10.4.3 below.

Once we know the replacement age N for a given record, we can assume that (a) all vehicles are replaced every N years or else (b) that 1/N of the vehicles are replaced each year. For a single fleet, something in between is probably most realistic. Since each record in the sample represents a number of actual fleets, however, assumption (b) seems most realistic. So, we assume that initially that 1/N of the vehicles fall in each of the N most recent vintage years, and that no vehicles are older than N. When added across the sample, this should give an accurate total vintage distribution. (Note that replacement age is assumed to be independent of fuel type - the same rule is applied to all fuel types.)

Table 10.4.3: Average Age at Replacement by Fleet Type and Body Type

Vehicle Class:	Total	Car/Wag	Co. Pick	FS Pick	Mini Van	FS Van	MD Truck
Agriculture	8.45	5.17	5.92	8.31	12.44	12.50	10.30
Construction	9.00	6.27	9.51	8.88	9.33	10.46	9.06
Manufacture	6.41	5.30	6.11	7.93	5.87	6.21	9.27
TransComm	5.11	5.67	5.30	9.27	9.23	9.74	7.28
Retail/Whole	7.15	4.86	7.48	7.97	6.39	7.68	11.11
Auto Indust	5.83	4.22	7.01	7.30	4.60	8.96	5.82
Ranking Ins	4.31	4.25	4.62	4.00	5.80	4.80	8.22
Service Ind	7.11	5.37	7.59	7.95	6.60	8.95	8.25
City/County	7.68	7.24	8.21	8.69	7.74	7.67	9.51
Local Service	8.82	8.59	9.00	8.66	6.24	5.56	9.60
Schools	11.10	5.76	13.92	12.91	9.77	12.4	14.77
Others	5.53	4.99	6.15	7.02	3.29	5.85	3.00
Total	7.68	6.19	8.32	8.83	6.51	9.07	8.91

Preliminary analysis of more sophisticated weighting schemes suggests that there will only be minor changes from the results using the weights described in this section.

11. FLEET VEHICLE ALLOCATION CHOICE

11.1 Objectives And Research Context

The potential demand for alternative-fuel vehicles (AFVs) operating on electricity, compressed natural gas, methanol, or other "clean" fuels, can be divided into residential (or personal-use) demand and fleet demand. Although our preliminary results indicate that fleets with ten or more vehicles comprise only about 5 percent of the vehicle stock, they may still be an important source of demand for AFVs. First, there are incentives and mandates emanating from United States clean-air and fuel-management legislation (US DOE, 1994) that are intended as direct stimulants of fleet demand. Second, manufacturers are likely to make financial concessions to fleets in order to meet low-emissions vehicle sales quotas mandated in California by the California Air Resources Board. Third, the on-site refueling capabilities and mechanical expertise available at many fleet sites are key factors in the adoption of the new technology. Finally, competitive fuel prices make certain types of AFVs cost-effective for certain types of fleet operations.

Although it is widely recognized that fleet demand is important to the growth of alternative-fuel vehicle technology, survey data suitable for developing fleet demand models have been generally unavailable before 1994 due to the difficulty of establishing a representative sample of both business and government organizations with fleet operations. The current study provides results from a large, broad-based sample of fleet sites in California.

The paper is organized as follows: Previous research is discussed in Section 2, followed by a description of the survey in Section 3. Fleet site characteristics are explored in Section 4. Vehicle utilization is analyzed in Section 5. Fleet operator awareness of clean fuel mandates and their near-term AFV purchase intentions are examined in Section 6. A model of vehicle choice is presented in Section 7 that provides insight into the attribute tradeoffs that fleet managers are likely to exhibit when making future vehicle acquisitions in the presence of AFVs. Finally, the conclusions drawn to date are reported in Section 8.

11.2 Previous Research

Research encouraged by the oil crises of the 1970's focused on the ability of fleets to use low-range alternative-fuel vehicles, and on how fleet managers might make trade-offs among factors such as mileage and operating cost (e.g., Berg, *et al.*, 1984, and Hill, 1987). However, a new set of research priorities emerged in the late 1980's. The introduction of the US Clean Air Act Amendments (US EPA, 1990) and the consideration of regional mandates in California (California Air Resources Board, 1992) created a need to enumerate the number of fleet vehicles and how they were used at sites (e.g. Wachs and Levine, 1985). Although the US Census Bureau details truck inventories every five years, their report falls far short of providing the information needed for policy planning (Census, 1990).

Since 1990, there has been a growing number of marketing-based studies commissioned by fuel suppliers such as electric and natural gas utilities, equipment manufacturers, and others in emerging AFV industries (e.g. Runzheimer, 1993; Macro, 1994). In addition, there is a small, but increasingly valuable body of findings accumulating from alternative-fuel vehicle fleet trials (e.g. Batelle, 1994).

It is expected that fleets might have a greater capacity than households to re-assign vehicles to different routes and drivers, thus better accommodating limited range alternative-fuel vehicles. On the other hand, issues of safety, insurance cost, and risk associated with a new technology may be more salient to fleets because of their corporate liability.

Several findings consistently emerge from previous descriptive studies of fleet demand for alternative-fuel vehicles (AFVs):

(1) Operating characteristics are critical. Fleet purchase decisions are based on objective criteria such as direct cost, reliability, and job suitability (Berg, 1985, Miaou et al., 1992), and AFV demand is related to the availability of on-site refueling and the operational uses of the vehicles. Certain vehicle classes, such as vans or pick-up trucks, are more likely candidates for alternative fuels because of their refueling patterns and lower annual mileage. (Berg *et al.*, 1984)

(2) Larger fleets are more likely to innovate. Vehicles can be more readily re-assigned among tasks, and on-site refueling and service is more likely to be available. There is also evidence that large firms reach decisions differently and might be more willing to experiment and "risk" new automotive technology.

(3) Government and public utility fleets are more predisposed than commercial fleets to adopt alternative fuel vehicles. Since many studies pre-date important new regulations on AFV adoption, it is not known whether such fleets are reacting to mandates or to other factors, but mandates are likely to become of increasing importance to all fleets (Easton, 1991).

11.3 Survey Method

The survey sample was obtained from vehicle registration data for the State of California. Rule-based algorithms were developed to exclude households with large numbers of registered vehicles, and to identify slight differences in registration names and addresses as likely fragments of the same fleet site. The final sample was based upon a proportionate sample of vehicles registered to sites with 10 or more registrations. It excluded fleets registered to state and federal government agencies, rental fleets, and fleets composed only of large trucks (>14,000 lb. Gross Vehicle Weight (GVW)).

A two-part survey instrument was administered to fleet operators between February and June, 1994. The response to an initial CATI (Computer Assisted Telephone Interview) was 71%, once an eligible fleet manager could be identified. Information from this survey was used to customize a mail survey, which had an effective response rate of 78%.

The customized follow-up mail questionnaire was composed of three main parts (Golob, *et al.,* 1995):

> (1) Survey data were collected on the basis of seven different <u>vehicle classes</u> defined by body type and size (cars, minivans, full-size vans, compact pickups, full-size pickups, small buses, and medium-duty trucks with less than 14,000 lb. gross vehicle weight). Detailed questions were asked about vehicle acquisitions and operations for the vehicle class with the most vehicles at the site, and for a second vehicle class, which was assigned at random from a list of the other remaining vehicle classes operated there (if any). Information was therefore collected for a maximum of two vehicle classes: this restriction was used in order to reduce the survey length and minimize non-response. Questions included the number of vehicles and their average annual vehicle miles traveled (VMT) by usage category, how they are maintained, and the manner in which the vehicles are disposed of and replaced.
>
> (2) A stated choice task (a type of conjoint analysis) was presented for each of the vehicle classes identified in part (1). In each task, the respondent was asked to allocate future fleet acquisitions from a set of three hypothetical future vehicles defined according to an experimental design. The experimental design manipulated vehicle fuel type (gasoline, electric, compressed natural gas, and methanol), vehicle capital cost, operating costs, range between refueling, refueling times, fuel availability, cargo capacity, and emission levels.
>
> (3) Finally, we gathered extensive information on attitudes, intentions, and fleet decision making parameters. The attitudinal questions involved importance scales for a series of AFV acquisition criteria, AFV purchase intention, and opinions about the reliability and safety of different fuel types. We also assessed knowledge and awareness of AFV mandates.

The final sample consists of 2,711 CATI and 2,131 mail surveys. Most analyses are based on 2,023 responses that exclude 108 sites that had less than 10 vehicles.

11.4 Fleet Site Characteristics

11.4.1 Fleet Sectors

It is not generally known how vehicles are distributed by various industries, because most available samples of business establishments do not generate a representative sample of those that operate fleets. With respect to industrial sectors, our sample was broken down based on a simplification of SIC codes. There were also fleets from governments and schools. In all, we divided the sample into 12 major fleet sector categories. The tables presented in this section are not weighted, so they may not reflect the underlying population of fleet operators.

TABLE 11.1: Sample Breakdown by Fleet Sector

Fleet Sector	Number of Fleet Sites	% of Total
Agriculture	94	4.6
Automotive Business or Service	66	3.3
Banking & Insurance	56	2.8
City & County Government	291	14.4
Construction & Contracting	263	13.0
Household Services and Trades	256	12.7
Manufacturing	230	11.4
Miscellaneous Industries	32	1.6
Retail & Wholesale Sales	133	6.6
Services for Business & Professional Orgs.	202	10.0
Schools (public & private)	195	9.6
Transportation & Communications	162	8.0
Unknown	*43*	*2.1*

Table 1 shows that city and county government agencies account for the largest proportion of fleet sites that were contacted (14.4%), but this may also reflect a greater willingness on the part of these fleet managers to participate in a University of California study. About 60% of the fleets in the sample were in five of the twelve sectors: government fleets (14.4%), construction and contracting (13.0%), household services and trades (12.7%), manufacturing (11.4%) and services for business (10.0%). The sample excludes rental company fleets and those of federal and state government agencies.

11.4.2 *Fleet Size*

There were approximately 136,000 vehicles represented in the sample, but their distribution across sites is highly skewed towards large organizations. While approximately 50% of the sample fleet sites had 25 vehicles or less, these sites account for only 13% of the total fleet vehicles. Half of the vehicles are in fleet sites of 200 vehicles or more.

11.4.3 *On-Site Refueling And Maintenance*

On-site refueling is a critical reason why fleets might adopt clean fuels in advance of households. Although 44% of the overall sample have on-site refueling facilities, the use of such facilities varies widely. Table 2 shows a breakdown of the sites within each organizational sector according to whether: (1) they currently have on-site refueling, (2) they do not have it now, but either had central refueling in the past or indicated that it was physically possible to have on-site refueling at their location, or (3) they indicated that it was *not* possible to have central refueling.

TABLE 11.2: On-Site Refueling Capability by Site Organization Type

Fleet Sector	On-site refueling capability (%)			
	has presently	not now/ feasible	not feasible	unknown
Agriculture	71	25	4	0
Automotive Business or Service	24	49	27	0
Banking & Insurance	14	11	66	9
City & County Government	76	20	4	0
Construction & Contracting	41	39	17	3
Household Services and Trades	20	40	34	6
Manufacturing	41	33	23	3
Miscellaneous Industries	28	38	28	6
Retail & Wholesale Sales	35	38	24	3
Services for Business & Professional Orgs.	25	32	40	4
Schools (public & private)	72	21	5	2
Transportation & Communications	42	27	29	3
Total sample	43.8	30.8	22.4	2.9

Fleets that use on-site refueling most frequently are those in agriculture (71%), city and county government (76%) and school (72%) sectors. Fleet sites with considerably less on-site refueling include those in the construction (41%), manufacturing (41%), and transportation/communication (42%) sectors. Fleet sectors that are least likely to have on-site refueling capability are banking and insurance, and business and household services and trades. Such sectors might have smaller fleets, they may be based in dense urban areas, the vehicles might be taken home by employees at night, or the vehicles might be driven in less predictable patterns.

11.4.4 On-Site Maintenance

In the absence of a developed AFV service infrastructure, fleets might have to rely on their on-site capabilities in the near term. However, this is dependent upon cost factors, the ability to train mechanics, and procedures for obtaining parts. Forty percent of the fleet sites in the sample had the capacity to service at least two different vehicle classes on-site, while 33% of the sites always contracted out for service. The remaining sites in the sample serviced only one of two vehicle classes on-site.

Table 3 lists the maintenance locations for a site's primary vehicle class and one other vehicle class they operate (if any). Fleet sites with small (shuttle) buses are most likely to perform maintenance for such vehicles on-site, while minivans are more likely to be serviced off-site. On-site maintenance is also more common for full-size pick-up trucks and medium duty trucks under 14,000 gross vehicle weight.

TABLE 11. 3: Maintenance Locations by Vehicle Class

Vehicle class	Total fleet sites	Primary maintenance location (%)		
		On-site or at another co. location	Contracted to outside garage/lessor	Other or unknown
Cars	823	42.9	44.2	8.7
Minivans	310	33.6	47.1	19.3
Full-size Vans	523	43.6	44.4	8.8
Compact Pickups	560	45.5	40.2	14.3
Full-size Pickups	1019	53.9	32.2	13.9
Small Buses	69	63.8	20.3	16.0
Trucks <14,000 lb. GVW	587	52.8	33.6	13.6

11.5 VEHICLE UTILIZATION

Vehicle miles of travel (VMT) and other components of fleet vehicle duty cycles are commonly regarded as the most critical component of AFV feasibility. However, aggregate measures of VMT are problematic because averages typically must be computed across the combination of multiple types of vehicles and multiple vehicle functions within a particular fleet. Thus, a decomposition of VMT by vehicle class and function, controlling for fleet site characteristics, is a useful means of assessing vehicle usage requirements.

Table 4 provides a breakdown of average annual VMT by fleet sector. Fleet sites in the transportation and communication sector record the highest VMT (approximately 36,000 miles per year per vehicle), followed by sites in the automotive sector, business services sector, and retail and wholesale trade sector. Schools record the lowest VMT (14,000 miles).

TABLE 11. 4: Average Annual Vehicle Miles Traveled for All Purposes by Site Organization Type

Fleet Sector	Average Annual VMT
Agriculture	22,300
Automotive Business or Service	28,300
Banking & Insurance	18,400
City & County Government	16,500
Construction & Contracting	24,500
Household Services and Trades	22,300
Manufacturing	23,700
Miscellaneous Industries	16,700
Retail & Wholesale Sales	27,900
Services for Business & Professional Orgs.	28,000
Schools (public & private)	14,000
Transportation & Communications	36,000

A regression model was computed to explain annual average VMT simultaneously as a function of: (1) utilization category, (2) vehicle class, and (3) fleet site characteristics. The regression results are listed in Table 5. The dependent variable is scaled in units of 1,000 miles. The R^2 for the regression is 0.094. The constant of 16,420 miles provides a baseline VMT from which comparisons can be made. VMT varies widely by industry sector, with the lowest VMT reported by schools. Average VMT for government agencies is not significantly different from the constant, while VMT for the remaining sectors are all greater than this constant.

VMT is negatively associated with very large site-size (sites with 500 or more vehicles), supporting the contention that large organizations are better able to rotate their vehicles, or allocate them across multiple drivers. Organizations that have 20 or more fleet sites are also less likely to have a higher VMT per vehicle. However, there is a very large and significant coefficient for the variable which measures how prevalent the primary vehicle class is relative to all other vehicle classes at the site.

Fleet sites that have a limited number of vehicle classes exhibit more extensive vehicle usage, compared to fleets that have a wider variety of vehicle types. It is likely that organizations with a single vehicle-type have a more specialized function (e.g. courier services). These regression results confirm that small buses log considerably more miles than other vehicles types, as do vehicles used in courier services, sales-calls, and transportation of people. Significant terms were found for the interactions of vehicle class and utilization category; these are potentially important fleet market segmentation variables.

11.6 AFV Mandate AWareness AND near-term purchase intention

Respondents were asked whether or not they believed that their site was subject to regulations requiring the use of alternative-fuel vehicles. Overall, 28% believed that there was legislation requiring their organization to use alternative fuel vehicles. By sector, 50% of the local and county governments perceived regulation, while only 23.3% of the commercial fleet managers perceived that their site was regulated.

A binomial probit model (Maddala, 1983) was estimated to explain differences in awareness of AFV regulation as a function of fleet site characteristics, the dependent variable being coded as: 0 = not aware, 1 = aware. The coefficient estimates for this probit model are listed in Table 6. The pseudo-R^2 value was 0.23. City and county government fleets were more likely to perceive that their site is subject to AFV mandates. Manufacturing organizations and schools were also more likely to perceive regulation, and other important predictors of awareness were the presence of on-site refueling, and the size of the fleet.

The propensity to purchase a clean fuel vehicle <u>within the next two years</u> was measured in the survey on a five-point scale, where the mid-point choice was "somewhat likely". The specific wording was: "What is the likelihood that one or more alternative fuel vehicles will be purchased for this location within the next two years?" Reliability analysis based on comparing results with a similar question asked in the follow-up mail survey eliminated 125 respondents.

An appropriate regression method for determining differences among fleet sites in terms of stated AFV purchase intentions is the ordered-response probit model (also known as the "ordered probit model"), developed by Aitchison and Silvey (1957) and Ashford, (1959). The ordered-response probit model respects the dependent variable as an ordinal scale, not requiring the tenuous assumption of equal intervals between the semantic scale points (Maddala, 1983). Results are listed in Table 7. The pseudo R^2 value was 0.22.

TABLE 11.5: Regression of Average Annual VMT as a Function of Vehicle Utilization Category, Vehicle Class, and Site Characteristics

Explanatory Variable	Coefficient	*t*-statistic
Constant	16.42	10.6
Fleet sector dummies *(base: city and county government)*		
Agriculture	5.89	3.2
Automotive Business or Service	7.97	3.6
Banking & Insurance	-2.57	-1.0
Construction and Contracting	5.72	4.3
Household Services and Trades	4.71	3.4
Manufacturing	2.54	1.8
Retail & Wholesale Sales	6.75	3.2
Services for Business & Professional Orgs.	4.50	3.1
Schools	-3.36	-2.4
Other fleet site characteristics *(base: 20 - 119)*		
Site Size 10-19 (dummy)	-2.57	-3.2
Site Size 120-499 (dummy)	-2.87	-2.4
Site Size 500 or more (dummy)	-4.18	-1.8
Site is Organization's Only Site in CA (dummy)	-1.41	-1.8
Organization has 20 or More Sites in CA (dummy)	-5.74	-2.6
On-site Refueling Present (dummy)	-1.58	-2.0
Vehicle class dummies *(base: compact pickups)*		
Cars	2.00	1.5
Minivans	2.75	1.8
Full-size Pickups	3.01	2.5
Small Buses	12.21	4.0
Trucks <14,000 lb. GVW	3.76	2.4
Fraction of fleet that is the primary vehicle class	12.50	8.9
Utilization category dummies *(base: "other" uses)*		
Courier	16.23	4.6
Pickup/Delivery	4.68	3.2
Haul Equipment	-1.98	-1.6
Service/Maintenance	0.056	0.0
Sales Calls	10.56	4.5
Transport People	14.45	9.0
Employee Use	0.335	0.2
Utilization X type interaction dummies		
Full-size Pickup X Service/Maintenance	-3.60	-2.0
Car X Employee Use	-4.23	-1.8
Truck < 14,000 lb. GVW X Pickup/Delivery	-4.55	-1.7
Car X Sales Calls	-6.49	-2.2
Truck < 14,000 lb. GVW X Service/Maintenance	-6.98	-2.7

TABLE 11.6: Binomial Probit Model of Belief that Site is Subject to AFV Mandates
Base categories are agriculture and site size 20-29

Explanatory Variable	Coefficient	*t*-statistic
Automotive Business or Service	-0.050	-2.40
City & County Government	0.131	5.14
Construction & Contracting	0.024	1.00
Household Services and Trades	-0.072	-3.06
Manufacturing	0.059	2.52
Retail & Wholesale Sales	-0.050	-2.21
Schools	0.089	3.77
Site size 10-19	-0.139	-5.99
Site size 15-19	-0.121	-5.10
Site size 30-59	0.033	1.34
Site size 60-119	0.131	5.44
Site size 120 or more	0.187	7.39
Organization Has More Than One Site in CA	0.047	2.23
On-site Refueling Present (dummy)	0.129	5.46

TABLE 11.7: Ordered-Response Probit Model of Stated Intention
to Purchase Alternative-Fuel Vehicles
Base categories are agriculture and site size 20-29

Explanatory Variable	Coefficient	*t*-statistic
Automotive Business or Service	-0.035	-1.58
City & County Government	0.180	7.03
Construction & Contracting	-0.090	-3.74
Household Services and Trades	-0.067	-2.82
Manufacturing	-0.061	-2.60
Retail & Wholesale Sales	-0.032	-1.42
Schools	-0.009	-0.39
Site size 10-19	-0.028	-1.17
Site size 15-19	-0.040	-1.67
Site size 30-59	-0.000	-0.01
Site size 60-119	0.079	3.26
Site size 120 or more	0.232	9.09
Site is Organization's Only Site in CA	-0.041	-1.91
On-site refueling present (dummy)	0.121	5.10

Larger fleets are more likely to intend to make an AFV acquisition, even when differences in decision making styles and awareness of AFV mandates are taken into account. It is likely that size is a proxy for several factors (Golob, *et al.*, 1995): First, larger firms have greater ability to absorb risk and liabilities associated with a new vehicle. Second, at an operational level, they find it easier to rotate drivers and vehicle assignments in order to accommodate limited range vehicles. Finally, larger firms might be more attracted to the potentially favorable publicity and image associated with use of clean fuels.

Fleet sector (Table 1) is another effective predictor of near-term AFV interest. City and county government is the only sector that is positively inclined to acquire AFVs. All other sectors, including manufacturing and construction, display a tendency of not intending to acquire AFVs. Fleet operators at sites in these sectors may perceive that current AFVs will not meet their duty-cycle needs, such as heavy delivery and hauling. On-site refueling is also a significant predictor of purchase-intention. Firms that have on-site refueling view it as more practical and feasible to operate alternative-fuel vehicles, given that most alternative fuels are currently not readily available at public service stations.

The model results for near-term AFV purchase intention parallel the results from the probit model of perceived awareness of AFV mandates (Table 6), except for schools and manufacturing sectors. Government sites appear more likely to acquire alternative-fuel vehicles, as do sites with on-site refueling, and sites with larger fleets. School fleets and manufacturers were aware of the mandates, but the purchase-intention model indicates that these fleets are unlikely to acquire AFVs in the near-term.

11.7 THE STATED PREFERENCE VEHICLE CHOICE MODEL

11.7.1 Methodology

The mail-out portion of the survey was sent to the person who was identified in the initial CATI contact as responsible for acquisition of the vehicles at the sampled fleet site. In most instances this was the same respondent that was interviewed by phone, but for some sites a different person was identified. For instance, the vehicle acquisition manager might be at a different location, e.g., the company headquarters. Complicated contact protocols were followed to establish identities and to make appropriate introductions.

Managers responsible for vehicle acquisition were asked to complete a maximum of two stated preference (SP) choice tasks for the vehicle classes discussed in section 3. In each task, they were asked to allocate their future fleet purchases for a given vehicle class (e.g., car, minivan, etc.) by using a set of three hypothetical future vehicles defined according to an experimental design. For each vehicle class, there were three different fuel type versions available (out of a total of four: CNG, gasoline, electric, methanol). The format of this task is similar to the survey instruments used in household stated choice tasks (Bunch, et al., 1993; Golob, et al., 1993), but the respondents in the Fleet Survey were allowed to choose varying numbers of vehicles to make up their entire fleet for each vehicle class. The experimental design manipulated the three available fuel types by randomly selecting three from the list of four types included in our study (i.e., gasoline, electric, compressed natural gas, and methanol). For respondents having the (maximum) two vehicle classes, the fuel types were chosen so as to be different for the two tasks. This approach ensures that a respondent would be exposed to all four fuel types. In addition, this has the feature of allowing estimation and testing of models that do not assume the independence from irrelevant alternatives (IIA) assumption with respect to vehicle fuel type.

An experimental design was also used to manipulate the values of the other "generic" and "fuel specific" vehicle attributes, including vehicle capital cost, operating costs, range between refueling, refueling times, and fuel availability. These design variables are listed in Table 8. The specific operating characteristics of the hypothetical vehicles varied from survey to survey according to an experimental design approach (Bunch, Louviere, and Anderson, 1994) that makes use of orthogonal fractional factorials for generating the first alternative in the choice set, followed by a shifting procedure to generate additional choice alternatives in the choice set. The page of one questionnaire containing a stated choice task is reproduced in Figure 1. This is an example of one of the sixty-four different experimental design treatments used: a wide variety of vehicle descriptions was seen by the survey respondents.

The indicated number of vehicles assigned to each fuel type by the respondent was converted to a fraction of the total number of vehicles for that vehicle body type and used as a weight in a maximum likelihood estimation procedure. A weight of zero was assigned to fuel types that were not picked at all by the respondents.

11.7.2 Choice Model Results: Generic Vehicle Attributes

The multinomial conditional logit model (Maddala, 1983) effectively explained vehicle allocation choices. In order to test the validity of the multinomial logit model (MNL), tests of the independence of disturbances across the fuel types (the model IIA properties) were conducted (Hausman and McFadden, 1984). The need to capture non-IIA properties with a nested multinomial logit or similar model was rejected in favor of the simpler MNL. This model fits the stated choice data well, with a log-likelihood (initial) = -5087.2, and a log-likelihood (model) = -4455.9 with 34 degrees of freedom and 2131 observations. This corresponds to a pseudo-R^2 of 0.12. The coefficients are listed in Table 9.

The coefficient for capital cost is statistically highly significant, and has the expected sign. The interaction terms involving capital cost and fleet sector dummy variables indicate that city and county government fleet sites are slightly less sensitive to the capital cost of the vehicles compared to most other sectors.

TABLE 11.8: Stated Choice Task Design Variables

Variable	Acronym
generic variables	
Capital cost of vehicle in $	capital cost
Vehicle range in miles	refueling range
Number of refueling stations relative to gas stations (gasoline = 1)	station density
Tailpipe emissions relative to new 1993 gasoline vehicles	emissions
electric - specific variables	
Operating cost with overnight recharging in cents/mile	EV off-peak cost
Operating cost with day-time recharging in cents/mile	EV peak cost
Number of vehicles with similar fuel type on California roads	EV penetration
Hybrid dummy (0 = battery only / 1 = with gas range extender)	EV hybrid
On-site recharging time in hours	EV on-site time
EV service station recharging time in minutes	EV station time
Cargo capacity compared to gasoline vehicles	EV cargo
compressed natural gas - specific variables	
Operating cost in cents/mile	NGV operating cost
Number of vehicles with similar fuel type on California roads	NGV penetration
Dual fuel dummy: (0 = NGV only; 1 = can also run on gasoline)	NGV dual fuel
Cost of installing NGV slow-fill refueling on-site in $	NGV slow-fill cost
Cost of installing NGV fast-fill refueling on-site in $	NGV fast-fill cost
On-site slow-fill refueling time in hours	NGV slow-fill time
On-site fast-fill refueling time in minutes	NGV fast-fill time
Service station refueling time in minutes	NGV station time
Home refueling unit installation cost in $	NGV home-fill cost
Cargo capacity compared to gasoline vehicles	NGV cargo
methanol - specific variables	
Operating cost in cents/mile	MV operating cost
Number of vehicles with similar fuel type on California roads	MV penetration
Cost of installing methanol refueling on-site in $	MV on-site cost

As expected, range was found to be an important vehicle attribute, and fleet sites where vehicles are used for transporting people have a significantly lower coefficient for range. The ratio of the range coefficient (.00219) to the capital cost coefficient (-.0000265) indicates that the "trade-off" between range and capital cost is approximately $80.00 per mile. In other words, for a given choice alternative, if the range is reduced by one mile, then the associated decrease in utility can be compensated for by reducing the capital cost by $80; such a trade-off will leave the choice probability unchanged. Adding 25 miles of range is equivalent to a $2,000.00 cost premium. Certain sectors, particularly government and manufacturing sites, have a much lower dollar value for range. Fleet sites with personnel transport functions have a higher dollar value for range.

The choice model is specified with one operating cost variable for gasoline vehicles, NGVs and MVs, and two operating cost variables for EVs: operating cost for off-peak (night-time) recharging and operating cost for peak (day-time) recharging. All the coefficients have the correct negative sign. The non-EV operating cost and capital cost coefficients imply that fleet acquisition managers are indifferent between a capital cost increase of approximately $2,200 for a reduction in operating cost of $.01 per mile. The coefficients for EV operating cost indicate that fleet managers are less sensitive to EV operating costs relative to operating costs for other fuels.

The availability of alternative fuel stations off-site was also important to fleet managers, indicating that fuel infrastructure should be an important element of policies aimed at encouraging the adoption of alternative-fuel vehicles. However, reduced tailpipe emissions was found to be a significant predictor of vehicle choice only for the government and school sectors. This indicates that fleet operators in other sectors may be guided by economic and other practical concerns, rather than purely environmental factors, in their vehicle selections. Perhaps local government agencies and schools are the equivalent of "green" consumer in the commercial sector.

11.7.3 Choice Model Results: Fuel-Specific Effects

Even after controlling for range, capital, and operating costs, fleet managers clearly prefer gasoline vehicles over alternative fuels. Gasoline was defined to be the base fuel, and the choice-specific constants for the other three fuels are negative. However, there are many significant interaction terms involving the fuel-choice-specific constants and fleet site characteristics, indicating that there are considerable differences in preferences by market segment.

Agricultural sites have strong aversions toward electric vehicles (EVs), as do sites operating trucks from 6,000 to 14,000 lb. gross vehicle weight. School fleet operators are less negative about EVs, possibly due to their more intense environmental concerns. This is consistent with the awareness among school fleet operators of AFV mandates (Table 6), and the sensitivity of their choice to tailpipe emissions.

Several fleet market segments find compressed natural gas vehicles (NGVs) to be just as attractive as gasoline vehicles, based on their fuel-specific choice constants. These segments include: Large fleets with at least 120 vehicles at the surveyed site, schools, and city and county governments. It is likely that firms with larger fleets have had more exposure to NGVs, are subject to various AFV regulations, and can potentially accommodate on-site refueling. Conversely, preference for NGVs is weakest among fleets in the banking, insurance and real estate sector, potentially because of a low incidence of on-site refueling fuel (Table 2) and relatively low vehicle usage levels (Table 4).

TABLE 11.9: Conditional Logit Model Of Vehicle Allocation Choice

Explanatory Variable	Coefficient	*t*-statistic
Capital cost and fleet sectors = agriculture, automotive business or services, banking & insurance, household services, retail & wholesale sales, services for business & professional orgs., schools, or transportation & communication	-.0000265	-4.78
capital cost and fleet sector = city and county gov.	-.0000235	-2.12
capital cost and fleet sector = construction	-.0000143	-1.31
capital cost and fleet sector = manufacturing	-.0000239	-1.88
Range and utilization category = all except transport people	0.00219	6.39
Range and utilization category = transport people	0.00152	2.77
Station density	0.213	2.27
Operating cost (NGV, methanol, gasoline)	-0.0583	-4.91
Emissions and fleet sector = city/county gov. or = school	-0.409	-2.70
NGV dual fuel	0.294	3.59
EV off-peak cost	-0.0129	-0.41
EV peak cost	-0.0162	-1.62
gasoline on-site refueling available	0.267	3.49
EV on-site refueling time in hours.	-0.0688	-1.66
EV station time	-0.00468	-1.57
NGV station time	-0.0253	-2.49
Cargo capacity (EV and NGV)	0.147	1.31
EV constant	-0.895	-2.51
EV constant and vehicle class = compact pick up	0.289	2.14
EV constant and utilization category = transport people	0.484	3.39
EV constant and vehicle class = trucks =<14,000# GVW	-0.395	-2.47
EV constant and utilization cat. = service/maintenance	0.349	3.23
EV constant and fleet sector = schools	0.769	4.16
EV constant and fleet sector = agriculture	-0.632	-1.82
NGV constant	-0.363	-2.43
NGV constant and fleet site size >=120 vehicles	0.424	3.04
NGV constant and fleet sector = city and county gov.	0.297	2.34
NGV constant and fleet sector = schools	0.439	2.71
NGV constant and fleet sector = retail and wholesale	-0.261	-1.49
NGV constant and fleet sector = banking, ins., real est.	-0.754	-1.95
MV constant	-0.261	-2.95
MV constant and fleet sector = schools	-0.297	-1.70
MV constant and fleet sector = transport. and comm.	-0.268	-1.65
MV constant and fleet sector = agriculture	0.342	1.84

The dual-fuel capability of operating NGVs on gasoline substantially increased their acceptability. The coefficient of the dual-fuel variable implies that fleet managers are indifferent between a $11,000 increase in capital cost and adding dual-fuel capability. However, fleet managers rated cargo space as important, so the reduction in cargo space to accommodate dual-fuel capability partially offsets the dual-fuel advantage. Finally, refueling time at a service station is also an important variable.

Methanol is the least unattractive of the non-gasoline fuels, as indicated by comparing the fuel-specific choice constants. Many fleet managers are familiar with methanol, and some methanol vehicles are available today. Because all methanol vehicles presented in the stated choice tasks were flexible-fuel, they can also operate on gasoline, which is clearly a preferable attribute. However, preference for methanol vehicles is significantly lower for both school and transportation and communication fleets. In the case of schools, this could reflect a common concern about safety. In contrast, the agricultural sector was more-predisposed towards methanol than other sectors. There are several plausible explanations for this, including similarities (and possible confusion) between methanol and ethanol, the ease of conversion between gasoline and methanol, and geographic differences in air quality.

11.8 Conclusions and directions for future research

This investigation has provided new information on preferences for electric and other alternative-fuel vehicles among a wide spectrum of fleet managers. These preferences should be important to governmental policy planners and vehicle manufacturers, because fleet demand is a critical component in US Federal clean air and energy legislation and California mandates for the electric and low emissions vehicles. The underlying survey used a complex contact protocol and multiple-stage interview process in order to interview both managers responsible for fleet operations and those responsible for vehicle acquisition decisions.

The descriptive analysis pinpointed vehicle utilization as a significant parameter. Although the average fleet annual vehicle miles of travel (VMT) across all fleet sites was 16,420 miles, there was substantial variation between fleet sectors, from a high of 36,000 miles by fleets in the transportation and communications sector, to a low of 14,000 miles by schools. VMT further varies by vehicle class, indicating that fleet operations are highly differentiated. This differentiation was apparent in fleet managers' awareness of alternative-fuel mandates and in their plans for near-term purchases of alternative-fuel vehicles.

The stated choice model results also showed that there were major differences in preferences among fleet market segments. For example, schools were less negative toward electric vehicles and compressed natural gas vehicles, but more negative toward methanol vehicles, relative to the other segments. There were substantial differences among fleet market segments in terms of attribute trade-offs.

The choice model results provide information on attribute trade-offs. For all fleets on average, the trade-off between range and capital cost is approximately $80.00 per mile. The availability of alternative fuel stations off-site was important to fleet managers, indicating that fleets are willing to trade-off costs for fuel infrastructure, or that such infrastructure can compensate for limited vehicle range. However, reduced tailpipe emissions were found to be a significant predictor of vehicle choice only for the government and school sectors. This indicates that private fleet operators' vehicle selections are not influenced by environmental factors.

The stated choice model provides a basis for forecasting fleets' demand for alternative-fuel vehicles. These forecasts require weights to expand the survey sample to the entire fleet population, and registration files of the California Department of Motor Vehicles are being used to develop these weights. Preliminary results show that for the six-county greater Los Angeles region there are approximately 10 million household vehicles, 430,000 fleet vehicles operated by the fleets of the type covered in this report. This suggests that the fleets covered in this paper will need to purchase a disproportionate number of alternative-fuel vehicles if they are to be important contributors to meeting clean-fuel mandates.

Once the vehicle registration files are processed, we can get more information about our sample fleets' current vehicle holdings. In particular we can get the make, model, and vintage of each vehicle in the fleet. This information can be used to more closely link the fleets' stated choices to their revealed preferences as evidenced by their past vehicle purchases. Eventually these data could be used to fit joint stated and revealed preference models similar to our household models (Brownstone et. al., 1994).

We plan on following all of the sample's fleet vehicles between two "snapshots" of the registration file taken one year apart. This will allow a better measure of the fleets' vehicle replacement policies. In particular, we will be able to see which fleets purchase new or used vehicles. This information is critical for forecasting the short-run dynamics of fleet purchase behavior.

FIGURE 11.1: Example of the Stated Preference Choice Allocation Survey Task Showing One of 64 Experimental Treatments

19. Assume that you must now replace your entire fleet of CARS AND STATION WAGONS by using the three types of CARS AND STATION WAGONS described in the table below.

CARS AND STATION WAGONS

Fuel Type	Gasoline	Electric	Natural Gas (CNG)
Dual Fuel Ability			Can also run on gasoline.
Capital Cost Per Vehicle	$17,000	$14,000 (includes recharge unit)	$16,000
Vehicle Range	250 miles	100 miles	275 miles on CNG
Operating Costs	6 cents per mile	4 cents per mile for overnight recharging. 12 cents per mile for daytime recharging.	4 cents per mile
On-Site Refueling Refueling Time	On-site refueling not available Not Applicable	Recharging unit comes with each vehicle for on-site use. 3 Hrs. for full charge	Not Applicable Not Applicable
Service Station Refueling Refueling Time	Gasoline available at current stations 7 min. to fill empty tank	5 recharge stations for every 10 gasoline stations 60 min. for full charge	1 CNG station for every 10 gasoline stations 5 min. to fill empty CNG tank
Home Refueling Refueling Time	Not Available.	Can recharge at home overnight.	CNG home refueling units cost $4,000 6 Hrs. to fill empty CNG tank
Tailpipe emissions	25% of new 1993 gasoline car emissions	Zero tailpipe emissions	40% of new gasoline car emissions

How would you replace your entire fleet of CARS AND STATION WAGONS from the three vehicle choices described in the preceeding table? Under each fuel type indicate the number of vehicles you would require for each use.

Replacement of CARS AND STATION WAGONS

VEHICLE USAGE	Gasoline	Electric	Natural Gas (CNG)
SALES OR CUSTOMER VISITS	______	______	______
SHUTTLE / RIDESHARE / COMMUTE	______	______	______
Other uses: ______	______	______	______
Total:	______	______	______

If you ruled out any vehicle type in the above table, please describe why: ____________________

12. FLEET VEHICLE AND FUEL DEMAND

12.1 Introduction

This chapter is laid out as follows:

Section 12.2 outlines the structure of the fleet vehicle forecasting module, the types of data inputs necessary, and the types of output produced.

Section 12.3 documents the fuel-type choice model for new fleet vehicles which is at the heart of the fleet vehicle forecasting module.

Section 12.4 describes the process for creating a representative sample of fleets and vehicles for forecasting, including expansion and specification of fleet characteristics.

Section 12.5 presents initial forecast results from the fleet module, and illustrates the types of outputs which are produced.

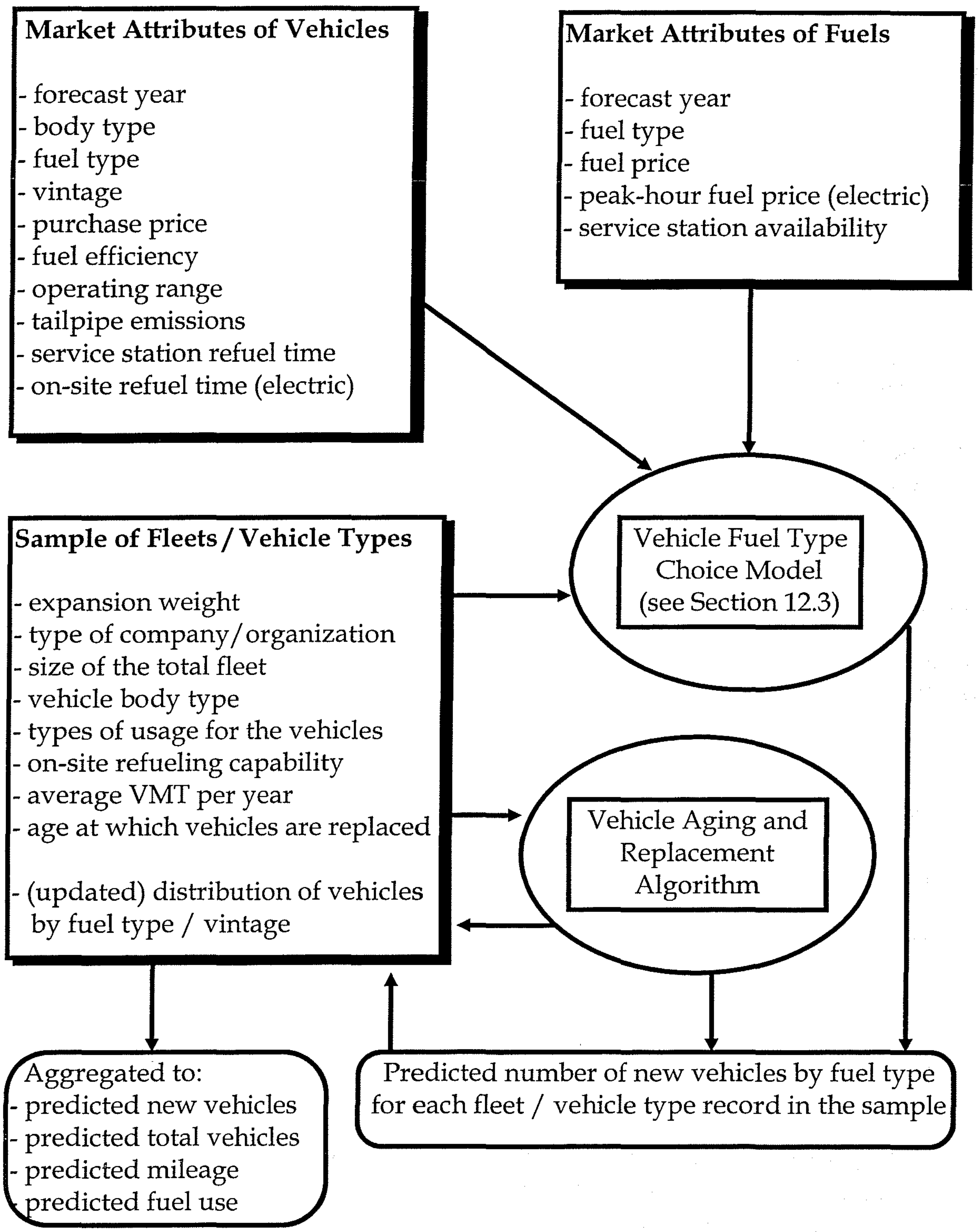
12.2 Structure of the Fleet Vehicle Forecasting Module

Figure 12.1 illustrates the main inputs, models and outputs of the fleet module. There are three main input files:

(1) The vehicle technology file.
This is the same file format as is used in the personal vehicle forecasting module-- see Table 3.3. Two extra vehicle classes - medium-duty trucks and shuttle buses - were added for potential use in the fleet model; however, our survey data only support forecasts for medium-duty trucks. The attributes in the vehicle technology file used by the fleet models are summarized in Figure 12.1. The file contains additional attributes, such as acceleration and top speed, which are used in the personal vehicle model but not in the fleet model. For a more detailed discussion of the contents of this file, see section 3.4.

(2) The fuelsforecast file.
Again, this is the same file format as is used in the personal vehicle module--see section 3.5 and Table 3.4. For each forecast year, the file contains the price and service station availability of each fuel: gasoline, methanol, CNG and electricity. A higher peak-period price is also used for electricity, since it is assumed that most emergency or "opportunity" recharging will occur during the day.

Figure 12.1 Schematic Diagram of the Fleet Vehicle Forecasting Module



(3) The sample of fleets and vehicle types.
The final input file has been derived from the 1994 Fleet Survey sample. Each record in the sample is a particular fleet/vehicle body type combination. In the survey, detailed data was collected for the main body type in each fleet, plus one other body type at random if the fleet actually contained more than a single type. Thus, a maximum of two records per actual fleet is contained in the sample. Each record contains the variables listed in Figure 12.1, including the type of organization and the size of the total fleet, and the on-site refueling capability, plus, for the specific vehicle type, the types of usage (sales, delivery, etc.), the average annual miles traveled per vehicle, and the age at which most vehicles are replaced. Finally, each record contains the base year (1994) joint distribution of vehicles of that body type by fuel type and vintage - from new up to 19+ years old.

Before the fleet sample is used in forecasting, it is expanded to represent the actual 1994 population of fleets and fleet vehicles in the study area. At the same time, any missing data are "filled in" for key characteristics such as average VMT and vehicle replacement age. These data expansion and completion procedures are described in Section 12.4.

Figure 12.1 shows the three input files described above feeding into the Vehicle Fuel Type Choice Model. This is a discrete choice model, based on the 1994 Fleet Survey stated preference experiment, which predicts the split of new fleet vehicle purchases across competing fuel types. This model is described further in Section 12.3.

The other main algorithm needed to complete the module is one which simulates vehicle aging and replacement from year to year. Aging and replacement rates in the model are a function of the current vintage distribution in each fleet and the age at which that fleet replaces older vehicles of a specific type with new ones. Thus, if a certain fleet replaces full size pickups at age 6, than any pickup trucks in that fleet which are less than 6 years old will become one year older during the simulation year, and any which are already six years old will be replaced by new vehicles - with fuel types predicted by the vehicle choice model. In this way, it may take a number of years for electric and alternative fuel vehicles to gain a significant predicted market share, as only a fraction of vehicles are replaced each year.

This representation of the fleet vehicle aging / replacement process contains a few important simplifying assumptions:

(a) that a given fleet replaces all vehicles of a certain type at the same age;
(b) that fleets purchase only new vehicles and no used vehicles; and
(c) that replacement rates are not influenced by the attributes of the vehicles and fuels on the market.

The first assumption is not a major issue, since the replacement rates across the expanded sample as a whole are representative of the actual fleet population. The second and third assumptions may not be entirely accurate, but they are necessary approximations given that data was not available to estimate a full transactions-based model structure as was done for household vehicles. In any case, the idea of a fixed vehicle replacement schedule seems much more realistic for commercial fleets than for households.

In Figure 12.1, both the aged vehicles and the new vehicles predicted by the fuel type choice model feed back to update the distribution of vehicles by fuel type and vintage for each record in the fleet file. The updated distributions are also aggregated across fleets in each forecast year to report the numbers of new vehicle, total vehicles, miles traveled and energy consumed by fuel type. Electricity consumption is also predicted by time of day, using the same recharging algorithm as is used in the personal vehicle model. The aggregation and output of forecast results is illustrated in Section 12.5.

12.3 The Vehicle Fuel Type Choice Model

The mail-out portion of the 1994 California Fleet Site Survey was sent to the person identified in the CATI portion of the Survey as being responsible for acquisition of the vehicles at the sample fleet site. These managers were asked to complete a maximum of two stated preference (SP) tasks. In each task, they were asked to allocate their future fleet purchases for a given vehicle type (e.g., car, minivan, etc.) by using a set of hypothetical future vehicles defined according to an experimental design.. The experimental design manipulated the vehicle fuel type and vehicle attributes, including vehicle capital cost, operating costs, range between refueling, refueling times, and fuel availability.

Each vehicle acquisition manager was presented with a vehicle allocation task for the most prevalent vehicle class at their site and for a second vehicle class selected at random from those classes in operation at the site. In the case of sites operating only one vehicle class, only one stated choice task was presented. In each task there were three vehicle types available, randomly selected from the four types included in the study, namely, gasoline vehicles, methanol vehicles (MV's), compressed natural gas vehicles (NGV's), and electric vehicles (EV's). For fleet sites operating more than one vehicle body type, all four fuel types were represented in the two allocation tasks presented to the respondent.

For each choice task, the respondent was asked to choose varying numbers of vehicles to make up their entire fleet. In analysis, the indicated number of vehicles assigned to each fuel type by the respondent was converted to a fraction of the total number of vehicles for that vehicle body type and used as a weight in a maximum likelihood estimation procedure. A weight of zero was assigned to fuel types that were not picked at all by the respondents.

Multinomial logit analysis was used to estimate choice utility coefficients for the attributes presented in the SP design. The resulting coefficients in the model implemented in the fleet vehicle forecasting module are presented in Table 11.9 of Chapter 11. The key vehicle attributes such as capital cost, operating cost, and range all showed significant effects. The remaining constants for EV, MV and NGV give an idea of the preferences of various fleet types for electric and alternative fuel vehicles relative to gasoline vehicles, over and above any differences in the experimental design variables.

12.4 Expansion and Correction of the Fleet Sample Data

The 1994 California Fleet Survey sample contains 1987 fleets with valid data on the numbers of vehicles owned by body type. Note that this sample is for the entire state. As of yet, no selection has been made for fleets only within the SCE service region.

The main stratification criterion used in recruiting the fleets was the total fleet size. Thus, fleet size was also used as the main variable in expansion to capture any systematic differences in sampling rates. Table 12.2 shows the expansion factor necessary in each fleet size class to meet the target. The target values represent the SCE region only, and are based on data supplied by the California Energy Commission. (CEC) The expansion factors are in the range 2.1 to 3.3 in all classes except the smaller fleets, where the factor is 7.74. Either (a) there was a sampling/screening bias against small fleets or (b) small fleets are more common in the SCE region than elsewhere in the state. Based on the sampling procedures followed, the first reason seems most likely.

The last two columns in Table 12.1 show that the average fleet sizes within each size category are similar in the sample and in the target data. The total unexpanded sample average is higher (65 vs. 46 vehicles) because of the lower proportion of small fleets in the sample.

After each fleet was expanded using the appropriate size class factor from Table 12.1, a further expansion was done to arrive at the correct total number of vehicles in each body type class. For each of the six body classes distinguished in the fleet survey (shuttle buses are grouped with full size vans), the target number of vehicles was determined from the body type distribution in the CEC target data for years 1985 to 1992 (the years with the most reliable data) applied to the target total number of fleet vehicles (432604).

The results in Table 12.2 show ratios of the targets to the sample totals in the range 1.5 to 2.2 for all classes except medium-duty trucks, which has a much lower factor of 0.76. Respondents may have a included a wider range of truck sizes in their survey estimates than what had been intended. In general, however, the factors are greater than 1.0 because the survey sample data includes only 2 body types per fleet, at most, while many fleets actually have more than that. The very similar adjustment factors for all classes except medium-duty trucks suggest that there was very little sampling bias for particular types of vehicles. In total there are 3743 fleet / vehicle type records in the sample, with each record representing an average of 116 vehicles after expansion and adjustment.

Table 12.1 Expansion by Fleet Size Class (Targets in SIZETARG.TXT)

Fleet Size Category	Fleets in Sample	Fleets in Target Data	Ratio Target to Sample	Avg. Size in Sample	Avg.Size in Target Data
10-19 veh.	744	5759	7.74	14	15
20-59 veh.	777	2497	3.21	31	38
60-119 veh.	238	561	2.36	78	85
120-499 veh	192	408	2.13	212	239
>500 veh.	36	118	3.26	995	1150
TOTAL	1987	9343	4.70	65	46

Table 12.2 Adjustment Factors by Vehicle Class (Targets in CLASTARG.TXT)

Vehicle Body Type Class	Sample Records	Sample Vehicles*	Target Vehicles	Ratio Target to Sample	Adjusted Veh/Record
compact car / station wagon	799	84159	173755	2.06	217
mini/compact van	283	9925	15022	1.51	53
full size van/shuttle buses	559	29143	64245	2.20	115
compact pickup	542	20622	36689	1.78	68
full size pickup	992	57320	120970	2.11	122
medium-duty truck	568	28846	21923	0.76	39
TOTAL	3743	230015	432604	1.88	116

*Note: Sample vehicles are shown as adjusted by the expansion factors form the fleet size groupings as described in Table 12.1.

In addition to the expansion weight, three important fleet / vehicle type characteristics for forecasting are:

***(1) The average miles driven per vehicle per year*.** This was taken directly from the reported figure on the survey record. In cases where no response was given, VMT for that particular fleet / vehicle type was estimated using a regression equation, as reported in Golob et al. (1995 WCTR).

(2) The current distribution of vehicles across fuel types: This was also taken directly from the survey questions regarding electric and alternative fuel vehicles already owned. As one might expect, there were very few such non-gasoline vehicles reported in the data.

(3) The current distribution of vehicles across vintages: No reliable external data could be found to estimate this distribution, which is a key input to the vehicle aging and replacement algorithm. (The CEC target data tends to have inconsistent classification of vehicle types across vintages.) An alternative approach was to use the data directly from the survey records - respondents were asked at what age vehicles of each type are typically replaced. Of the 3743 fleet / vehicle type records in the sample , 2957 (80%) have a valid response, either directly in terms of the age at replacement or else indirectly in terms of the VMT at replacement divided by a valid response for average VMT (the regression -based VMT estimates were not allowed as valid in this case.) In cases where both types of responses were given, an average of the two was used. Valid responses for replacement age were assumed to be those between 0 and 20 years.

For the cases with missing data, a regression-based estimate of replacement age did not seem appropriate. Since we are trying to estimate the vintage distribution, an estimate which is a regression to the mean would change the shape of the distribution. As an alternative method, the distribution across replacement ages was tabulated using the valid responses for each body type/fleet type combination, and a random draw from the appropriate distribution was made to fill in each case of missing data. The numbers of vehicles and the average age at replacement for each vehicle type / fleet type combination are shown in Tables 12.3 and 12.4 below.

Once we know the replacement age N for a given fleet / vehicle type record, we can either assume that (a) all vehicles are replaced every N years, or else that (b) 1/N of the vehicles are replaced each year. For a single fleet, neither is likely to be wholly realistic. Since each record in the expanded sample represents several actual fleets, however, assumption (b) seems most realistic. So, it is assumed that initially, in the base year, 1/N of the vehicles fall in each of the N most recent vintage years, and that no vehicles are older than age N. When added across the sample, this method gives an accurate total vintage distribution. (Note: Replacement age is assumed to be independent of fuel type - the same rule is applied to all fuel types.)

Table 12.3 Expanded Numbers of Vehicles by Fleet Type and Vehicle Type

Vehicle Type / Fleet Type	Car / Wagon	Comp. Pickup	Full size Pickup	Mini Van	Full size Van	Med-duty Truck	TOTAL
Agriculture	949	989	9191	88	425	697	12339
Construction	4799	3421	21939	99	2893	3109	36259
Manufacture	20308	2861	9381	2286	4878	3982	43695
Transp/Comm	13567	1839	15130	3952	22544	1900	58932
Retail/Wholesle	6965	2568	2415	1006	3299	3113	19296
Auto Industry	4890	1259	2583	197	907	572	10409
Banking/Insurnc	16623	110	114	377	718	39	17981
Service Industry	12119	3435	7679	1447	6376	1857	32912
City/County Gvt	77795	8077	25553	1156	4509	2356	119446
Local Services	2024	5985	15173	900	8127	2221	34430
Schools	5920	2081	9785	733	6890	485	25895
Others	7797	4064	2027	2781	2750	1593	21012
TOTAL	173755	36689	120970	15022	64245	21923	432604

Table 12.4 Average Age at Replacement by Fleet Type and Vehicle Type

Vehicle Type / Fleet Type	Car / Wagon	Compact Pickup	Full size Pickup	Mini-Van	Full size Van	Md-duty Truck	TOTAL
Agriculture	6.17	8.92	8.31	12.44	12.50	10.30	8.48
Construction	8.27	9.51	8.88	9.33	10.46	9.06	9.00
Manufacture	5.30	6.11	7.93	5.87	6.21	9.27	6.41
Transp/Commu	5.67	5.30	8.27	8.23	9.74	7.28	8.11
Retail/Wholesle	4.86	7.48	7.97	6.39	7.68	11.11	7.16
Auto Industry	4.22	7.01	7.30	4.60	8.96	5.82	5.83
Banking/Insrnce	4.25	4.62	4.00	5.80	4.80	8.22	4.31
Service Industry	5.37	7.59	7.95	6.60	8.95	8.25	7.11
City/County Gvt	7.24	8.21	8.69	7.74	7.67	9.51	7.68
Local Services	8.59	9.00	8.96	6.24	8.56	9.60	8.82
Schools	5.76	13.92	12.91	9.77	12.14	14.77	11.10
Others	4.99	8.15	7.02	3.23	5.85	3.00	5.53
TOTAL	6.18	8.32	8.83	6.51	9.07	8.91	7.68

After the data has been completed, and the current vehicles are split across vintages assuming 1/N of the vehicles in age up to N (the replacement age), resulting fleet sample file FLEET3.TXT has the following 115 fields per record:

1- 20: the number of gasoline vehicles in vintage years 1 to 20, from newest to oldest
21- 40: the number of methanol vehicles in vintage years 1 to 20
41- 60: the number of CNG only vehicles in vintage years 1 to 20
61- 80: the number of dual-fuel CNG vehicles in vintage years 1 to 20
81-100: the number of electric vehicles in vintage years 10 to 20

101: the vehicle type class
102: the fleet organization type
103: the unexpanded size of the actual fleet, in total
104: the average VMT per vehicle per year, in miles
105: a dummy for current on-site refueling facilities
106: a dummy if on-site refueling facilities are possible
107: a dummy for use in courier, express mail
108: a dummy for use in other types of deliveries
109: a dummy for use in hauling equipment
110: a dummy for use in service and maintenance
111: a dummy for use in sales
112: a dummy for use in shuttling passengers
113: a dummy for use by employees outside working hours
114: a dummy for use in other purposes
115: the age at which vehicles are replaced, in years

12.5 Fleet Forecast Results

To forecast fleet vehicle demand for various fuel types, the fleet file is used along with the model coefficients, and the vehicle and fuel technology files. The model application was programmed in both Pascal and Matlab to provide a useful way of comparing results and finding errors.

Assuming that the fuel type choice distribution from the models applies only to newly purchased vehicles and not to the entire fleet, the results for 1995 to 2010 are as summarized below. All AFVs start to enter the market in 1997. Methanol seems unrealistically attractive, with a share of over 20% of new vehicles in each year. CNG first gets about 10% and jumps to 20% in 1998. Dual CNG is constantly around 5%, and electric increases from 2% up to over 5% in 2010. In terms of total vehicles, gasoline retains a share of over 80% until 1999, and the overall shares remain fairly stable around 62% after that.

Table 12.5 New Vehicle Percentage Shares

Fuel	Gasoline	Methanol	CNG only	CNG Dual	Electric
1995	100	0	0	0	0
1996	100	0	0	0	0
1997	81.57	0.81	9.24	5.99	2.39
1998	61.86	7.33	21.96	4.94	3.91
1999	61.57	8.14	21.34	4.79	4.17
2000	61.72	8.01	21.05	4.76	4.46
2001	61.51	7.99	20.98	4.77	4.76
2002	61.36	7.97	20.82	4.77	5.09
2003	61.25	7.96	20.63	4.75	5.41
2004	61.29	7.94	20.59	4.76	5.42
2005	61.32	7.93	20.54	4.76	5.45
2006	61.32	7.91	20.56	4.77	5.44
2007	61.32	7.91	20.57	4.77	5.44
2008	61.30	7.89	20.58	4.77	5.47
2009	60.97	8.35	20.50	4.71	5.46
2010	61.24	7.86	20.62	4.78	5.51

Table 12.6 Total Vehicle Percentage Shares

Fuel	Gasoline	Methanol	CNG only	CNG Dual	Electric
1995	98.86	0.43	0.67	0.01	0.03
1996	99.06	0.33	0.56	0.01	0.03
1997	95.70	0.41	2.24	1.16	0.48
1998	89.13	1.73	6.11	1.89	1.15
1999	83.40	2.97	9.25	2.60	1.77
2000	78.24	4.11	12.02	3.27	2.36
2001	73.82	5.16	14.30	3.83	2.89
2002	70.19	5.99	16.13	4.32	3.36
2003	67.21	6.70	17.59	4.73	3.77
2004	64.82	7.28	18.70	5.10	4.11
2005	63.11	7.79	19.48	5.27	4.35
2006	62.04	8.13	19.84	5.44	4.54
2007	61.33	8.42	20.04	5.53	4.69
2008	60.79	8.59	20.22	5.61	4.79
2009	60.32	8.83	20.33	5.66	4.87
2010	60.02	8.92	20.42	5.71	4.93

12.6 How to Run Fleet Forecast Program

The file containing the forecast codes for the fleet is called FLEET5.M. To run this program follow these steps: (1) open MATLAB application, (2) type **pwd** and hit return key, (3) type **cd d:** and hit return key <if d is the hard drive where fleet programs are stored>, (4) type **fleet5** and hit return key. The program looks for various input files (all in the same subdirectory as where FLEET5.M is stored and creates output file containing the fleet forecast for fuel/vintage as shown in Table 12.5. All the input files as well as the output files are shown in Appendices B and C. The key is to store all these files including FLEET5.M in the same subdirectory. An example of an execution of FLEET5.M program is shown in Appendix B.

13. ACKNOWLEDGMENTS

The authors wish to thank those who have worked diligently to make this study a success:

Lyn Long, Jane Torous, Seyoung Kim, Soheila Crane, Camilla Kazimi, Dawn Hagerty, Paula Nohalty, Weiping Ren, Hongyan Sheng, Matt Weatherford, Ziggy Bates, Ann Marie DeFeo, Arleene Parsons, and Will Recker of the University of California, Irvine. Support for Camilla Kazimi and Weiping Ren was also provided by the University of California Transportation Center. Kazimi's Ph.D. Dissertation (Kazimi, 1995) is the basis for Chapter 6 of this report, and Ren's Ph.D. Dissertation (Ren, 1995) is the basis for Chapter 8.

Allan Biegaj, Mitchell Chan, Luke Lau, and Susie O'Bryant of the University of California, Davis.

Mark Bradley of the Hague Consulting Group and Bradley Research and Consulting.

Satoshi Fujii and Toshiyuki Yamamoto of Kyoto University.

Ernest Morales, Richard Rice, Depak Nanda, Debbie Brodt, and Tony del la Sota of Southern California Edison Co.

Gary Occhiuzzo, Mike Jaske, and Chris Kavalek of the California Energy Commission.

The authors are solely responsible for all of the views expressed here and for any errors.

14. REFERENCES

Aitchison, J. and S. Silvey (1957). The generalization of probit analysis to the case of multiple responses. *Biometrika*, 44: 131-140.

Ashford, J.R. (1959). An approach to the analysis of data for semi-quantal responses in biological response. *Biometrics*, 15: 573-581.

Batelle (1994). Vehicle Exhaust Emissions -- Early Mileage Results. Clean Fuel Demonstration Project, Statistical Analysis Report No. 4. Batelle Memorial Institute, Columbus, OH, for the South Coast Air Quality Management District.

Bates, J. (1988). Econometric issues in stated preference analysis, *Journal of Transport Economics and Policy.*

Becker, Gary S., landis, Michael and Robert Michael (1977). An economic analysis of marital instability. *Journal of Political Economy*, 85: 1141-1187.

Beggs, S. and S. Cardell (1980). Choice of smallest car by multi-vehicle households and the demand for electric vehicles. *Transportation Research*, 14A: 389-404.

Beggs, S., S. Cardell and J. Hausman (1981). Assessing the potential demand for electric cars. *Journal of Econometrics*, 16: 1-19.

Ben-Akiva, M. and S.R. Lerman (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand.* MIT Press, Cambridge, MA.

Ben-Akiva, M. and T. Morikawa (1990). Estimation of travel demand models from multiple data sources. *Transportation and Traffic Theory.*

Bennet, Neil G., Bloom, David E. and Patricia H. Craig (1989). The Divergence of Black and White Marriage Patterns. *American Journal of Sociology*, 95 No. 3: 692-722.

Bentler, P.M. and Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88: 558-606.

Berg, M. (1985). The potential market for electric vehicles: results from a national survey of commercial fleet operators. *Transportation Research Record,* 1049.

Berg, M., M. Converse, D. Hill (1984). Electric Vehicles in Commercial Sector Applications: A Study of Market Potential and Vehicle Requirements. Institute for Social Research, University of Michigan for the Detroit Edison Company.

Berkovec, J. (1985). Forecasting automobile demand using disaggregate choice models. *Transportation Research,* 19B: 315-329.

Berkovec, J. and J. Rust (1985). A Nested logit model of automobile holdings for one vehicle households. *Transportation Research*, 19B: 275-285.

Blau, Francine D. and Marianne A. Ferber (1986). T*he Economics of Women, Men and Work*, Prentice-Hall, Englewood Cliffs, New Jersey.

Bloom, David E. and Anne R. Pebley (1982). Voluntary Childlessness: A Review of the Evidence and Implications. *Population Research and Policy Review* 1: 203-224.

Bloom, David E. and James Trussell (1984). What are the Determinants of Delayed Childbearing and Permanent Childlessness in the United States?. *Demography*, 21 No. 4: 591-611.

Bollen, K.A. (1989). *Structural Equations with Latent Variables*. Wiley, New York.

Boomsma, A. (1983). *On the Robustness of LISREL (Maximum Likelihood Estimation) against Small Sample Sizes and Nonnormality.* Amsterdam: Sociometric Research Foundation.

Booz, Allen, and Hamilton, Inc. (1983). Forecasting household demand for light duty motor vehicles. Report prepared for Oak Ridge National Laboratories, Contract No. 40X-40485C.

Bradley, M. and A. Daly (1983). Estimation of logit choice models using information from both stated preferences and revealed preferences. Discussion paper for the Duke International Symposium on Choice Modeling and Behavior, July.

Browne, M.W. (1982). Covariance structures. In D.M. Hawkins, ed., *Topics in Multivariate Analysis:* 72-141. Cambridge: Cambridge University Press.

Browne, M.W. (1984). Asymptotic distribution free methods in analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37: 62-83.

Brownstone, D.. D. S. Bunch and T. F. Golob (1994). A demand forecasting system for clean-fuel vehicles. Presented at the OECD International Conference, Towards Clean Transport: Fuel Efficient and Clean Motor Vehicles, Mexico City, March .

Bumpass, Larry L. (1990). What's Happening the the Family? Interactions Between Demographic and Institutional Change. *Demography*, 27 No. 4: 483-498.

Bumpass, Larry L., Sweet, James and Andrew Cherlin (1991). The Role of Cohabitation in Declining Rates of Marriage. *Journal of Marriage and the Family*, 53 No. 4: 913-927.

Bumpass, Larry L., Sweet, James and Teresa Castro Martin (1990). Changing Patterns of Remarriage. *Journal of Marriage and the Family*, 52 No. 3: 747-756.

Bunch, D. S., M. Bradley, T. F. Golob, R. Kitamura and G. P. Occhiuzzo (1993). Demand for clean-fuel personal vehicles in California: A discrete-choice stated preference study. *Transportation Research,* 27A: 237-253.

Bunch, D. S., J. J. Louviere, and D. Anderson (1994). A Comparison of Experimental Design Strategies for Multinomial Logit Models: The Case of Generic Attributes. Working Paper UCD-GSM-WP#01-94, Graduate School of Management, University of California, Davis.

Calfee, J.E. (1985). Estimating the demand for electric automobiles using fully disaggregated probabilistic choice analysis. *Transportation Research*, 19B: 287-302.

California Air Resources Board (1992). Approval of motor vehicle pollution control devices (new vehicles). *Title 13, Div. 3, Chapter 1, Art. 2, California Code of Regulations.*

California Air Resources Board (1994). Zero-Emission Vehicle Update: Draft Technical Document for the Low-Emission Vehicle and Zero-Emission Vehicle Workshop. CARB, Sacramento, March.

Chang, H. L. and P. Jovanis (1990). Formulation accident occurrence as a survival process. *Accident Analysis and Prevention*, 22 No. 5: 407-419.

Cherlin, Andrew J. (1992). *Marriage, Divorce, Remarriage*, Cambridge, MA, Harvard University Press.

Cohanim, S., M. Hoggan, R. Sin, L. W. Jones, M. Hsu, S. Tom (1991). Summary of Air Quality in California's South Coast and Southeast Desert Air Basin, 1987-1990, South Coast Air Quality Management District, August 1991.

Crow R. and Ratchford B. (1977) Classical qualitative choice and new goods: An application to electric automobiles. Unpublished paper, Electric Power Research Institute, Palo Alto, California.

Dickson E. and B. Walton (1977). A scenario of battery/electric vehicle market evaluation. Unpublished Stanford Research Institute Report, SRI, Menlo Park, CA.

Duncan, G. J. and M. S. Hill (1985). Conceptions of Longitudinal Households. *Journal of Economic and Social Measurement*, 13: 361-375.

Easton Consultants, Inc. (1991). Natural Gas Vehicle Fleet Market Study. Unpublished report prepared for the American Gas Association and the Natural Gas Vehicle Coalition.

Efron, B. and R. Tibshirani, 1993. *An Introduction to the Bootstrap.* Chapman and Hall, New York.

Ermisch, J. (1988). An Economic Perspective on Household Modelling. Chapter 3 in *Modelling Household Formation and Dissolution*, edited by Nico Keilman, Anton Kuijsten and Ad Vossen, Clarendon Press, Oxford: 23-39.

Farrell, M.J. (1954). Demand for passenger cars in the United States. *Royal Statistical Society Journal,* 117A.

Flinn, C. J. and J. J. Heckman (1982). Models for the Analysis of Labor Force Dynamics. *Advances in Econometrics*, 1: 35-95.

Flinn, C. J. and J. J. Heckman (1983). Erratum and Addendum to Volume 1: Models for the Analysis of Labor Force Dynamics . *Advances in Econometrics*, 2: 219-223.

Fowkers, T. and M. Wardman (1988). The Design of stated preference travel choice experiments, with special reference to inter-personal taste variations. *Journal of Transport Economics and Policy*, 22.

Giuliano, G., K. Hwang, D. Perrine, and M. Wachs (1991). Preliminary evaluation of Regulation XV of the South Coast Air Quality Management District. Presented at Annual Meeting of Transportation Research Board, Washington DC, January 13-17.

Glick, Paul C. (1990). American Families: As They Are and Were. Sociology and Social Research, 74 No. 3: 139-145.

Glick, Paul C. and Sung-Ling Lin (1987). Remarriage After Divorce, Recent Changes and Demographic Variations. *Sociological Perspectives*, 30, No. 2: 162-179.

Glick, Paul C. and Arthur J. Norton (1977). Marrying, Divorcing, and Living Together in the U. S. Today. *Population Bulletin*: 3-39.

Golob, T.F. (1990). The dynamics of travel time expenditures and car ownership decisions. *Transportation Research*, 24A: 443-463.

Golob, T.F., D.S. Bunch and D. Brownstone (1995). A vehicle usage forecasting model based on revealed and stated vehicle type choice and utilization data. Presented at Seventh World Congress on Transport Research, Sydney, Australia, July 16-21.

Golob, T.F. and L.D. Burns (1978). Effects of transportation levels of service on automobile ownership. *Transportation Research Record No. 673*: 137-144.

Golob, T.F., S. Kim and W. Ren (1995). A structural model of vehicle use in multi-vehicle households. Presented at Annual Meeting of Transportation Research Board, Washington, DC, January 23-28. *Transportation Research (in press).*

Golob, T. F., R. Kitamura, M. Bradley and D. S. Bunch (1993). Predicting the market penetration of electric and clean-fuel vehicles. *The Science of the Total Environment,* 134: 371-381.

Golob, T. F., J. Torous and S. Crane (1995). Precursors of demand for alternative-fuel vehicles: results from a survey of 2,000 fleet sites in California. In *Environmental Vehicles '95: Proceedings of the January 23-25 Conference, Dearborn, MI.* ESD, The Engineering Society, Ann Arbor, MI.

Golob, T.F., J. Torous, D. Brownstone, S. Crane, D.S. Bunch and M. Bradley (1995). Commercial fleet demand for alternative-fuel vehicles: Results from a stated-choices survey of 2,000 fleet operators in California. Presented at Seventh World Congress on Transport Research, Sydney, Australia, July 16-21.

Gourieroux, C., A. Holly, and A. Monfort (1982). Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters, *Econometrica*, 50, No. 1.

Greene, D.L. (1989). Motor fuel choice: An econometric analysis. *Transportation Research,* 23A: 243-254.

Greene, D.L. (1990). Fuel choice for multi-fuel vehicles. *Contemporary Policy Issues,* 8: 118-137.

Hall, Jane, Arthur Winer, Michael Kleinman, Frederick Lurmann, Victor Brajer, and Steven Colome (1992). Valuing the Health Benefits of Clean Air. *Science*, 225: 812-817.

Hausman, J., B. Hall and Z. Griliches (1984). Econometric models for count data with an application to the patents-R&D relationship. *Econometrica,* 52: 909 - 938, 1984.

Hausman, J. and D. McFadden (1984). Specification tests for the multinomial logit model. *Econometrica*, 52: 1219-1240.

Heaton, Tim B. and Cardell K. Jacobson (1994). Race Differences in Changing Family Demographics in the 1980s. *Journal of Family Issues,* 15, No. 2: 290-308.

Hensher, D.A. (1982). Functional measurements, individual preferences and discrete choice modeling: Theory and application. *Journal of Econometric Psychology,* 2: 323-335.

Hensher, D.A. (1985). An econometric model of vehicle use in the household sector. *Transportation Research*, 19B: 303-314.

Hensher, D.A. (1992). Integrating revealed preference and stated response data into a jointly estimated hierarchical mode choice model. Presented at the Seventeenth Australian Transport Research Forum Conference, Canberra, 7-9 October.

Hensher, David A. (1994). The Timing of Change for Automobile Transactions: A Competing Risk Multispell Specification. working paper presented at the International Conference on Travel Behavior, Santiago, Chile, June 1994.

Hensher, D.A., P.O. Barnard, and T.P. Truong (1988). The Role of Stated Preference Methods in Studies of Travel Choice, *Journal of Transport Economics and Policy*, 22.

Hensher, D.A., P.O. Barnard, and F.W. Milthorpe (1990). An Empirical Model of Household Automobile Holdings. *Applied Economics*,.

Hensher, D.A., P.O. Barnard, and F.W. Milthorpe (1990). An Empirical Model of Household Automobile Holdings. *Applied Economics.*

Hensher, D.A. and Manefield, T. (1982). A structured-logit model of automobile acquisition and type choice. *Proceedings of the 7th Australian Transport Research Forum, Hobart, Tasmania:* 625-654.

Hensher, David A. and Fred L. Mannering (1994). Hazard-based Duration Models and Their Application to Transport Analysis. *Transport Reviews*, 14 No. 1: 63-82.

Hensher, D.A. and V. Le Plastrier (1985). Towards a dynamic discrete-choice model of household automobile fleet size and composition. *Transportation Research,* 19B: 481-495.

Hensher, D.A. and N.C. Smith (1986). A structural model of the use of automobiles by households. *Transport Reviews,* 6: 87-111.

Hensher, D.A., N.C. Smith, F.W. Milthorpe and P.C. Barnard (1992). *Dimensions of Automobile Demand.* Elsevier, Amsterdam.

Hill, D. (1987). Derived demand estimation with survey experiments: Commercial electric vehicles. *The Review of Economics and Statistics,* 69: 277-285.

Hill, D. (1993). Demand for Commercial Electric Vehicles: A Regional Pilot for a National Survey. Unpublished manuscript, Survey Research Institute, University of Toledo.

Hill, M. (1992). *The Panel Study of Income Dynamics: A Users Guide.* Sage Publications, Newberry Park, CA.

Hocherman, I., J. Prashker, and M. Ben-Akiva (1983). Estimation and use of dynamic transaction models of automobile ownership, *Transportation Research Record No. 944:* 134-141.

Huang, D.S. (1964). A Microanalytic model of automobile purchase, *Research Monograph, No. 29.* Bureau of Business Research, Graduate School of Business, University of Texas at Austin.

Huang, D.S. (1966). A multi-cross-section investigation of demand for automobiles, *Research Monograph, No. 31.* Bureau of Business Research, Graduate School of Business, University of Texas at Austin.

Janosi, P.E. de, (1959). Factors influencing the demand for new automobiles, *Journal of Marketing*, Vol. 23.

Johnson, T. (1975). The Structure of markets for new and used automobiles. Unpublished doctoral thesis, Department of Economics, University of Washington.

Johnson, T. 1978). A Cross-section analysis of the demand for new and used automobiles in the United States, *Economics Inquiry*, 16: 531-548.

Johnson, N.L. and S. Kotz (1970a) *Distributions in Statistics, Continuous Univariate Distributions -- 1.* John Wiley & Sons, New York.

Johnson, N.L. and S. Kotz (1970b) *Distributions in Statistics, Continuous Univariate Distributions -- 2.* John Wiley & Sons, New York.

Jöreskog, K.G. and D. Sörbom (1993a). *LISREL 8 User's Reference Guide.* Scientific Software, Chicago.

Jöreskog, K.G. and D. Sörbom (1993b). *PRELIS 2 User's Reference Guide.* Scientific Software, Chicago.

Kain, J., and M. Beeseley (1965). Forecasting car ownership and use, *Urban Studies.*

Kain, J., and G. Fauth (1977). Forecasting auto ownership and mode choice for U.S. metropolitan areas. Department of City and Regional Planning, Harvard University, Cambridge, MA.

Kalbfleisch, J. D. and R. L. Prentice (1980). The Statistical Analysis of Failure Time Data, Wiley Publishing, New York.

Karfisi E., C. Upton, and C. Agnew (1978). The Impact of electric passenger automobiles on utility system loads, 1985 - 2000. Report EA-623 to Electric Power Research Institute, research project 751-1.

Kazimi, C. (1995). A microsimulation model for evaluating the environmental impact of alternative-fuel vehicles, Unpublished Ph.D. Dissertation, Department of Economics, University of California, Irvine.

Kazimi, C. and D. Brownstone (1995). Competing risk hazard models for demographic transitions. To be presented at the World Congress of the Econometrics Society, Tokyo, Aug. 21-29.

Kroes, Eric P., and Robert J. Sheldon (1988). Stated preference methods: An introduction. *Journal of Transport Economics and Policy*, 22.

Krupnick, Alan and Paul Portney (1991). Controlling Urban Air Pollution: A Benefit Cost Assessment. Science, 252: 522-528.

Lancaster, T. (1990). The Econometric Analysis of Transition Data, Cambridge University Press.

Lancaster, Tony (1979). Econometric Methods for the Duration of Unemployment. Econometrica, 47 No. 4: 939-956.

Land, Kenneth and Andrei Rogers (1982). M*ultidimensional Mathematical Demography, proceedings from the Conference on Multidimensional Mathematical Demography,* University of Maryland, March 1981, edited by Kenneth Land and Andrei Rogers, Academic Press, Inc.

Lave, C. and J. Bradley (1980). Market share of imported cars: a model of geographic and demographic determinants, *Transportation Research,* 14A: 379-388.

Lave, C., and K. Train (1979). A Disaggregate model of auto-type choice, *Transportation Research*, 13A: 1-9.

Lerman, S. and M. Ben-Akiva. A behavioral analysis of automobile ownership and modes of travel. Report No. DOT-05-3005603, prepared by Cambridge Systematics, Inc., for the U.S. Department of Transportation, Office of the Secretary, and the Federal Highway Administration.

Lillard, Lee A. (1993). Simultaneous Equations for Hazards. *Journal of Econometrics*, 56: 189-217.

Louviere, J. (1988). Conjoint analysis modeling of stated preferences: A review of theory, methods, recent developments and external validity. *Journal of Transport Economics and Policy*, 12: 93-120.

Louviere, Jordan J., Conjoint analysis modelling of stated preferences. *Journal of Transport Economics and Policy*, pp. 93-119, January 1988.

Macro International (1992). NGV Market Assessment. Unpublished Report, conducted for the New England Gas Association.

Maddala, G.S. Limited dependent variable models using panel data. *The Journal of Human Resources*, XXII.3, pp. 307-338.

Miaou, S., P. Hu, J. Young (1992). Fleet Vehicles in the United States: Composition, Operating Characteristics, and Fueling Practices. Paper prepared by Oak Ridge National Laboratory for US Department of Energy. ORNL-6717.

Mannering, F. (1983). An econometric analysis of vehicle use in multi-vehicle households. *Transportation Research,* 17A: 183-189.

Mannering, F. L. and K. Train, Recent Direction in Automobile Demand and Modelling, *Transportation Research,* Vol. 19B, No.4, pp.215-236, 1985.

Mannering, F. L. and C. Winston (1985). Dynamic empirical analysis of household vehicle ownership and utilization. *Rand Journal of Economics*, 16: 215-236.

Martin, Teresa Castro and Larry L. Bumpass (1989). Recent Trends in Marital Disruption. *Demography*, 26 No. 1: 37-51.

McFadden, D. L., C. Winston and A. Borsch-Supan (1985). Joint estimation of freight transportation decisions under non-random sampling. In A.F. Daugherty, ed. *Analytical Studies in Transport Economics:* 137-160. Cambridge University Press, Cambridge.

McCarthy, P.S. and R. Tay (1989). Consumer valuation of new car attributes: an econometric analysis of the demand for domestic and Japanese and Western European imports. *Transportation Research,* 23A: 367-375.

Merz, Joachim (1991). Microsimulation - A Survey of Principles, Developments and Applications. *International Journal of Forecasting*, 7: 77-104.

Meyer, Bruce D. (1990). Unemployment Insurance and Unemployment Spells. *Econometrica*, 58 No. 4: 757-782.

Mogridge, M. (1978). The Effect of the oil crisis on the growth in the ownership and use of cars. *Transportation,* 7: 45-68.

Moore, Kristin A. and Thomas M. Stief (1991). Changes in Marriage and Fertility Behavior, Behavior Versus Attitudes of Young Adults. *Youth & Society*, 22 No. 3: 362-386.

Morgan, S. Phillip, Lye, Diane N. and Gretchen A. Condran (1988). Sons, Daughters, and Risk of Marital Disruption. *American Journal of Sociology*, 94 No. 1: 110-129.

Morikawa, T. (1993). Estimation of discrete choice models from serially correlated RP and SP data. Discussion paper for the Duke International Symposium on Choice Modeling and Behavior, July.

Nesbitt, K. (1993). What Can Alternative Fuel Mandates for Fleets Really Accomplish? Paper presented at the conference, Transportation and Energy: Strategies for a Sustainable Transportation System, Asilomar Conference Center, Pacific Grove, CA, August 22-25.

Parrish, M. (1994). EPA to Toughen Emissions Rules in the Northeast. Los Angeles Times, The Times Mirror Company, September 14, 1994: D1.

Pyke, Karen D. (1994). Women's Employment as a Gift or Burden, Marital Power Across Marriage, Divorce, and Remarriage. *Gender & Society*, 8 No. 1: 73-91.

Ren, W. (1995). A vehicle transactions choice model for use in forecasting vehicle demand for alternative-fuel vehicles conditioned on current vehicle holdings, Unpublished Ph.D. Dissertation, Department of Economics, University of California, Irvine.

Ren, W., D. Brownstone, D.S. Bunch and T.F. Golob (1995). A personal vehicle transactions choice model for use in forecasting demand for future alternative-fuel vehicles. Presented at Annual Meeting of the Transportation Research Board, Jan. 23-28, Washington. *Research in Transportation Economics* (in press).

Rima and Van Wissen (1987). *A Dynamic Model of Household Relocation, A Case Study for the Amsterdam Region*, Free University Press, Amsterdam.

Rinduss, Ronald R. and Craig St. John (1983). Social Determinants of Age at First Birth. *Journal of Marriage and the Family*, pp.553-565.

Runzheimer International (1993). Business fleet refueling assessment: final report. Runzheimer International Ltd., Northbrook, IL, for the Gas Research Institute.

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*, 28, No. 2.

Stanford Resarch Institute (1978). *Electric Vehicle News*, May.

South Coast Air Quality Management District (SCAQMD) and Southern California Association of Governments (SCAG) (1994). Air Quality Management Plan: Meeting the Clean Air Challenge, SCAQMD, Diamond Bar, CA, August.

Spanier, Graham B. and Paul C. Glick (1981). Marital Instability in the United States: Some Correlates and Recent Changes. *Family Relations*, 31: 329-338.

Stanley, Sandra C., Hunt, Janet G. and Larry L. Hunt (1986). The Relative Deprivation of husbands in Duel-Earner Households. *Journal of Family Issues*, 7 No. 1: 3-20.

Steinberg, Dan and Philip Colla (1993). CTM: A Systat Compatible Version of the Continuous Time Modeling Program. Salford Systems, San Diego.

Train, K. (1980). The potential market for non-gasoline-powered vehicles. *Transportation Research*, 14A: 405-414.

Train, K. (1980a). A Structured logit model of auto ownership and mode choice. *Review of Economic Studies*: 357-370.

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

Treachman, Jay D. and Mark D. Hayward (1993). Interpreting Hazard Rate Models. *Sociological Methods & Research*, 21 No. 3: 340-371.

Thorton, Arland (1988). Cohabitation and Marriage in the 1980s. *Demography*, 25 No. 4: 497-508.

US Department of Commerce: Bureau of the Census. (Census) (1990). *Truck Inventory and Use Survey*. US Government Printing Office. Washington.

U.S. Department of Energy (DOE) (1994). *Taking an Alternative Route*. Argonne National Laboratory, Energy Systems and Decision Information Systems Division, Argonne, IL.

U.S. Environmental Protection Agency (EPA) (1990). *Clean Air Act Amendments of 1990: Detailed Summary of Titles*. U.S. Government Printing Office, Washington.

Wachs, M. and Levine, N. (1985). Vehicle Fleets in the South Coast Air Basin: Results of a Survey Performed for the SCAQMD. Urban Innovations Group, Los Angeles.

Wardon, M. (1988). A Comparison of revealed preference and stated preference models of travel behavior. *Journal of Transport Economics and Policy*, 22.

Wharton Econometric Forecasting Associates, Inc. (1977). An Analysis of the automobile market: modeling the long-run determinants of the demand for automobiles, Vol. 1-3. Prepared for the U.S. Department of Transportation, Transportation Systems Center, Cambridge, MA.

Winston, C. and F. Mannering (1984). Consumer demand for automobile safety. *American Economic Review,* 74, 316-319.

van Wissen, L.J. and T.F. Golob (1992). A dynamic model of car fuel type choice and mobility. *Transportation Research*, 26B: 77-96.